LUIGI FABBRIS
Editor

Effectiveness of University Education in Italy

Employability, Competences, Human Capital

Physica-Verlag

A Springer Company

Effectiveness of University Education in Italy

Luigi Fabbris (Editor)

Effectiveness of University Education in Italy

Employability, Competences, Human Capital

With 71 Figures and 138 Tables

Physica-Verlag
A Springer Company

Editor

Prof. Luigi Fabbris
University of Padua
Statistics Department
Via C. Battisti 241
35121 Padua, Italy
luigi.fabbris@unipd.it

Library of Congress Control Number: 2006931360

ISBN-10 3-7908-1749-X Physica-Verlag Heidelberg New York
ISBN-13 978-3-7908-1749-2 Physica-Verlag Heidelberg New York

Physica-Verlag is a part of Springer Science+Business Media

springer.com

Typesetting: Camera ready by the author
Cover-design: Erich Kirchner, Heidelberg
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig

SPIN 11803072 Printed on acid-free paper 43/3100YL 5 4 3 2 1 0

Preface

The Italian university system, along the mainstream of the European higher education systems, is adopting efficiency and effectiveness as criteria for process programming and result managing.

For this purpose, both the national system headed by the Ministry of Education, University and Research, and the universities, within their managerial autonomy, activated since the 1990s a systematic assessment of their research and teaching activities and of other services.

The assessment on a large scale of aspects of the educational system started in Italy when the Ministry imposed in 1999 that the teaching of each course of Bachelor and Master programmes be evaluated by the student and that synthetic marks be sent to a National Board for University System Evaluation (CNVSU) for comparative and intervention purposes.

The compelling need to assess the course quality not only demolished the atavic distrust that part of the academics have had against the assessment of the teaching, but also started, in a concrete way, the participation of students to activities dedicated to them. Moreover, the widespread consciousness of teaching assessment, as a tool for empowering the higher education system, entered the Ministry and universities' agenda.

Even though the use of the course evaluations, for intervention purposes, still is limited to the most fanciful and fearless faculties and universities, the process has started and it will not be easy to circumvent.

A second assessment strategy is that of the so-called "external effectiveness" of university education. This process was started by some scouting universities and is now customary to the university system. It consists of the assessment, some time after degree achievement, of the graduates' employment and professional career.

The employability assessment is a strategic upgrade. First, it focuses upon the external-to-university life, and academic self-reference is much smaller a risk than it would be from just an internal evaluation. Second, it refers to the outcomes instead of the university educational processes. Third, professionalism becomes a daily 'gold standard', instead than just a faded backstage element of university education.

The search of a possible correspondence between education and labour implies the ascription of a social value, as opposed to just individual value, to the university education. Employability and successful career attainments should be in the background of the university aims together with that of developing culture and civic spirit of students. All this should be kept in mind while designing study programmes and the services that take care of students before, during and after their university paths.

The ascription of social values to education hints that an athenaeum is a social entity integrated in its economic and social contexts. This concept pervades the European strategies endorsed by the EU Member Countries in Bologna (1999), Lisbon (2000) and Barcelona (2002). These strategies aim at integrating education with wider economic and social strategies at all levels, European, national and local, and that of considering social and working development while designing the study programmes.

The labour society demands graduates who possess not only a theoretical education, but also adequate professional competences, namely a mix of knowledge, skills and attitudes that contribute to their success at work and in the social life at large.

Hence, work values and educational aims seem to be the two sides of the same coin. The higher the value ascribed to fast employment, in a good professional position and with chances of using the competences achieved, the higher the mirror value attributed to education.

In this volume – starting from the analysis of some Italian universities – we wanted to unveil if and how it is possible to measure and infer upon the quality of the education achieved by graduates by surveying their transition to the labour market and the short and mid term professional exits.

The volume is partitioned according to the issues dealt with. First, we present a set of papers on the effectiveness of university education, then another on employability of graduates with a discussion on the opportunity to consider job quality, rather than the employment rate, as the main assessment indicator, and finally a set of papers on competence analysis for external assessment purposes.

The first paper of this volume explores the indicators of quality evaluation that can be drawn from surveys on teaching assessment, on graduates' placement, on teachers and employers, and on linkage between databases. Those wishing to provide their own data for education assessment may find it useful.

In the volume, we mixed up study cases and methodological issues. The study cases are the analysis of the educational effectiveness of the Universities of Cagliari, Florence, Foggia, Milan-Bicocca, Padua and Naples "Federico II" performed by researchers of the same universities. Other study cases concern the teaching of Statistics and that of Social work at the university level. We can state that this volume is a university-driven contribution for the definition of concepts, models and methods for education's quality evaluation.

Those who like indulging with the modelling of assessment data can read almost all papers. The Authors adopted for their analyses the 'fuzzy' approach, structural equation models, linear, logistic and Boolean regression models, chain graph models, principal strata approach to causal analysis, forward search for robust approach to linear regression, latent growth curves and other growth models, various approaches to factor analysis (exploratory, confirmatory, categorical, Rasch, multiple correspondence), segmentation analysis with various criterion variables, cluster analysis methods, semantic marking of symbolic objects and other textual analysis approaches, survival analysis with Kaplan-Meier parametric and non-parametric methods, utility function analysis, and the Delphi-Shang method for data collection.

There is also a paper on the compatibility criteria between occupations and university curricula. It includes a basic scheme for comparing work requirements and the professional counterparts achievable at school. Compatibility is based on competences required for work and those achievable at school. In other words, if each job is described in terms of professional competences and each juxtaposed curriculum is described in terms of comparable competences, competences become the keys for the job-to-curriculum compatibility evaluation.

We close our volume with two papers that deal with the evaluation of the so-called ‘human capital’ that is the indicator of long-term external outcome of an educational programme. The possibility of using human capital as an external assessment indicator is a controversial issue. However, the papers support the idea that it is possible to compute it with current data. Its large-scale applicability for the differential evaluation of university programmes’ effectiveness remains, instead, a research issue.

The researches presented in this volume gave some methodological answers and uncovered political questions. We touched policy issues and highlighted the interpenetration between methodological and political choices. Let us remind some that are crucial for both methodology and policy:

- For comparative evaluation purposes, should we consider the profiles of students that enter a study programme, and how? Should we reward a school that offers an education to less gifted students, or one that emphasizes quality, even with strict entry selection rules?
- Which are the ways to provide the students with a flexible education, the teaching of very basic subjects or a more balanced skill-promoting programme? As training for a job forms a student less than training on the job, is there an appropriate balance between skill-development during the school years and that at work?
- Because the assessment of education may be realized at the levels of single study programmes, faculties, athenaeums and the country, a complex informative system should be installed to disentangle the specific effects and understand if, where and how to intervene. Nonetheless, how should one account for the social and economic environments of the various educational levels?

The papers of this volume had been written, together with others published elsewhere, within a research funded by the Italian Ministry of Education, University and Research. The research was titled: “The transition from university to work and the valorisation of graduates’ competences: models and methods for multidimensional analysis of determinants”.

Its acronym was OUTCOMES, whose spell is “*Occupation as a University Target and Careers of Outgoing-graduates Maximizing Educational Skills*”. This slogan expresses some of our feelings with regard to university education.

We are particularly grateful also to the University of Foggia for the grant that allowed us to publish these papers.

Let me finally thank all the Authors for their scientific contributions and all those who helped me in preparing this volume. In particular, I wish to thank Prof. Francesco d’Ovidio for his pagination work.

Padua, June 2006 The editor

Luigi Fabbris
Statistics Department, University of Padua, Italy

Contents

Informative Sources for the Evaluation of the University Education Effectiveness in Italy

Stefano Campostrini, Simone Gerzeli

Dept. of Applied Statistics and Economics "Libero Lenti", University of Pavia, Italy

Summary. The evaluation of the effectiveness of a study programme refers to its outcomes, and may be measured with the level of satisfaction of the programme objectives. In an educational programme, we can recognize three macro objectives referred to students:
i) the formation of specific competences;
ii) general competences and individual cultural development;
iii) capability in finding a suitable job.
The measurement of these components is based on the construction and analysis of several indicators from surveys on teaching assessment, or on placement, or on teachers and employers, and, too, on linkage between databases.
Keywords: Education effectiveness; Evaluation; University; Informative sources.

1. The evaluation of effectiveness

The evaluation of the effectiveness of education involves several elements of complexity, from the definition of effectiveness itself to the multiplicity of programmes that pursue educational objectives. We will not linger over the delicate question of the definition of effectiveness, accepting, as an operational definition "*the degree to which the educational objectives are achieved*".

Hence, the evaluation of effectiveness implies a clear identification of the objectives and of the expected results, as well as of their measurement. The complexity of an educational programme makes the definition of the objectives even more important: only a clear, specific, precise declination of the objectives allows us to define the concept of effectiveness to which the programme refers. We could say that the definition of the objectives also contains the definition of effectiveness.

Thus, if an initial definition of the objectives leads to a precise determination of the possible results, a subsequent effort must be undertaken to define effectiveness, by indicating *what it possible to measure is*. Often at the initial stage just proxy measures are used, measures close enough to the concepts expressed by the objectives.

In defining objectives, particularly in the social context, we often use concepts (for example, the “quality of teaching”) which are not directly measurable. However, each of these can be made operational by identifying the components of the concept itself, which make it easier to identify possible measures, which nevertheless are proxy ones.

In the evaluation process, inadequate conclusions may be drawn if the outcomes are undefined. This is why it is risky to pay attention to only one outcome indicator and to evaluate the effectiveness with even small variations of this indicator.

The multi-dimensional nature of the social world implies a multiplicity of indicators. The basic idea behind evidenced-based approaches refers, on the one hand, to the impossibility of precisely measuring the complex situation the programme is on about and, on the other, the importance of choosing evidence indicators in order to make sufficiently well informed decisions.

A possible guideline for an operative definition of effectiveness may help also in identifying the information needed for its evaluation: how to make explicit the central feature of the definition of the objectives specific to the university programme and how it may be measured. The process is typically circular: from an initial definition of objectives, we determine which of these can be made operational and measured, and this often leads to the redefinition of the objectives themselves.

Thus, going back to the definition of effectiveness as the “degree to which the objectives are achieved” it may be added[1] “*provided the objectives are clearly defined*”.

The theory thus indicates that it is difficult to adhere to a single definition of effectiveness whatever this may be, that is valid for every project and every context; but this is due to the very same definition of the objectives that the educational programme intends to set forth.

An important corollary of this issue is that no evaluation of effectiveness is possible if the programme itself has not stated what concept of effectiveness it refers to. Hence, each evaluative project should explicitly mention which concept of effectiveness is referred to, if it is a partial or overall measure of it, and to what extent the “proxy” indicators are valid, reliable and sensitive.

---

[1] This should also solve the problem for defining effectiveness as the “satisfaction of needs”. It is the definition of the objectives that should account for how much and in what way the programme needs are to be satisfied; this should not be dealt with through the search for their (ex post) effective satisfaction outside of the objectives set forth by the project.

2. The components of university education effectiveness

In order to manage the complexity of programme evaluation, we attempt to determine (macro) objectives typical to the evaluation process without discussing the peculiar aspects of each of these. This approach aims to identify the macro-categories that are measurable.

University education typically unfolds in ways that are often quite diverse depending on the framework of application, and involves profoundly different objectives. Nevertheless, it is possible to identify some common features, often because it is not the basic objectives that change but their relative importance. Thus, for example, the objective of finding a job cannot be the primary of university education, but at the same time, it cannot be ignored. Likewise, in a high-level university education (for example, a Master programme) the cultural development of the person, while not a priority, is an important objective anyway.

In Table 1, we outline the macro objectives that an educational programme could typically seek to achieve. We will try to bring out typical features defined by both the theory and the evaluative practice.

Table 1. A framework for defining effectiveness in university education programmes.

Objectives	*What we can measure*	*How*
Learning / development of specific knowledge	Satisfaction of the students during and after their studies Self-perception of educational advancement Assessment of knowledge/skill acquisition Perception of employers	Customer satisfaction surveys on teaching and placement Partial and final verifications *Ad hoc* surveys of employers Impact studies of different programmes Surveys of teachers Surveys of experts
Development of cross-occupation competences/ "formation of a person"	Satisfaction of students during and after their studies Self-perception of (*ex post*) educational achievement/ usefulness of cross-occupation competences Perception of employers Enrolment in more advanced studies	Customer satisfaction surveys on teaching and placement *Ad hoc* surveys on employers Surveys on teachers Surveys on experts
Employment capabilities	Employability, such as time required to enter the labour market Quality of work (contract / position / professional nature) Coherence of the job with the studies	Placement surveys Comparison of archives Surveys on employers Surveys on teachers Surveys on experts

Each of the items presented in the table would deserve a full and deep discussion. This exceeds the aims of this paper, which seeks to take advantage of this synthetic presentation to discuss the possible sources for evaluating the effectiveness of University education, their pros and relative limitations. Moreover, the list is not meant to be exhaustive.

The scheme obviously does not consider specific cases and the classification cannot capture all the complexities of particular educational programmes. On the other hand, such an approach perhaps cannot reveal other important issues faced by university education[2]. If we consider only three macro objectives, we accept the risk of leaving out many important aspects. However, given the complexity of the problem, we have sought some solid footing to start from, with the possibility of subsequently widening the scope and detail of these considerations.

3. The sources for evaluating effectiveness

We will now examine the main problems and possible methodological solutions that can be encountered when using the typical instruments for evaluating the effectiveness of educational programmes. We begin with an outline in order to discuss, though briefly, each element.

Table 2. Typical approaches to gather information on the effectiveness of education programmes, and related problems

Sources and instruments for evaluating effectiveness	*Possible problems*
Evaluation surveys of teaching methods	Subjective evaluation
Placement studies	Subjectivity Timing of data collection Selection bias
Linkage of archives	Data availability
Surveys on experts	Subjectivity / lack of quantitative elements
Surveys on professors	Subjectivity / partiality
Surveys on employers	Subjectivity Coverage and rates of response

[2] For instance, we lose the trade-off between quality of education and the increase in the number of educated individuals, a problem that is dealt with and not yet resolved by the reform of the educational system in Italy.

3.1 Evaluation surveys of teaching methods

The evaluation of teaching is quite a well-known approach around Europe. In Italy its application is recent (late 1990s), but it is routinely used in a most educational programmes, and is compulsory in public universities.

Even though the data-gathering instruments and the subsequent construction of indicators have become more refined, one would be wrong in thinking that the validity and sensitivity of these elements allow for indications of absolute effectiveness of the educational programmes being studied.

In fact, the specific role of this evaluation approach, especially when we limit ourselves to gathering information from assessments by students, can only represent a general stimulus for the individual teachers and a sort of alarm signal with regard to certain particularly critical situations.

It would be deceiving (and in many regards counterproductive) to attribute to these types of evaluation the role of an absolute judgement on the quality of teaching and, even more, on the single teacher. In fact, the reliability and sensitivity of the instruments can certainly allow us to make valid considerations at the macro level as well as regarding the “tails” of the distributions (the left one in particular, the relevant negative cases), while the validity of comparison in the central part of the distribution can be questionable. I.e., a teacher who receives an average score of 3 out of 10 is certainly a problematic case, while it is not necessarily true that the performance of a teacher with a score of 7.5 is below that of one whose score is 8.

Thus, the methodological solution to the problems linked to the subjectivity and partiality of the evaluations does not involve adopting some technical tricks but limiting ourselves to the role that evaluations of this type can play and understanding the importance of interpreting the results in light of other possible information, following an integrated approach (Nevo, 2001; Bernardi *et al*., 2005).

3.2 Placement studies

Placement studies are also being used in different educational contexts. Typically, at a distance of six months, and one to two years from graduation, students are interviewed regarding their satisfaction as well as to obtain information on their employment situation. It must be emphasized that opinions given spontaneously and after some work experiences (even unsuccessful ones) can provide an interesting point of view.

Obviously, all *ex post* assessments provide us with a lot of information to define overall judgements, and are not in competition with the approach presented in the previous section, which focuses more on formative evaluation.

Timing is certainly a critical element: studies too close to the end of the educational experience do not sufficiently reflect the students’ rethinking of their educational lives. However, if the surveys are too far from this experience, the rethinking can itself be strongly influenced by exogenous factors.

Once again, the specific evaluative objective indicates the appropriate timing with respect to this trade-off. In any event, experience seems to show a minimum distance of six months and a maximum of two years.

As far as the employment situation is concerned, we need again to pay attention to the value and the role that this type of informative source can play. The case of vocational training is emblematic. In many regional systems, evaluations on the employment capacity of formative programmes have been considered as points in favour for those seeking financing for their educational programmes. This use of the evaluation of results in selection processes is certainly positive and should be encouraged.

There is a mixture of aspects concerning both the definition of objectives and methodology. It is clear that, with regard to different study programmes available in a territory, a selection problem could hide the effectiveness – in terms of employment capabilities – of such study programmes. Students who choose subjects that are viewed as important by the labour market will certainly have more chances of success, for example in terms of employment rate six months after the end of the educational programme.

This typical problem of impact evaluation can be overlooked if we consider that the ability to select worthier and more appropriate students is itself part of the educational programme. Thus, we again have a problem of the definition of objectives: this would surely be a tenable approach.

Should we reward a school that offers a university education to less gifted students, or one that emphasizes quality, even with a rigid entry selection? If we do not consider selection problems, which can hide a result (positive or negative), we run the risk of wrongly judging the outcomes of a placement survey.

3.3 Linkage of databases

Linkage of databases could certainly be an interesting approach to overcome most of the above-mentioned problems. At a small cost we could compare the performance of persons that received a university degree with those without it (or who have had a different degree) by simply observing the social security agency status (INPS in Italy) after a certain period. When this has been applied to similar contexts, there have been encouraging results (see, among the others, Battistin *et al*., 2001).

Apart from the methodological problem of linkage, the issue here is often a “political” one, involving the willingness of different institutions. The data exists, but what is often missing is the willingness to link the data for comparative purposes. A non-trivial problem is data privacy, although it is frequently used as an excuse.

Basic information for the evaluation, in terms of internal effectiveness or output data, could also be found in the university administrative archives. These databases allow us to construct indicators to measure the output of the university education, for instance the number of students and degrees, the number of courses, etc.

Of course, these archives are useless for outcome evaluation, although some information on effectiveness can be obtained.

In Italy, the universities' capability of carrying the students to final degree is taken by the Ministry for University as an indicator of performance of the universities.

3.4 Studies on experts

Often the data on the effectiveness of educational programmes is limited to those directly involved (e.g. students and subsequently to graduates), who undoubtedly are aware of the progress in their educational experience, but can provide only a conditioned point of view. A useful complement to this data can be "qualitative" studies on experts.

Though not useful to quantify effectiveness, this information can nevertheless be precious not only to complete the data that has been gathered but also to provide a key to interpreting the success or the failure of an educational programme.

3.5 Surveys of teachers

In many countries (and Italy is one of these) teachers are rarely surveyed for their opinions, almost as if researchers were afraid to inconvenience those on the other side of the fence. Actually, teachers can also provide interesting information, if not on the final effectiveness of programmes, at least on those elements of the *process* that favour or obstruct such programmes.

In fact, it is becoming increasingly necessary to include process elements in *ex post* evaluations: these allow us to understand not only if an outcome has been achieved but also how and why, and under what conditions. Those who manage this process can certainly provide this information. Thus, it is not simply a question of asking teachers to do a sort of self-evaluation of their work but to involve them in a more comprehensive evaluation programme.

In this regard, we must quote the disputes between those who believe that an evaluation should be internal to the concerned organisation (i.e. the organisation that must "learn") and those who feel it should be external, requiring third-party evaluators. It depends on the evaluative objectives, on who requires the evaluation, and the point of view the study will take (Chen, 1996; Campostrini, 2001).

A third-party evaluator is certainly necessary to guarantee evaluations that are independent of the context and free of peculiar interests, among the other targets, when it is used to formulate judgements for programme funding.

In the same way, we cannot deny the validity of internal evaluations, undertaken as a standard practice by the organization itself (Patton, 1997).

3.6 Survey of employers

As far as the educational objective of labour market entry and specific skills training are concerned, the point of view of employers is surely important, if not fundamental.

The problems in this context are technical and cultural. In fact, the basic limit to employer surveys is the difficulty of getting employers to collaborate. It is not only the problem of interviewing who are busy, often on the move and have little time available for activities that take them away from work. It is not only the problem of selecting the most appropriate person to interview in highly complex organizations. The problem is (also and above all) cultural.

This is the case in Italy, where employers have shown little interest in educational programmes. Recently there have been some notable changes, but these involve training on the job more than training for a job. Even companies that believe strongly in training programmes, as opposed to basic education, prefer to train their personnel themselves to provide specific skills than turn to some training agency.

Some changes are taking place, and it could be worthwhile to study the methodological problems for this type of survey and to launch some pilot projects. In the meantime, we feel that the qualitative approach involving experts presented above still represents the best proposal.

4. Final remarks

In this paper, we have tried to systematise the complexity of an evaluation project of university educational activities. Our purpose was to provide a possible framework, discussing the typical sources for evaluation. This has been done with reference to the Italian situation. Giving these limitations, our work can represent only a possible starting point for approaching the matter.

Further research is needed to better define the effectiveness of educational measures, in particular university ones. Although some approaches (such as customer satisfaction or placement surveys) are well developed both from a theoretical point of view as well as from the methodological one, some others (e.g. surveys on employers or data linkages) need to be further studied.

References

BATTISTIN E., GAVOSTO A., RETTORE E. (2001) Why do subsides firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy. In: LECHNER M., PFEIFFER F. (eds) *Econometric Evaluation of Labour Market Policies*, Physica-Verlag, Heidelberg: 153-182.

BERNARDI L., CAMPOSTRINI S., PARISE N. (2005) L'impatto della valutazione della didattica sull'Università: il caso dell'Ateneo patavino. *Rassegna Italiana di Valutazione*, **31**: 103-123.

CAMPOSTRINI S. (2001) Metodi e tecniche nella valutazione di politiche e programmi nella realtà del nuovo millennio, *SIS 2001 Convegno Intermedio Processi e Metodi Statistici di Valutazione*, CISU, Roma: 65-74.

CHEN H.T. (1996) A comprehensive typology for program evaluation, *Evaluation Practice*, **17(2)**: 121-130.

NEVO D. (2001) School evaluation: internal or external?, *Studies in Educational Evaluation*, **27**: 95-106.

PATTON M.Q. (1997) *Utilization Focused Evaluation. The New Century Text.* Thousand Sage, Oaks.

A Fuzzy Measure of Satisfaction for University Education as a Key for Employment[1]

Corrado Crocetta* Giuseppe Delvecchio**

*Department of Economic, Mathematic and Statistic Sciences, University of Foggia, Italy
** Department of Electrotechnics and Electronics, Polytechnic of Bar, Italy

Summary. In this paper, we propose a fuzzy approach to measure the degree of satisfaction of graduates on the suitability of university education for working purposes. The indicators proposed come from two factors: the importance attributed to six aspects of university education and the quality of the service perceived by graduates. We use these indices to evaluate the first results of a telephone survey on a sample group of graduates at the University of Foggia. Using factor analysis, we show that all the aspects of university education can be summarised in just one dimension.
Keywords: Fuzzy indicators, Customer satisfaction, Graduates, University of Foggia.

1. Introduction

The evaluation of *customer satisfaction* implies the collection of one or more marks allotted to a service by its users. The allotment of these scores requires subjective evaluations, which may be though of as values on an ideal continuum and may be suitably summarised.

In this paper, we propose a *fuzzy* method to measure the satisfaction of graduates for their university education in terms of preparation for the labour market. Thus, we evaluate *ex post* the university education in order to improve the formative offer according to labour market needs.

---

[1] This paper is the result of the joint research of the two authors. C. Crocetta was responsible for the final editing of Sections from 1 to 4, 6 and 7, whereas G. Delvecchio was responsible for Sections 3 and 5.

Indeed, the didactic autonomy of universities imposes serious thoughts about the objectives and contents of the educational pathways through which the competences develop.

2. The survey

The University of Foggia has recently started the monitoring of placement and satisfaction of its graduates with respect to the effectiveness of the education received for preparing their insertion into work.

We will use the database formed through the integration of the data available in the archives of Student secretariat of the University of Foggia and that obtained from an *ad hoc* telephone survey.

Because of some difficulties in interviewing graduates by telephone (some of them graduated seven years before), we had to undergo some limitations in the data collection design. The initial purpose was to contact all graduates of the University of Foggia. We made up to seven attempts to contact the students at different hours before considering them unreachable. We carried out the 2,133 interviews from April 28th to May 26th, 2003. The percentage of completed interviews reached the 72.9% of the eligible population.

The questionnaire is composed of four parts: the first one is dedicated to graduates who worked before graduation, the second to those who are still seeking employment, the third to current or past professional experiences and the last to graduate satisfaction with services received, considering both the job searching and professional experiences.

In this paper, we will focus on the evaluation of graduate satisfaction with respect to the services received.

In order to measure the importance that each person assigns to the different aspects of university education and the quality of the teaching offered by each faculty, the interviewees were asked to give a score from 0 to 100 to both, importance and quality of service, with respect to the following aspects:

- basic teaching,
- specialist teaching,
- practical activity (practice, laboratories, job placement, etc.),
- use of equipment for improving expertise and research,
- *forma mentis,*
- methods and techniques learned during the course.

The answers to this part of the questionnaire required particular commitment from the interviewers because many interviewees found it difficult to answer the questions properly and required the interviewer to explain the exact meaning of questions. We chose a 0-100 scale in order to have a wide range of scores, thus providing greater accuracy without using decimal numbers.

It is not possible to carry out mathematical operations between values assigned to the importance and the quality to obtain a satisfaction index, because they are inconsistent between themselves and with satisfaction. For this reason, to summarise the two (quality and importance) scores, it is inappropriate to use the arithmetic mean. In the following (Section 3), we propose to estimate this value with a model-free fuzzy approach.

3. The fuzzy approach

In order to identify a fuzzy system it is to be chosen (Dubois & Prade, 1980; Bualoti *et al.*, 1999; Delvecchio, 2002):

- the type of system, and the way to activate fuzzy subsets (see Section 3.1),
- the *memberships* associated to the measures entering the system (Section 3.2),
- the *memberships* associated to the measures leaving the system (Section 3.3),
- the input-output association rules (Section 3.4),
- the type of defuzzifier (Section 3.5).

3.1 The fuzzy system used

In the fuzzy logic, the units (importance attributed, quality perceived and index of satisfaction with respect to six aspects of university education) are divided into fuzzy subsets.

We will indicate a low importance subset with $\mathbf{IA}_L$, a medium with $\mathbf{IA}_M$, a high with $\mathbf{IA}_H$, and a very high one with $\mathbf{IA}_{VH}$. The analogous subsets associated to the perceived quality is denoted with $\mathbf{Q}_L$, $\mathbf{Q}_M$, $\mathbf{Q}_H$, $\mathbf{Q}_{VH}$, and those associated to the index of satisfaction with *ten* subsets $\mathbf{I}_1$, $\mathbf{I}_2$, …, $\mathbf{I}_{10}$. Our fuzzy system is based on a method of fuzzy inference known as “Mamdani”[2]. The basic idea is to incorporate the experience of the researcher who designs the system[3].

---

[2] The main methods of fuzzy inference are those of Mamdani and Sugeno. The former (Mamdani & Assilian, 1975) is the most popular: it has the advantages of being intuitive and adaptable to human inputs. The latter (Sugeno, 1985) can be used to model any inference system in which the membership functions in output are linear or constant: it has the advantages of being computationally efficient, of working well with linear techniques or techniques of optimisation and adaptation, and of adapting well to mathematical analysis.

[3] «*Fuzzy systems allow us to construct systems in the ordinary language. To supply some fuzzy systems the common sense contained in the heads and maybe in the language of Cro-Magnon and Neanderthal man was sufficient*» (Kosko, 1995).

We have to establish the membership functions and the matrices of rules: i.e. the matrices that associate the value of the index (Section 3.4) to a particular combination of importance and quality. From a set of linguistic rules that describe the system, we obtain an algorithm (i.e. a system of fuzzy rules) whose words are defined as *fuzzy sets*.

The main advantages of this approach are the possibility of making feasible the knowledge based on empirical rules and on intuition, and also that of not requiring a process model (Kosko, 1995).

We applied the *correlation minimum encoding rule*, i.e., the fuzzy subset *A'* of a general fuzzy set *A*, activated by a determined value x_0 of the non fuzzy variable *x*, is obtained from *H* by cutting the upper part at $\mu_H(x_0)$, where $\mu_H(x)$ is the *membership function* of the input data associated to *H* (Kosko, 1992).

3.2 Memberships associated to the data entering the system

To define the functions that describe the degree of membership of input data to the fuzzy subsets we chose – for the sake of simplicity – triangular functions distributed in [0, 100].

In Figure 1, we visualise the *memberships* associated to importance. The ‘low’ fuzzy subset is associated to the membership constituted by the triangle of vertices (0, 0), (0, 1), (33.3, 0); the ‘medium’ one is associated to the membership made up of the triangle of vertices (0,0), (33.3, 1), (66.6, 0), and so on[4].

The quality issue is represented with similar memberships.

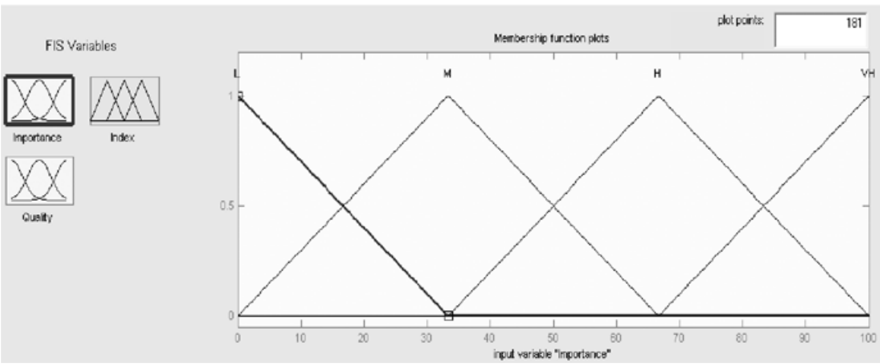


Figure 1. Representation of memberships associated to educational importance in our fuzzy system, as visualised in the implemented software.

4 The elements are not attributed to an ordinal class, because each element will belong to the four *fuzzy* subsets at the same time, but with a different degree of membership.

3.3 Memberships associated to data leaving the system

In Figure 2, we represent the memberships and their visualisation associated to the index of satisfaction (Mathworks, 1999). In this case, too, we decided to use the triangular type of membership for each of the ten *fuzzy* classes.

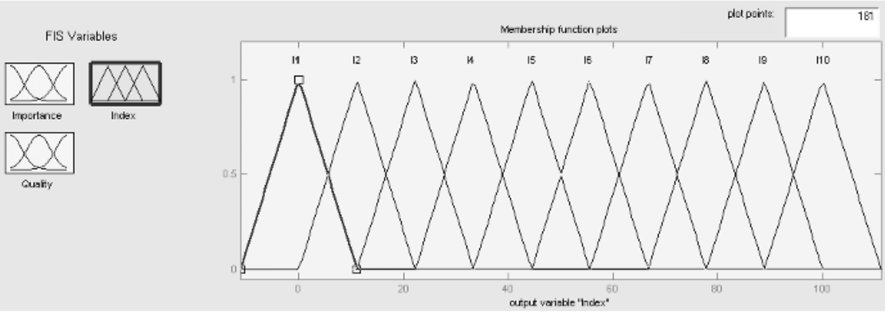


Figure 2. Representation of memberships associated to the index

3.4 Rules of input/output association

In the fuzzy approach, even the input/output rules of association are based on the operator's experience. For instance, in the case of a risk deriving from two concurring sources, a popular model is multiplicational: i.e. the total risk equals to the product of the risks of the two components (Fiorentini & Romano, 1995; Delvecchio, 2002).

However, the rules for the definition of importance and the quality of the labour market and to the education received from the university cannot be mathematical, but psychological. To define the rules, we used the opinion of a group of experts composed of a job psychologist, an anthropologist and two social statisticians [5] who agreed on the following criteria (Table 1):

Table 1. Matrix of rules constructed with expert opinions.

Perceived quality	Attributed importance			
	IA_L	IA_M	IA_H	IA_{VH}
Q_L	I_4	I_3	I_2	I_1
Q_M	I_4	I_4	I_5	I_5
Q_H	I_4	I_5	I_7	I_8
Q_{VH}	I_4	I_6	I_9	I_{10}

5 We express our thanks to professors Amelia De Lucia; Francesco Delvecchio, Giancarlo Tanucci and Ernesto Toma who collaborated to define the rules.

- the scores given to the importance of the aspects of the university education are independent of those assigned to the perception of quality;
- satisfaction is maximum when the highest level is attributed both to importance and quality;
- the bottom of satisfaction is when the quality is at the lowest level and the importance given to this aspect is highest;
- the satisfaction cannot decrease if the perceived quality increases, whatever the importance attributed.

3.5 The defuzzifier

The method chosen for defuzzifying (Cammarata, 1994) assumes as output the abscissa of the barycentre of the area subtended by the function that describes the fuzzy set output (the method is thus named *centroid*[6]).

In Figure 3, we describe an example of activation of rules relative to an aspect of university education, whose importance score is 10, quality score is 80, and satisfaction index is 41.9. Each row reports the three memberships (importance, quality, index of satisfaction) of the 16 rules described in Table 2. At the bottom right (under the sixteenth rule in the third column of the index of satisfaction) there is the output subset (obtained by "consolidating" the subsets in the third column, and outputting 16 rules), and the value of the index (the abscissa of the barycentre of the subtended area).

Let us consider, for instance, rule n. 7: if the importance is medium and the quality is high, then the index is I_5. The seventh row of Figure 3 shows diagrams of the three triangular memberships of medium importance, of high quality and of index I_5. The values in abscissa 10 and 80, by "activating" the subsets, determine the cutting of those triangular membership functions, thus generating some trapezoidal memberships (first and second diagram). The memberships of the output subset (third diagram) is obtained by cutting the respective triangular function of the I_5 index near the lower height of the previous two trapezoidal memberships that regard importance and quality (rule of *correlation minimum encoding*, see Section 3.1).

The global satisfaction index is the arithmetical mean of the *fuzzy* indices derived for the six aspects of university education analysed.

In Figure 4, we show how the index of satisfaction constructed with our fuzzy technique is connected to importance and the perceived quality. The relationship is not linear and when the perceived quality is low, the index of sat-

---

[6] This method (which must not be confused with the cluster analysis one), is popular because it does not fail with several maxima or a maximum limit, and it considers the information of the output set (Cammarata, 1994; Kosko, 1995). According to Kosko, in the FAM (Fuzzy Associative Memory) technique, this method is appropriate because ‹‹*we feel or perceive the centre of mass*››.

isfaction decreases with increases of importance; however, when quality if high, the index increases with increases of importance.

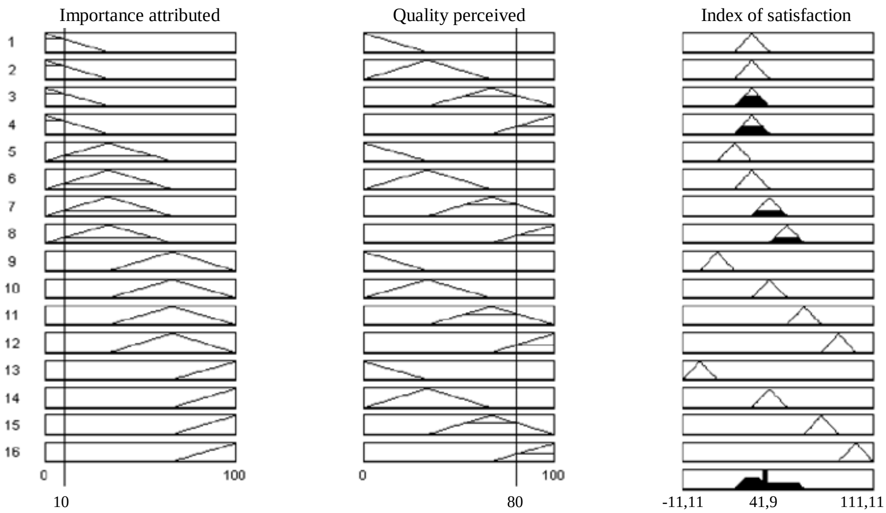


Figure 3. Example of activation of the rules in a fuzzy scheme (the score of the importance is 10, see the abscissa of the first column, and the perceived quality is 80, see the abscissa for the second column).

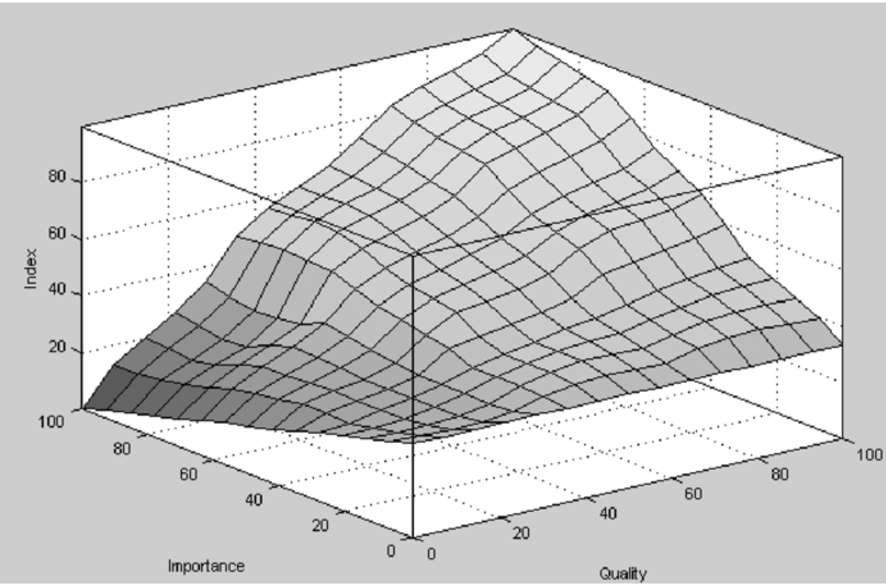


Figure 4. Representation of the index of satisfaction, obtained with fuzzy rules, according to the importance attributed and the quality perceived

Table 2. Average (M) and coefficient of variation (CV) of the scores assigned to importance, quality and index of satisfaction

Aspects of university education	Importance attributed		Quality perceived		Index of satisfaction	
	M	CV	M	CV	M	CV
1. Basic teaching	75.4	27.3	64.0	31.6	66.8	27.7
2. Specialised teachings	73.1	30.5	61.1	35.8	64.0	30.9
3. Practical activities (practice, labs, job placement)	76.5	34.8	49.2	53.8	53.8	44.5
4. Use of equipment for expertise & research	74.5	35.9	47.6	54.8	52.8	44.6
5. Forma mentis	77.8	26.5	67.4	32.5	69.9	28.9
6. Methods and techniques	72.6	30.5	58.3	37.3	61.5	32.3
Global index of satisfaction					62.9	24.2

For low values of importance, satisfaction does not change with changes of quality, whereas when importance is high, satisfaction increases rapidly with the increase of quality perception.

4. Index of satisfaction

Following the methodology described above, we computed – for each respondent – six indices of satisfaction for the university education. In Table 2, we show the means and the coefficients of variation of the six aspects.

On average, the importance of the different aspects rated higher than the perceived quality. The global index of satisfaction indicates that graduates are reasonably satisfied (62.9).

The interviewees attribute great importance to *forma mentis*, practical activity and basic teaching. This opinion is popular among researchers, considering the fact that the coefficients of variation of the fifth and first aspect are significantly lower than the others. The index of satisfaction regarding practical activity, however, shows a great variability.

Therefore, according to the interviewees, the university should give basic knowledge and integrate it with practical activities (such as internship, labs, and job placement) in order to prepare graduates to face the problems of working life. In general, graduates at Foggia University prefer less specialised courses.

The relations between the scores attributed to importance and to the perceived quality, for the six aspects, are evident in the correlation matrices (Table 3). All the values of these matrices are positive, and this indicates that there is a general agreement between all the variables. The only exceptions are practical activities and use of equipment for which there is a stronger linear relation both as regards to the importance attributed (r=0.75) and the perceived quality (r=0.69).

Table 3. Correlation matrices between the scores regarding the importance attributed and the quality perceived for six aspects of university education.

Importance attributed to the aspects of university education	Aspects					
	1	2	3	4	5	6
1. Basic teaching	1.00	0.52	0.31	0.29	0.43	0.43
2. Specialised teachings		1.00	0.48	0.45	0.45	0.50
3. Practical activities			1.00	0.75	0.48	0.55
4. Use of equipment				1.00	0.51	0.56
5. Forma mentis					1.00	0.63
6. Methods and techniques						1.00
Quality perceived attributed to the aspects of the education	Aspects					
	1	**2**	**3**	**4**	**5**	**6**
1. Basic teaching	1.00	0.55	0.34	0.27	0.38	0.41
2. Specialised teachings		1.00	0.48	0.38	0.39	0.47
3. Practical activities			1.00	0.69	0.34	0.48
4. Use of equipment				1.00	0.32	0.45
5. Forma mentis					1.00	0.54
6. Methods and techniques						1.00

Table 4. Correlation and partial correlation coefficients between the scores regarding the importance attributed, quality perceived and index of satisfaction.

Aspects of the university education	Correlation coefficients between			Partial correlation coefficients between	
	impor-tance and quality	importance and satisf. index	quality and satisf. index	Importance and satisfaction index | quality	Quality and satisf. index| importance
1. Basic teaching	0.46	0.67	0.94	0.78	0.96
2. Specialised teaching	0.46	0.61	0.95	0.63	0.95
3. Practical activity	0.34	0.41	0.97	0.33	0.96
4. Use of equipment	0.35	0.39	0.97	0.21	0.96
5. Forma mentis	0.51	0.63	0.97	0.64	0.97
6. Methods& techniques	0.39	0.53	0.96	0.59	0.96

Let us now verify if there is a relation between satisfaction index, importance attributed and perceived quality (Table 4).

The correlation coefficients between attributed importance and perceived quality for the different aspects vary between 0.34 and 0.51. The most correlated are those regarding the *forma mentis* and the basic and specialised subjects. The correlation between importance and quality for practical activities, use of equipment and methodology taught is less intense.

As regards the importance attributed to educational aspects and the corresponding indices of satisfaction, a larger correlation is observed between the theoretical aspects as compared with the practical ones.

The perceived quality has a much higher correlation with the index of satisfaction than the one regarding the importance attributed. The explanation may be found in the type of fuzzy rules used to construct the indices. We constructed the indices using fuzzy type rules of association and not combining the opinions.

According to the rules established by the experts, it was not possible to foresee the existence of a strong linear relation between the quality and satisfaction perceived (as measured by the *fuzzy* indices). Indeed, the rules assume a nonlinear link (Figure 4), so if the importance attributed to an aspect is low, the satisfaction remains constant even if the quality changes.

The high correlation coefficient between perceived quality and satisfaction may depend on the fact that if a person evaluates positively all the considered aspects, the overall satisfaction will be positive even if some aspects are marginally important.

The partial correlation coefficients between importance and the satisfaction indices highlight that, for a given quality, the overall satisfaction grows while the importance attributed to aspects increases. This agreement is larger for the more theoretical aspects of education.

For an analogous reason, the positive correlation between perceived quality and the indices of satisfaction, given the importance, allows us to state that satisfaction depend on the perceived quality.

It may be seen that the quality-and-importance scores are concentrated in the triangle limited by coordinates (0, 0), (100, 100) and (0, 100), an area within which the relations between the satisfaction index, the importance and quality are linear.

The correlation matrix between indices of satisfaction (Table 5) is similar to those regarding importance and the quality (Table 4). Still there is a strong relation between satisfaction and practical activities and specialised teaching,

Table 5. Correlation matrix between the indices of satisfaction regarding the six aspects of university education considered.

Indices of satisfaction of the aspects of university education	Indices of satisfaction						
	I1	I2	I3	I4	I5	I6	I
I1 - Basic teaching	1.00	0.50	0.29	0.25	0.36	0.40	0.59
I2 - Specialised teaching		1.00	0.43	0.36	0.36	0.47	0.69
I3 - Practical activity			1.00	0.69	0.32	0.47	0.76
I4 – Use of equipment				1.00	0.34	0.44	0.74
I5 - Forma mentis					1.00	0.57	0.67
I6 - Methods and techniques						1.00	0.79
I - Index of satisfaction (I)							1.00

and a similar agreement exists between the use of equipments for providing expertise and practical activities.

The link between the fundamental aspects of education, such as *forma mentis* and basic teaching, with practical activity and the use of equipment is rather weak. Instead, the satisfaction index is positively correlated with all the educational aspects.

5. Graduate satisfaction

Because the University of Foggia was only recently established, it seemed appropriate to investigate whether the indices of satisfaction vary according to study course, year of degree achievement, graduation mark, and employment position (Tables 6 and 7).

Table 6. Average scores of the importance attributed to six aspects of university education, by faculty, year of graduation, final graduation mark and current working position.

Faculty, year at grad., classes of grade and working position	***Aspects of university education***					
	Basic teaching	Specialist teaching	Practical activity	Use of equipment	Forma mentis	Methods & techniques
Faculty						
Agriculture	81.3	80.8	79.4	79.0	82.2	79.1
Economics	74.7	73.2	73.6	71.2	76.9	72.1
Law	75.2	71.8	77.0	75.7	77.5	71.6
Medicine and Surgery	75.9	78.2	85.3	78.9	84.1	79.3
Solar year at graduation						
Until 1995	79.2	74.1	83.4	88.8	87.3	79.4
1996	79.0	76.0	78.0	70.0	81.0	75.0
1997	78.0	75.0	79.0	76.0	81.0	74.0
1998	79.0	75.0	80.0	76.0	82.0	77.0
1999	78.0	76.0	77.0	76.0	80.0	75.0
2000	75.0	73.0	75.0	75.0	77.0	73.0
2001	75.0	72.0	74.0	74.0	76.0	71.0
2002	74.0	73.0	78.0	75.0	76.0	70.0
2003	69.0	66.0	69.0	66.0	72.0	70.0
Final graduation mark						
66 – 99	66.6	64.7	66.7	64.8	69.6	64.7
100 – 105	75.3	73.3	75.2	72.1	77.2	72.3
106 - 110 L	79.3	75.8	78.7	76.6	81.1	74.9
Current working position						
Employed	78.6	75.3	76.5	75.0	79.8	73.9
Not employed	72.3	71.0	76.4	74.1	75.9	71.3
University of Foggia	75.4	73.1	76.5	74.5	77.8	72.6

Table 7. Average scores of the quality attributed to six aspects of university education, by faculty, year of graduation, final graduation mark and current work position.

Faculty, year of grad., classes of grade and working position	**Aspects of the formation at university**					
	Basic teaching	Specialist teaching	Practical activity	Use of equipment	Forma mentis	Methods & techniques
Faculty						
Agriculture	68.2	66.0	53.1	49.9	67.2	62.0
Economics	63.6	59.8	47.5	45.0	67.3	59.0
Law	63.8	61.2	50.2	50.3	67.4	57.5
Medicine and Surgery	65.2	64.5	49.7	44.6	68.7	57.9
Solar year at graduation						
Until 1995	75.7	67.2	54.8	56.5	76.7	65.2
1996	64.1	62.2	55.2	45.6	66.7	59.0
1997	62.8	61.2	49.8	47.3	67.2	57.0
1998	66.0	59.7	50.2	47.4	70.7	58.4
1999	66.1	60.4	49.5	49.9	68.7	59.1
2000	63.6	60.6	47.2	47.9	68.0	59.0
2001	63.9	61.3	49.0	49.1	65.9	57.4
2002	62.3	61.7	47.5	45.1	65.3	56.7
2003	62.1	61.2	51.8	45.9	67.2	61.3
Final graduation mark						
66 – 99	61.6	59.2	48.1	49.2	66.3	58.3
100 – 105	63.5	60.4	48.3	46.2	66.9	57.7
106 - 110 L	67.9	64.2	51.2	47.6	69.5	59.2
Current working position						
Employed	65.4	61.6	49.5	47.8	68.5	59.2
Not employed	62.6	60.7	48.9	47.4	66.3	57.4
University of Foggia	64.0	61.1	49.2	47.6	67.4	58.3

It is evident that the opinions of graduates about the importance they attribute to education always exceed the quality they feel with it. In particular, the difference between those opinions (attributed importance and perceived quality) is generally greater for the aspects regarding the use of equipment and practical activity but lower for basic and specialist teaching.

Table 8 shows an overall positive opinion (provided by the global satisfaction index *I*) for all the faculties. The most satisfied graduates seem to be those of the Agriculture Faculty, whereas those of the other faculties are slightly more critical. Obviously, the judgement expressed for the aspects of university education varies according to faculty: for graduates in Agriculture the highest satisfaction level refers to basic teaching, whereas for graduates in Medicine and Surgery, Law and Economics the most important aspect is *forma mentis*.

Graduates of almost all the faculties are unsatisfied with their practical training and expertise for the labour market, thus requesting a greater commitment of the University management to take care of it.

Table 8. Indices of satisfaction regarding the 6 aspects of university education considered and index of global satisfaction, by faculty, year of graduation, working position and final graduation grade.

Faculty, year of grad., classes of grade and working position	***Indices of satisfaction***						
	I1 Basic teaching	***I2*** Specialised teaching	***I3*** Practical activities	***I4*** Use of equipment	***I5*** Forma mentis	***I6*** Methods & techn.	***I*** Global index
Faculties							
Agriculture	72.2	70.0	58.5	55.3	71.2	67.0	65.7
Economics	66.5	63.0	52.1	50.3	69.7	62.5	62.7
Law	66.5	63.7	54.2	55.3	69.7	60.2	62.9
Medicine and Surgery	68.0	68.4	55.4	50.5	71.5	62.0	62.7
Year of Graduation							
Up to 1995	76.3	70.2	56.2	59.4	79.9	68.0	69.5
1996	67.8	64.8	59.6	51.3	71.0	61.7	63.0
1997	66.7	64.8	54.3	53.3	70.8	61.7	63.9
1998	68.9	63.6	55.6	53.4	73.4	62.6	64.4
1999	69.2	64.1	54.5	55.1	71.3	62.5	64.3
2000	66.6	63.4	52.8	53.2	70.0	61.6	62.8
2001	66.8	64.2	53.2	54.4	68.6	60.8	62.6
2002	64.9	64.3	52.0	50.1	67.8	59.6	61.4
2003	63.6	61.8	54.2	49.4	68.1	62.7	61.5
Final graduation grade							
66 - 99	63.7	61.7	52.4	54.2	68.2	60.4	61.6
100 - 105	66.5	63.3	53.1	51.6	69.5	60.9	62.6
106 - 110 L	71.0	67.3	55.7	52.6	72.2	63.3	65.0
Curr. working position							
Employed	69.0	64.9	54.1	53.2	71.3	62.7	64.0
Not Employed	64.8	63.1	53.5	52.4	68.5	60.3	61.9
University of Foggia	66.8	64.0	53.8	52.8	69.9	61.5	62.9

In Table 8, by observing the distribution of the *satisfaction indices* as regards year of graduation, some interesting regularity and some even more eloquent peculiarities can be observed.

We have to specify that some people who graduated at Foggia University in first years after its foundation transferred from other universities, and had the opportunity of using equipment and the organisation of other long established universities. The unexpected positive opinions expressed by these graduates on various aspects of university education can be easily understood. In the first years, the family-like environment resulting from the small number of students and the relaxed relation with lecturers may have positively influenced the students.

The global satisfaction index assumes higher values for graduates of the first years and decreases during the last ones. The less theoretical aspects of university education (*use of equipment and practical activity*) always differ from the others. However, considering the evolution of these scores in time, it is inevitable to ask why the significant investments made to equip the young University of Foggia with adequate structures did not affect the evaluation of those who graduated in the last years.

Of course, as the refreshers increased the intrinsic validity of the teaching and the presence of adequate structures increased their importance, graduates became more demanding, and evaluative scores lowered. The evaluation was good until the University obtained its independence from the University of Bari in 1999.

The final graduation marks had also influenced the students' opinions: those who have achieved a better university education were more positively oriented on all aspects, the only exception being training on the use of equipment. This confirms that the lower the opportunity of using university resources, the lower the educational performances.

The only aspects that even graduates with the best marks evaluate insufficiently are those regarding practical activity and use of equipment.

Finally, employed graduates express more positive opinions than the unemployed ones on all the aspects of University education. Also between these two categories, there is little discrepancy of evaluation regarding practical activities and use of equipment.

6. Factor analysis

By comparing the different indices of satisfaction, it emerges the existence of a latent structure influencing the different aspects considered.

To make this structure clear and minimise the number of evaluative dimensions, we applied a factor analysis on the index scores. Since there is only one eigenvalue greater than 1, we decided to retain the first factor alone, which explains 58.3% of the variability. In Table 9, we present the standardised factor loads and the communalities of each index of satisfaction, by faculty.

From the factor load analysis (Fabbris, 1997), we can see that practical aspects are juxtaposed to theoretical-methodological ones, and name the factorial axis *theory vs practice*.

The factorial loads of the Athenaeum all exceed 0.5 and are distributed in an almost uniform manner.

We can conclude that the university education quality, evaluated by the satisfaction indices, lead back to a single latent factor spanning from practical to the theoretical-speculative aspects.

Indeed, the factorial loadings regarding the faculties undergo only slight variations compared to the one obtained for the whole Athenaeum. The presence of a *theory-practice* factor is evident for all the faculties.

Table 9. Factor loadings and communalities of the satisfaction indices regarding university education, by faculty.

Aspects of university education	Faculties				University of Foggia
	Agriculture	Economics	Law	Medicine & surgery	
	Factor loadings				
1. Basic teaching	0.60	0.64	0.65	0.69	0.65
2. Specialist teaching	0.57	0.74	0.73	0.76	0.73
3. Practical activities	0.77	0.73	0.75	0.86	0.76
4. Use of equipment	0.73	0.66	0.71	0.80	0.70
5. *Forma mentis*	0.56	0.64	0.71	0.72	0.68
6. Methods & techniques	0.79	0.73	0.80	0.84	0.78
	Communalities				
1. Basic teaching	0.36	0.41	0.42	0.48	0.42
2. Specialist teaching	0.33	0.55	0.53	0.57	0.53
3. Practical activities	0.59	0.53	0.57	0.73	0.57
4. Use of equipment	0.53	0.44	0.51	0.64	0.49
5. *Forma mentis*	0.32	0.41	0.51	0.52	0.47
6. Methods & techniques	0.63	0.54	0.64	0.70	0.61

Nevertheless, some specifications are needed. For graduates in Agriculture there is a clear cut between theoretical aspects and practical ones and, in this case, the index of satisfaction regarding *forma mentis* has a lower factorial weight than the one regarding basic teaching.

For graduates in Economics, the factorial loads regarding specialist teaching assume an apparently anomalous position. Those related to the indices regarding methods and techniques, practical activities and specialist teaching almost coincide. This is reasonable if we consider that in this faculty, practical activities concern almost exclusively the specialist teaching.

The factor loads make it possible to position each faculty and the entire Athenaeum along a continuum (Figure 5). Even if the deviations are small, indicating a basic homogeneity between the considered faculties, we can see that the graduates in Law and Agriculture are at the extremes of the axis. We can conclude that the satisfaction of graduates in Agriculture is more influenced by their opinion on practical activities than it is for Law graduates.

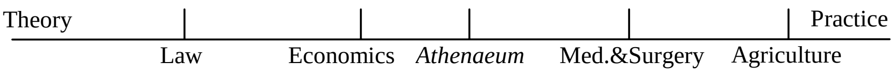


Figure 5. Factorial weights regarding the theory-practice factor

7. Conclusions

Thanks to the *fuzzy* approach, we have summarised in a single index the scores attributed by the University of Foggia graduates on the importance and quality of education for employment and career purposes.

The index depends on the *fuzzy* rules provided by experts and thus has a high degree of subjectivity. Subjectivity is both the strength and weakness of the fuzzy method because, on the one hand, it allows the solution of complex problems with incomplete information, but, on the other, it does not guarantee objectivity of the results obtained.

Although fuzzy techniques have shown enormous progress in the last few years, applications in the field of customer satisfaction are few and there is a lack of specific methodological suggestions for this kind of application.

Our analysis has also pointed out that graduates express positive opinions on the quality of the education received, although there are some critical points.

The lowest graduate satisfaction was for the following aspects: use of equipment for specialisation and research, and practical activities (practical exercises, labs, work placement, etc.). The lack of space and adequate structures for research depend on the recent establishment of the Athenaeum. We hope these problems can be solved in the next few years.

References

BUALOTI R., DELVECCHIO G., TROVATO M. (1999) A fuzzy reasoning approach for distribution automation, *Proceedings of the IEEE Budapest Power Tech'99 Conference (*29 August – 2 September, 1999, Budapest, Hungary: 249.

CAMMARATA S. (1999) *Sistemi fuzzy: un'applicazione di successo dell'intelligenza artificiale*, Etaslibri, Milano, 1994.

DELVECCHIO G. (2002) Un approccio fuzzy per la valutazione del rischio da mobbing. In: PUGGIONI G: (ed) *Modelli e metodi per l'analisi di rischi sociali e sanitari*, Cleup, Padova: 248-266.

DELVECCHIO G. (2003) A method for estimating the environmental impact of a wind farm by fuzzy logic techniques, *Proceedings of the International Conference on Renewable Energy and Power Quality - ICREPQ 2003, 9-11 April, 2003,* Vigo, Spagna.

DUBOIS D., PRADE H. (1980) *Fuzzy Sets and Systems: Theory and Application*, Academic Press Inc., San Diego (CA).

FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.

FIORENTINI C., ROMANO A. (1995) Riflessioni sulla valutazione del rischio ai sensi del Decreto Legislativo 626/94, *Sicurezza*, dicembre: 21-23.

KOSKO B. (1992) *Neural Networks and Fuzzy Systems: a Dynamical System Approach to Machine Intelligence*, Prentice-Hall, Englewood Cliffs.

KOSKO B. (1995) *Il fuzzy-pensiero, teoria e applicazioni della logica fuzzy*, Baldini & Castoldi, Milano.

MAMDANI E.H., ASSILIAN S. (1975) An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies*, **7(1)**: 1-13.

MATHWORKS (1999) *Fuzzy Logic Toolbox for Use with MATLAB*, User's Guide Version 2, MathWorks, Inc..

SUGENO M. (1985) *Industrial Applications of Fuzzy Control*, Elsevier Science Publications Co.

A Multilevel Analysis of Graduates' Job Satisfaction

Leonardo Grilli, Carla Rampichini

Statistics Department "G. Parenti", University of Florence, Italy

Summary. In this paper, we analyse some aspects of job satisfaction by means of a multilevel factor model, decomposing the factor structure into the graduate and degree programme components, using data from a survey on the 1998 graduates of the University of Florence. Due to the ordinal scale of the response variables, we adopt a multilevel factor model for ordinal variables. The results show that the factor structures at the graduate and study programme levels are not the same, although they are similar; the study programmes with extreme factor scores should be selected for a deeper investigation.
Keywords: Factor model; Job satisfaction; Multilevel model; Ordinal variable.

1. External effectiveness at Florence University

Nowadays it is relevant for the Universities to improve their efficiency and effectiveness, in order to ensure a good allocation of public funds, guarantee the rights of the students and their families to have good services and educational programmes, and, nonetheless, state the relevance of the University as a cultural, social and economic institution.

With this aim, the University of Florence has developed an evaluation system in the last years (Chiandotto *et al.*, 2004). External effectiveness is evaluated with respect to the employment results, such as the employment rate, the time span to the first job, the probability to find a job consistent with the acquired skills. The analysis of job satisfaction is a relevant part of the University evaluation. In the Italian context this issue is treated, among the others, by Santoro & Pisati (1996), Bini (1999), Mazzolli (2000), Bartolozzi (2001).

The main goal of the paper is to analyse and summarise the aspects of job satisfaction by means of a multilevel factor model (Goldstein & McDonald, 1988; Longford & Muthén, 1992), decomposing the factor structure into the graduate and study programme components. To this end, the data are taken from a survey conducted on the 1998 graduates of the University of Florence, interviewed about two years after the degree. Due to the ordinal scale of the response variables, a multilevel factor model for ordinal variables (Skrondal & Rabe-Hesketh, 2004; Grilli & Rampichini, 2006) is specified.

The structure of the paper is as follows. In Section 2 the model is defined, while in Section 3 the results of the analysis of job satisfaction of the 1998 graduates of the University of Florence, taken from a telephone survey conducted, about two years after the degree, are presented. Section 4 concludes our paper.

2. The statistical model

Let $Y_{ij}^{(h)}$ be the h-th ordinal variable (h=1, …, H) observed for the i-th subject (i=1, …, n_j) belonging to the j-th cluster (j=1, …, J). In the following, the subject level will be referred to also with the term ‘within’ and the cluster level with the term ‘between’. In the application presented in Section 3 the clusters are the study programmes, the subjects are the graduates and the ordinal variables are the ratings on 5 items of the questionnaire (H=5).

A two-level factor model for ordinal variables can be set up by defining two components, namely:

- a threshold model which relates a set of continuous latent variables $\tilde{Y}_{ij}^{(h)}$ to the observed ordinal counterparts $Y_{ij}^{(h)}$;
- a two-level factor model for the set of continuous latent variables $\tilde{Y}_{ij}^{(h)}$.

As for the threshold model, let assume that each of the observed responses $Y_{ij}^{(h)}$, which take values in {1, 2, …, C_h}, is generated by a latent continuous variable $\tilde{Y}_{ij}^{(h)}$ through the following relationship:

$$\left\{Y_{ij}^{(h)} = c^{(h)}\right\} \quad \Leftrightarrow \quad \left\{\gamma_{c^{(h)}-1}^{(h)} < \tilde{Y}_{ij}^{(h)} \le \gamma_{c^{(h)}}^{(h)}\right\}, \tag{1}$$

where the thresholds satisfy the inequality

$$-\infty = \gamma_0^{(h)} \le \gamma_1^{(h)} \le \ldots \le \gamma_{C_h-1}^{(h)} \le \gamma_{C_h}^{(h)} = +\infty .$$

The factor model can now be defined on the set of latent variables. A general formulation is (Goldstein & McDonald 1988; Longford & Muthén 1992):

$$\tilde{Y}_{ij}^{(h)} = \mu^{(h)} + \left[\sum_{m=1}^{M_u} \lambda_{u,m}^{(h)} u_{mj} + \delta_j^{(h)}\right] + \left[\sum_{m=1}^{M_v} \lambda_{v,m}^{(h)} v_{mij} + \varepsilon_{ij}^{(h)}\right]. \tag{2}$$

In this model the cluster level has M_u factors with corresponding loadings $\lambda_{u,m}^{(h)}$, while the subject level has M_v factors with corresponding loadings $\lambda_{v,m}^{(h)}$. Note that even if $M_u = M_v$ the factor loadings are generally different, the factors may have different interpretations.

Now it is convenient to express the general two-level model (2) for the latent variables in matrix notation:

$$\tilde{\mathbf{Y}}_{ij} = \boldsymbol{\mu} + \left[\boldsymbol{\Lambda}_u \mathbf{u}_j + \boldsymbol{\delta}_j \right] + \left[\boldsymbol{\Lambda}_v \mathbf{v}_{ij} + \boldsymbol{\varepsilon}_{ij} \right], \tag{3}$$

where

- $\tilde{\mathbf{Y}}_{ij} = (\tilde{Y}_{ij}^{(1)}, \cdots, \tilde{Y}_{ij}^{(H)})'$ is the vector of response variables
- $\boldsymbol{\mu} = (\mu^{(1)}, \cdots, \mu^{(H)})'$ is the vector of the means
- $\boldsymbol{\delta}_j = (\delta_j^{(1)}, \cdots, \delta_j^{(H)})'$ are the specific errors at cluster level
- $\mathbf{u}_j = (u_{1j}, \cdots, u_{M_u j})'$ are the common factors at cluster level
- $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij}^{(1)}, \cdots, \varepsilon_{ij}^{(H)})'$ are the specific errors at subject level
- $\mathbf{v}_{ij} = (v_{1ij}, \cdots, v_{M_v ij})'$ are the common factors at subject level
- $\boldsymbol{\Lambda}_u$ is the matrix of factor loadings at cluster level with h-th row $(\lambda_{u,1}^{(h)}, \cdots, \lambda_{u,M_u}^{(h)})$
- $\boldsymbol{\Lambda}_v$ is the matrix of factor loadings at subject level with h-th row $(\lambda_{v,1}^{(h)}, \cdots, \lambda_{v,M_v}^{(h)})$.

The standard assumptions on the item specific errors of model (3) are:

$$\boldsymbol{\delta}_j \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}_\delta), \qquad \text{where } \boldsymbol{\Psi}_\delta = diag\{(\psi_\delta^{(h)})^2\},$$

$$\boldsymbol{\varepsilon}_{ij} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi}_\varepsilon), \qquad \text{where } \boldsymbol{\Psi}_\varepsilon = diag\{(\psi_\varepsilon^{(h)})^2\},$$

while for the factors it is assumed that

$$\mathbf{u}_j \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_u), \qquad \mathbf{v}_{ij} \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_v),$$

where the covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_v$ are, in principle, unconstrained, but in the following we assume they are diagonal. Moreover, all the errors and factors are assumed mutually independent, so model (3) is equivalent to the following variance decomposition

$$Var(\tilde{\mathbf{Y}}_{ij}) = \left[\boldsymbol{\Lambda}_u \boldsymbol{\Sigma}_u \boldsymbol{\Lambda}_u' + \boldsymbol{\Psi}_\delta \right] + \left[\boldsymbol{\Lambda}_v \boldsymbol{\Sigma}_v \boldsymbol{\Lambda}_v' + \boldsymbol{\Psi}_\varepsilon \right]. \tag{4}$$

This amounts to a couple of factor models, one for the between covariance matrix and the other for the within covariance matrix (Muthén, 1994).

The outlined factor model (3) raises several identification issues, related to the two components: *(i)* the threshold model which relates the continuous latent variables $\tilde{Y}_{ij}^{(h)}$ to their observed ordinal counterparts $Y_{ij}^{(h)}$; and *(ii)* the two-level factor model for the continuous latent variables $\tilde{Y}_{ij}^{(h)}$.

The total relative communality for the h-th item can be computed as

$$\frac{\sum_{m=1}^{M_v}\left(\lambda_{v,m}^{(h)}\right)^2\sigma_{v,m}^2+\sum_{m=1}^{M_u}\left(\lambda_{u,m}^{(h)}\right)^2\sigma_{u,m}^2}{Var_T\left(\tilde{Y}_{ij}^{(h)}\right)}. \tag{5}$$

The proportion of total variance (relative communality) of the h-th item explained by the k-th subject-level factor, can be computed as

$$\frac{\left(\lambda_{v,k}^{(h)}\right)^2\sigma_{v,k}^2}{Var_T\left(\tilde{Y}_{ij}^{(h)}\right)}, \tag{6}$$

while the proportion of total variance of the h-th item explained by a cluster-level factor can be computed similarly. Note that, since the covariance between the h-th latent variable and k-th subject-level factor is $\lambda_{v,k}^{(h)}\sigma_{v,k}^2$, the corresponding correlation equals the square root of the proportion of total variance (6).

Finally, it is to be stressed that, even if all the estimable quantities are expressed in terms of the item-specific subject-level standard deviations $\psi_\varepsilon^{(h)}$ (Grilli & Rampichini, 2006), the interpretable quantities just described are unaffected by the item scale, since they are ratios of parameters within the same item.

3. Results

We used the ordinal multilevel factor model to analyse five items on job satisfaction of employed Florentine graduates.

Altogether, the considered data set includes 2,432 graduates from 36 study programmes, with a highly unbalanced structure: the minimum, median and maximum number of employed graduates by programme are: 3.0, 31.5 and 495, respectively.

Table 1. Univariate distributions of job satisfaction items. 1998 graduates, University of Florence.

Items	*Levels of satisfaction*					Total	
	1	2	3	4	5	%	N
a. Earning	7.8	23.9	38.1	20.5	9.7	100.0	2421
b. Career	11.0	28.2	32.6	18.0	10.2	100.0	2393
c. Consistency	24.5	27.5	24.2	12.5	11.3	100.0	2427
d. Professionalism	26.0	40.3	22.8	7.7	3.2	100.0	2420
e. Interests	21.5	32.7	28.2	10.8	6.8	100.0	2419

The question "*How much are you satisfied with the following aspects of your present job*?" required a response on a five point scale: *1*. absolutely satisfied, *2*. very satisfied, *3*. satisfied, *4*. unsatisfied, *5*. very unsatisfied. The five considered items are: *a*. earning, *b*. career's opportunities, *c*. consistency of job with degree programme curriculum, *d*. acquisition of competences (professionalism), *e*. correspondence with own cultural interests. The distributions of the items are reported in Table 1. Note that the number of responses for each item is different, due to partial non-response.

The main aim of the analysis is to describe and summarise the aspects of satisfaction measured by the five considered items, separately for the graduate and degree programme levels. The two-level factor model for ordinal variables defined in Section 2 is a useful tool to achieve this goal. In our application, the model is fitted by maximum likelihood with adaptive numerical integration, as implemented in the *GLLAMM* procedure of Stata (Rabe-Hesketh *et al*., 2004). Since the model fitting process is very time-consuming, it is useful to follow a step-by-step procedure:

1. *Univariate two-level models*. As a first step, it is advisable to fit a set of univariate ordinal probit variance component models, one for each item, using standard multilevel software. The estimated proportions of between variance ($\mathrm{ICC}^{(h)}$) allow us to evaluate if a two-level analysis is worthwhile, while a comparison of the thresholds among the items should give some hints about the restrictions to be imposed in the multivariate model.
2. *Exploratory non-hierarchical factor analysis*. In order to shade some light upon the covariance structure of the data, it is useful to estimate the matrix of product-moment correlations among the latent variables, i.e. the polychoric correlation matrix of the items, and to use this matrix to perform an exploratory non-hierarchical (i.e. single-level) factor analysis by means of standard software.
3. *Exploratory between and within factor analyses*. More specific suggestions for the two-level model specification can be obtained from separate exploratory factor analyses on the estimated between and within correlation matrices of the latent variables. The results of this two-stage proce-

dure are expected to be similar to that obtained from the full two-level analysis, as in the continuous case (Longford & Muthén 1992).

4. *Confirmatory two-level factor analysis.* The results of the exploratory two-stage factor analysis, as outlined in point 1, are used to specify one or more confirmatory two-level ordinal factor models as defined by equation (2) of Section 2. These models can be fitted with likelihood or Bayesian methods, and compared with reference to appropriate indicators. The exploratory two-stage factor analysis of point 1 provides fine initial values for the chosen estimation procedure, which may allow a substantial gain in computational time.

3.1 Univariate two-level models

The analysis begins by fitting five univariate ordinal *probit* variance component models. The results, obtained with GLLAMM, are reported in Table 2.

The between proportion of variance, expressed by the ICC, is significantly different from zero for all items. Note that the ICC value for the first three items is about 6-7%, which is measurable in a framework with categorical variables and indicates that a non-negligible part of variance can be explained by degree programme factors.

In a factor model for ordinal variables the thresholds can be left free, while fixing the item means and standard deviations. However, when all the items are on the same scale (and thus $C_h=C$ for each h) a more parsimonious specification can be achieved by assuming that the thresholds differ among the items only by a linear transformation, i.e. $(\gamma_c-\mu^{(h)})/\psi_{\varepsilon}^{(h)}$, where $\gamma_1,\ldots,\gamma_{c-1}$ is a set of thresholds common to all the items (Grilli & Rampichini, 2006).

In this application, the linear restriction on the thresholds is supported by the entries of Table 2. In fact, the differences between adjacent thresholds among the items are similar, except for the third one, which has smaller differences. This suggests that the third item has a higher variability, as also confirmed by the variances calculated after item scoring (Table 5).

Table 2. Univariate ordinal *probit* variance component models: estimated ICC and thresholds. 1998 graduates, University of Florence.

Items	*ICC (%)*	*Thresholds*			
		γ_1	γ_2	γ_3	γ_4
a. Earning	6.0	-1.53	-0.55	0.47	1.27
b. Career	7.4	-1.37	-0.38	0.52	1.24
c. Consistency	6.8	-0.69	0.05	0.71	1.21
d. Professionalism	2.2	-0.64	0.44	1.26	1.88
e. Interests	2.4	-0.77	0.14	0.98	1.54

In light of these remarks, the two-level confirmatory factor model of Section 3.4 will include only one set of thresholds, $\gamma_1,\ldots,\gamma_4$ while allowing the item means $\mu^{(h)}$ and item standard errors $\psi_\varepsilon^{(h)}$ to be freely estimated (with the exception of a reference item).

3.2 Exploratory non-hierarchical factor analysis

The second step requires the estimation of the matrix of product-moment correlations among the latent variables, i.e. the polychoric correlation matrix (see Table 3), whose entries are all significant.

We performed an exploratory maximum likelihood factor analysis on this matrix. The results of this analysis (Table 4) suggest the presence of two factors: a *cultural* factor (labelled *Factor 1*), that explains chiefly the *Consistency-Professionalism-Interests* correlations, and a *status* factor (labelled *Factor 2*), explaining mainly the *Earning-Career* correlation.

Given the low proportions of between variance (ICC of Table 2), this structure is expected to be quite similar to the within structure, thought it may be very different from the between structure.

Table 3. Polychoric correlation matrix of the items. 1998 graduates, University of Florence.

Item	*a*	*b*	*c*	*d*	*E*
a. Earning	1.00				
b. Career	0.54	1.00			
c. Consistency	0.11	0.25	1.00		
d. Professionalism	0.28	0.45	0.54	1.00	
e. Interests	0.16	0.33	0.61	0.58	1.00

Table 4. Exploratory factor analysis on the polychoric correlation matrix of the items: varimax rotated factors and communalities. 1998 graduates, University of Florence.

Item	Factor pattern		*Communality*
	Factor 1	Factor 2	
a. Earning	0.08	0.65	0.43
b. Career	0.26	0.80	0.70
c. Consistency	0.77	0.07	0.60
d. Professionalism	0.68	0.34	0.58
e. Interests	0.78	0.16	0.63

3.3 Exploratory between and within factor analyses

The third step of analysis calls for the decomposition of the overall correlation matrix of the latent variables into the between and within components. This task would require the fitting of a two-level multivariate ordinal model with five random effects for each level, which takes too long to be fitted with numerical integration. Therefore, an approximate procedure is adopted, assigning a score to each item category. Various sophisticated scoring systems could be applied (Fielding, 1999), but given the preliminary nature of this step, the simplest scoring system is applied, assigning the rank value to each category. After scoring, the within and between covariance matrices can be estimated by fitting a multivariate two-level model for continuous responses. To this end, the MLwiN software with RIGLS algorithm (Goldstein *et al.*, 1998) is used, yielding restricted maximum likelihood estimates, which are better for the estimation of variance-covariance parameters than unrestricted ones.

The results are shown in Tables 5 and 6. As for Table 5, note the following points:

- it is clear from the last row of Table 5 that the third item (*Consistency*) has the higher variability, as yet noted in the univariate analysis (Table 2);
- the between proportions of variance are in line with ICC of Table 2;
- the between proportions tend to be higher for covariances than for variances.

As for Table 6, note the following points:

- the total correlation matrix, which is obtained from the between and within components, is similar to the polychoric correlation matrix (Table 3), with a moderate attenuation;
- the structures of the between and within correlation matrices are quite different. Particularly, the between correlations are always higher than the within correlations: this means that the factor model, which explains the correlations, is suitable for the between level even more than might be appreciated by simply looking at the total correlation matrix;

Table 5. Two-level multivariate model on item scores: between variance-covariance percentage and total variance of items. 1998 graduates, University of Florence.

Item	*a*	*b*	*c*	*d*	*e*
a. Earning	5.90				
b. Career	12.96	8.63			
c. Consistency	21.02	13.09	7.37		
d. Professionalism	10.55	7.63	6.62	2.30	
e. Interests	9.57	4.75	6.07	2.76	2.36
Total variance	1.15	1.31	1.68	1.04	1.31

Table 6. Two-level multivariate model on item scores: correlation matrices. 1998 graduates, University of Florence.

ITEM	*a*	*b*	*c*	*d*	*e*
Between					
a. Earning	1.00				
b. Career	0.89	1.00			
c. Consistency	0.36	0.40	1.00		
d. Professionalism	0.72	0.69	0.79	1.00	
e. Interests	0.39	0.32	0.81	0.62	1.00
Within					
a. Earning	1.00				
b. Career	0.46	1.00			
c. Consistency	0.10	0.23	1.00		
d. Professionalism	0.23	0.39	0.48	1.00	
e. Interests	0.14	0.31	0.55	0.52	1.00
Total					
a. Earning	1.00				
b. Career	0.49	1.00			
c. Consistency	0.11	0.24	1.00		
d. Professionalism	0.25	0.40	0.49	1.00	
e. Interests	0.15	0.30	0.55	0.53	1.00

- the within correlation matrix is similar to the total correlation matrix, due to the low proportion of between variances and covariances.

The results of the exploratory maximum likelihood factor analyses performed on the within and between correlation matrices of Table 6 are reported in Tables 7 and 8, respectively.

As for the within structure (Table 7), Bartlett's test indicates that two factors are sufficient (p-value=0.5082). The factor patterns are similar to those found in the non-hierarchical analysis (Table 4).

Table 7. Exploratory maximum likelihood factor analysis on the within correlation matrix: varimax rotated factor loads and communalities. 1998 graduates, University of Florence.

Item	*Factor pattern*		*Communality*
	Factor 1	*Factor 2*	
a. Earning	0.07	0.59	0.35
b. Career	0.25	0.75	0.63
c. Consistency	0.72	0.07	0.53
d. Professionalism	0.64	0.32	0.50
e. Interests	0.74	0.16	0.58

Table 8. Exploratory maximum likelihood factor analysis on the between correlation matrix: factor loads and communalities. 1998 graduates, University of Florence.

Item	Factor pattern		Communality
	Factor 1	Factor 2	
a. Earning	0.00	1.00	1.00
b. Career	0.08	0.89	0.80
c. Consistency	0.93	0.36	1.00
d. Professionalism	0.57	0.72	0.84
e. Interests	0.71	0.39	0.66

As for the between structure (Table 8), while one factor is not enough, the estimation with two or more factors encounters a Heywood case. We decided to retain two factors, forcing the specificities to be non-negative. The second factor loads all items, while the first factor presents relevant loadings only for the last three items.

3.4 Confirmatory two-level factor analysis

Finally, in the light of the results of the preliminary analysis of Section 3.3, a two-level confirmatory factor analysis is performed using model (2). The model is fitted with *GLLAMM*, via adaptive numerical integration with five quadrature points. This is a flexible procedure, but as the complexity of the random part of the model increases, the computational time becomes very long. Since we are not particularly interested in decomposing item specificities, in order to reduce the computational effort the between error terms $\delta_j^{(h)}$ are omitted, so the variances of the remaining item-specific errors $\varepsilon_{ij}^{(h)}$ represent total specificities[1].

The within and between structures emerging from the exploratory analyses are not equally reliable: the within part is estimated on a large number of observations and Bartlett's test clearly indicates the presence of two factors, while the between part is estimated on only 36 degree programmes and the estimation is complicated by the presence of an Heywood case.

Therefore, for the within part of the model the two-factor structure suggested by the exploratory within factor analysis (Table 7) is retained, constraining to zero the loadings that were close to zero, that is the loading of *Earning* in the first factor and the loadings of *Consistency* and *Interests* in the second. As for the between structure, since the hints from the exploratory analysis are less clear, two configurations at this level have been tried:

[1] Grilli & Rampichini (2006) discuss the consequences of this choice.

(*i*) a one-factor unconstrained structure (model M1); and (*ii*) a two-factor structure (model M2), with unconstrained loadings in the first factor and two loadings equal to zero in the second factor (*Earning* and *Career*, see Table 8).

Models M1 and M2 are fitted by means of GLLAMM, using maximum likelihood with five-point adaptive quadrature. In both cases, convergence is achieved after a few iterations but the computational times are in terms of several days.

The likelihood ratio test comparing the models M1 and M2 clearly indicates that the second is better (LR statistic=95.6, *df*=3). The preferred model M2 has 27 estimable parameters: 4 item means $\mu^{(h)}$, 4 common thresholds γ_c, 4 specificities $\psi_{\varepsilon}^{(h)}$, 5 factor loadings $\lambda_{v,m}^{(h)}$ and 2 factor variances $\sigma_{v,m}^{(2)}$ at the student level (m=1,2), 6 factor loadings $\lambda_{u,m}^{(h)}$ and 2 factor variances $\sigma_{u,m}^{(2)}$ at the degree program level (m=1,2). The parameter estimates are reported in Table 9.

The interesting feature of the model is the covariance structure at both levels, which does not depend on the item means and thresholds and can be summarized by the communalities (Table 10). These values are obtained as suitable transformations of model parameters: specifically, the factor '*%Communalities*' are computed from formulae such as (6), the '*Total %Communality*' is obtained by summing the row values FW1, FW2, FB1 and FB2 (see equation (5)), while the last column of the Table is the percentage of total communality due to the between level. The following points should be noted:

- for the first three items the between component is greater for the communality (last column of Table 10) than for the total variance (ICC of Table 2);
- the last two items, *Professionalism and Interests,* are poorly explained by the factors at degree programme level;
- the first factor at the degree programme level, FB1, is interpretable as a status factor, while the second one, FB2, is essentially related to *Consistency*.

Table 9. Confirmatory two-level factor analysis: model M2 parameter estimates. 1998 graduates, University of Florence.

Item	*Loadings*				$\psi_{\varepsilon}^{(h)}$	*Mean*
	Within		*Between*			
	$\lambda_{v,1}^{(h)}$	$\lambda_{v,2}^{(h)}$	$\lambda_{u,1}^{(h)}$	$\lambda_{u,2}^{(h)}$		
a. Earning	-	0.80	0.70	-	1.24	0.21
b. Career	$1^{(*)}$	$1^{(*)}$	$1^{(*)}$	-	$1^{(*)}$	$0^{(*)}$
c. Consistency	3.45	-	0.24	$1^{(*)}$	1.55	-0.79
d. Professionalism	2.34	0.36	0.27	0.19	1.16	-1.42
e. Interests	3.09	-	0.01	0.30	1.15	-0.94
Factor variance	0.26	1.66	0.31	0.34		
Thresholds: $\gamma_1 = -2.71$, $\gamma_2 = -0.69$, $\gamma_3 = 1.13$, $\gamma_4 = 2.52$						

The symbol (*) denotes a fixed value.

Table 10. Confirmatory two-level factor analysis: communalities. 1998 graduates, University of Florence.

Item	*% Communality*					*% Between on Total*
	Within		*Between*		*Total*	
	FW1	*FW2*	*FB1*	*FB2*		
a. Earning	-	38.9	5.5	-	44.4	12.4
b. Career	8.2	51.4	9.5	-	69.1	13.7
c. Consistency	53.2	-	0.3	5.8	59.4	10.3
d. Professionalism	47.5	7.2	0.7	0.4	55.8	2.0
e. Interests	65.1	-	0.0	0.8	65.9	1.2

The factor scores of degree programmes are represented in Figure 1, where the labels concern the degree programmes having at least one score greater than 0.5 or less than -0.5. The points on the right side of the diagram indicate a high satisfaction on *Earning* and *Career*, while the points at the top denote a

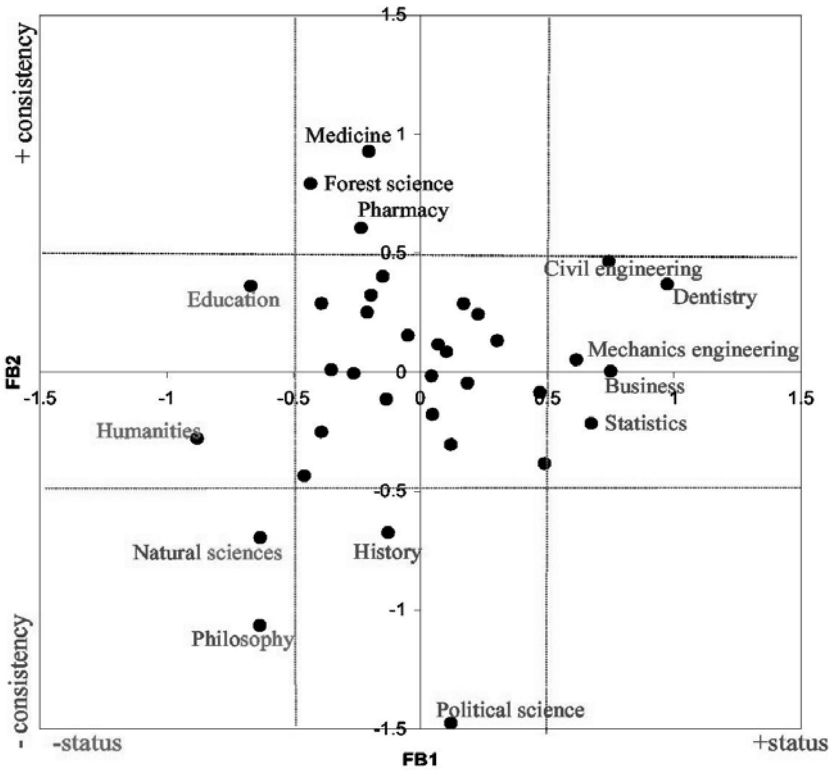


Figure 1. Estimated factor scores for the degree programmes

high satisfaction on *Consistency*. Note that some degree programmes are low only on one dimension (as Humanities on FB1 and Political Science on FB2), while there are two degree programmes lying in the left-down corner (Philosophy and Natural Sciences) with low satisfaction on both dimensions.

4. Concluding remarks

Our analysis showed that there are two relevant factors at both graduate and study programme level. The *status* factors, FW2 and FB1 on Table 9, essentially determine the same variables, *Earning* and *Career*. The other factors, FW1 and FB2 on Table 9, both have a high loading of *Consistency*, but the loadings of *Professionalism* and *Interests* are relevant only at the graduate level. This is in line with the subjective nature of such aspects of the job.

Looking at the estimated factor scores at the degree programme level (Figure 1), extreme cases should be selected for further investigation.

The analysis could be deepened by adding individual-level covariates. This extension is straightforward and does not require a significant additional computational effort.

At present, the major obstacle to a wide use of multilevel factor models for ordinal variables is due to software limitations. The GLLAMM procedure of STATA is very flexible, but it was extremely slow in the present application. Alternative software for fitting such models is M*plus* (Muthén & Muthén, 2003). Currently M*plus* cannot fit exactly the same model used in our analysis, but a trial with a slightly different version of the model shows that the modified EM algorithm implemented in M*plus* is considerably faster, achieving convergence in a few hours.

Anyway, even if fast estimation algorithms are available, it is advisable, especially in the case of ordinal response variables, to fit the multilevel factor model as the final step of the analysis, after having explored the data with simpler techniques.

References

BARTOLOZZI M. (2001) Analisi della soddisfazione sul lavoro tramite un modello ad equazioni strutturali, *Studi e note di economia*, **3**: 163-179.

BINI M. (1999) *Valutazione della Efficacia dell'Istruzione Universitaria rispetto al Mercato del Lavoro,* Rapporto di Ricerca, Osservatorio per la Valutazione del Sistema Universitario - Ministero dell'Università e della Ricerca Scientifica e Tecnologica, Roma.

CHIANDOTTO B., BACCI S., BERTACCINI B. (2004) *Profilo e sbocchi occupazionali dei laureati e diplomati dell'Ateneo fiorentino nell'anno 2000,* University of Florence. http://valmon.ds.unifi.it.

FIELDING A. (1999) Why use arbitrary point scores? Ordered categories in models of educational progress, *Journal of the Royal Statistical Society (A)*, **162**: 303-328.

GOLDSTEIN H., MCDONALD R. P. (1988) A general model for the analysis of multilevel data, *Psychometrika*, **53**: 455-467.

GOLDSTEIN H., RASBASH J., PLEWIS I., DRAPER D., BROWNE W., YANG M., WOODHOUSE G., HEALY M.J.R. (1998) *A User's Guide to MLwiN*, Institute of Education, London.

GRILLI L., RAMPICHINI C. (2006) Multilevel factor models for ordinal variables, *Structural Equation Modeling* (in press).

LONGFORD N., MUTHÉN B. (1992) Factor analysis for clustered observations, *Psychometrika*, **57**: 581-597.

MAZZOLLI B. (2000) *Valutazione dell'efficacia dell'ateneo fiorentino tramite un modello multilivello ad equazioni strutturali*, Ph.d. Thesis, Department of Statistics "G. Parenti", University of Florence.

MUTHÉN B. (1994) Multilevel covariance structure analysis, *Sociological Methods & Research*, **22**: 376-398.

MUTHÉN L.K., MUTHÉN B.O. (2003) M*plus User's Guide* (version 3). Muthén & Muthén, Los Angeles, CA.

RABE-HESKETH S., SKRONDAL A., PICKLES A. (2004) *GLLAMM Manual*. U.C. Berkeley Division of Biostatistics Working Paper Series, Working Paper 160.

SKRONDAL A., RABE-HESKETH S. (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*, Chapman & Hall/ CRC Press, Boca Raton.

SANTORO M., PISATI M. (1996) *Dopo la laurea. Status, sfide, strategie*, Il Mulino, Bologna.

Quality Assessment of the University Educational Process: an Application of the ECSI Model

Bruno Chiandotto, Matilde Bini, Bruno Bertaccini[1]

Statistics Department "G. Parenti", University of Florence, Italy

Summary. In a university, students represent the final users as well as the principal actors of the formative services. A measure of their perceived quality is essential for planning changes that would increase the level of the quality of these services. This perceived quality is analysed in this paper with the ECSI (*European Customer Satisfaction Index*) methodology. The ECSI, which implements a structural equation model, is aimed to represent the satisfaction of the students with some latent variables gauged through a set of observable indicators. We extend the ECSI to the data obtained from graduates of the University of Florence employed one year after graduation.
Keywords: Customer satisfaction; ECSI; Structural equation models; University education.

1. Introduction

We can say that a service is of good quality if it satisfies the consumers' needs (Fornell, 1992; Fornell *et al*, 1996). The development of appropriate activities aimed at improving the quality represents one of the main strategies of the organisations willing to increase their productivity and competitiveness (Montgomery, 1997). Hence, studies on customer satisfaction must be accomplished with the purpose of gathering the appropriate information on purchase experience in order to improve the quality of the service.

Customer satisfaction studies have become increasingly important due to the definition and use of new indices (CSI – *Customer Satisfaction Indices*) and barometers for the evaluation of large markets or that of the entire produc-

[1] The three authors of this paper contributed the initial idea, the structure and the design of the work, whereas M. Bini and B. Bertaccini elaborated and implemented the model.

tion of a country. These new tools derive from two kinds of analyses: *transaction-specific satisfaction* and *cumulative satisfaction* (Johnson *et al.*, 2000).

The initial interest, focused on single episodes of consumption, shifted towards the overall psychological satisfaction after the use of a service. The assessment is based, therefore, on the upgraded experience of multiple transactions between the consumer and productive organization.

The first model was the barometer proposed in Sweden in 1989 (SCSB – *Swedish Customer Satisfaction Barometer*), followed a few years later by the American index (ACSI – *American Customer Satisfaction Index*, 1994), then by the Norwegian barometer (NCSB – *Norwegian Customer Satisfaction Barometer*, 1996) and, finally, by the European Community index (ECSI – *European Customer Satisfaction Index*, see ECSI Technical Committee, 1998).

These models, based on validated theories concerning the consumers' behaviours, their satisfaction and the quality of the products purchased, consist of causal links among latent factors, each one representing the values of a specific set of measurable indicators. Their structure is under continuous review and is subject to modifications in relation to the context. The differences among the proposals are due both to the number of the latent factors and to the number of the causal nexuses involved in the analysis.

Our work is aimed to extend the use of CSI indices to the university education framework. We chose the ECSI model since its basic structure and the relative latent factors are consistent with it. In order to draw a picture of the quality of the educational programmes realised by universities, we analysed data on graduates employed one year after graduation.

For estimating the effects of ECSI we adopted the SEM - *Structural Equation Models* approach based on maximum likelihood estimation[2]. SEM procedures were preferred to PLS ones because the former

- model specification is more flexible,
- allow the significance testing of the omitted parameters (such as error covariances), loading of the latent variables, inclusion of ordinal and categorical variables, implementation of a two-level data structure, and handling of missing data,
- can be implemented by means of specific software (Müthen & Müthen, 2003).

We present the ASCI/ECSI models in Section 2, describe the data set used in our analysis in Section 3, and discuss the results in Section 4. Section 5 is devoted to some final remarks.

---

[2] The technique initially suggested for estimating latent variables in CSI models (Fornell, 1992) was Partial Least Squares-PLS (Wold, 1975). We abandoned this technique because it assumes normality for the estimation of latent variables. Nevertheless, the PLS approach may optimally predict the dependent variable, both with small samples and skewed distributions.

2. Structure of the ACSI / ECSI model

The structure of the CSI models includes some latent factors, each of which explains the variability of a set of, usually observable, indicators. The main feature of this approach is the connection between latent factors deriving from causal relationships between the sources and consequences of satisfaction.

The ECSI model has evolved from the ACSI one. In the European model, the expectations of the consumer, the quality and value perceived, the satisfaction and loyalty concepts are very similar to the American one. The models differ in two fundamental aspects: the ECSI does not include *complaints*, but includes *corporate image* as a latent variable; this latter aspect may exert a direct effect upon the consumer's expectation, satisfaction and loyalty.

The causes of satisfaction foreseen in these models are (for the relationships between factors see Figure 1):

- *Perceived Quality (*QUA*)*: it refers to the assessment of recent consumer experiences concerning the characteristics of the service (*perceived quality of hardware* – QUAHW), the assistance supplied both during, and after the consumer's experience (perceived quality of humanware – QUAUW). Both sub-dimensions are presumed to exert a direct, positive effect on the *overall satisfaction*;
- *Value (*VALU): it represents the value of the quality perceived in relation to price. This factor may positively determine the overall satisfaction and be influenced by the perception of quality;

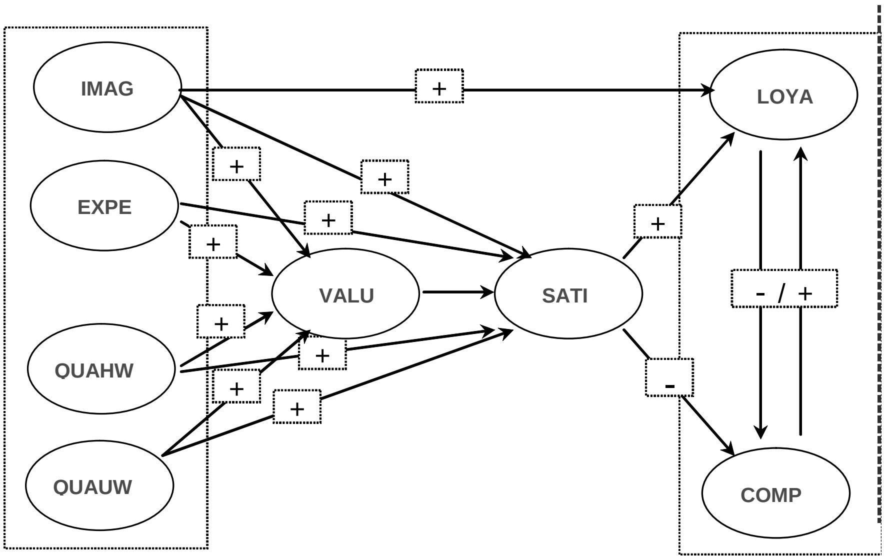


Figure 1. ACSI / ECSI Model: structure and expected relationships.

- *Image* (IMAG): it reflects the sensations generated by the product-brand-manufacturer association. This factor may positively influence value, overall satisfaction and loyalty, and produce a direct effect on perceived quality, nevertheless the ECSI model considers these two aspects exogenous;
- *Expectations* (EXPE): it is the level of quality that the purchaser/user expects to receive and is the result of previous consumption/use experiences. It is considered an exogenous factor capable of positive influence on both value and overall satisfaction.

  The consequences of satisfaction are:
- *Complaints* (COMP): it refers to the type and extent of the complaints and, above all, to the manner in which they are dealt with. In the ACSI model, an increase in the level of overall satisfaction is expected to decrease the incidence of complaints;
- *Loyalty* (LOYA): it is a proxy variable of the power of profit of the organisation since it generates the intention to purchase again, tolerance of variation in price, intention to recommend the service to others. High levels of image and overall satisfaction may activate the *loyalty* of the consumer.

We considered the European version of the customer satisfaction index as the most appropriate for representing the subject matter because adequate informative sources were available (see Section 3).

Based on the information available and the preliminary analyses performed and, above all, on previous knowledge of the phenomenon, we defined the structural section of the reference model for the following analyses (see Figure 2).

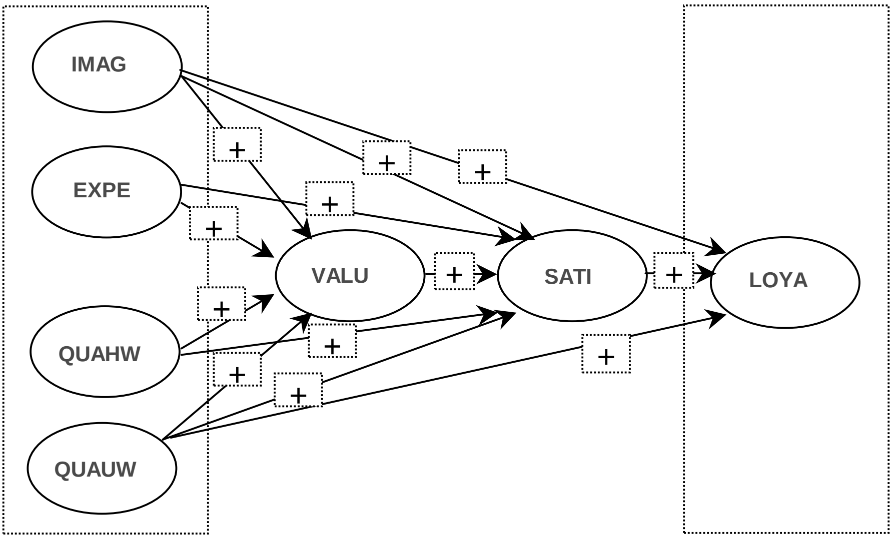


Figure 2. ECSI model expectations for the assessment of the quality of university education.

3. The database

Customer satisfaction surveys are based on questionnaires specifically devised for the purpose and aimed at covering all aspects that may be connected with the causes and consequences of the consumer/user experience.

For assessing the quality of education issued by the University of Florence we created a database containing the data from the ALMALAUREA survey on graduates and that on '*Professional opportunities and Florence University students graduated in the year 2000*' produced by the VALMON GROUP in collaboration with the Statistics Department 'G. Parenti' of Florence[3]. The latter data have been collected with a telephone interview lasting about 20 minutes and using a *CATI - Computer Aided Telephone Interviewing* system[4].

The cognitive requirements of the two basic surveys (Chiandotto *et al.*, 2003, 2004) were different from those entailed in the assessment of customer satisfaction. However, the surveys allowed the definition of a complete picture of the quality of university process by students at the graduation stage as well as some time upon degree achievement.

However, the enticing prospect of being able to combine information surveyed at the time of graduation with that obtained one year after it (which also referred to certain aspects of the graduates' employment) actually implied the exclusion from the analysis of graduates who did not found an occupation within one year after graduation.

The variables capable of assessing certain aspects of satisfaction are:

- the reason for enrolling in university and choosing a particular course,
- the quality of the relationships with professors, assistant teachers, non-teaching staff, fellow students,
- the quality of university facilities (classrooms, laboratories, libraries, canteens),
- satisfaction about occupation (consistency between job and studies, between job and own cultural interests, acquired skills) and overall satisfaction;
- the hypothesis of entering university again and doing the same subjects;
- the intention to continue with educational activities.

Nearly all the variables are dichotomous or five-grade ordinal. Thanks to the number of eligible interviewees, we could delete the records with incomplete information. Thus, the selected cases totalled 1,753.

---

3 ALMALAUREA survey involves graduates of the summer session monitored three (or five) years after the end of their studies. The Florentine survey extended the observation of graduates to the years 1998, 1999, 2000 and 2001 one year after graduation. The Valmon Group, coordinated by B. Chiandotto, is composed of students and professors of the Department of Statistics, Florence University.

4 A mail questionnaire was sent to graduates who were not possible to contact by telephone. The purpose of this survey was to detect possible differences with graduates contacted by telephone.

Table 1. Description of the measurement section of the theoretic ECSI model

Latent structures	***Observed variables***
IMAG:	*Reason for attending University: MOTISCR1, MOTISCR2, MOTISCR3*
EXPE:	*Reason for choosing specific subjects: MOTCOR1, MOTCOR2, MOTCOR3, MOTCOR4, MOTCOR6*
QUAHW:	*Evaluation of classrooms, libraries, laboratories and canteen (scale 1-5): STRAULE, STRBIB, STRLIB, STRMENSE*
QUAUW:	*Evaluation of relationship with professors, assistants, non-teaching staff, fellow students, and overall judgement (scale 1-5): RAPDOC, RAPCOL, RAPNDOC, RAPSTUD, GIUDIZIO*
VALU:	*Satisfaction for coherence with studies and conformity with own cultural interests (scale 1-5): SODDCOER, SODDICUL*
SATI:	*Satisfaction for acquired skills, and overall judgement (scale 1-5): SODDPROF, SODDTO*
LOYA:	*Hypothesis of re-attendance, intention to continue studies and educational activities: IPREISC, OKUNIV, OKCOR, INTSTUD, ATTFORM*

In Table 1, it is presented the measurement section of the ECSI model of the relationships between the directly observed variables and latent structures.

4. The ECSI model for the quality of university education assessment

As suggested by some Authors (see Bollen, 1989), a model for confirmatory factor analysis should be used at a first stage to validate a structural equation model[5], that is, to evaluate the quality of the indicators employed as a gauge of

[5] The structural equation models originate from the convergence of two scientific traditions, the *ergonometric*, which derives from the theory of a network of causal links between variables, and *psychometric*, which comes from the notion of latent variables. The structural equation model is usually referred to as the LISREL model (Jöreskog, 1973; Jöreskog, 1990; Jöreskog & Sörbom, 1993; Corbetta, 2002). This model assumes two components: a *structural* part, designed to explain the causal links between latent variables, and a *measurement* one for evaluating the variability on the observed variables. The model may be expressed as: $\boldsymbol{y} = \boldsymbol{\eta} + \mathbf{B}\boldsymbol{\xi} + \boldsymbol{\zeta}$ (structural) and $x = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$, $y = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}$ (measurement), where $\boldsymbol{y}$ and $\boldsymbol{x}$ are, respectively, endogenous and exogenous vectors of observed variables, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are vectors of underlying latent variables, $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are matrices of coefficients and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are the measurement error terms. In the SEM it is normally assumed that the data follow a multivariate normal distribution, so that the

latent components and eliminate alien sources of variability.

Nevertheless, the use of this procedure would have been inappropriate for our purposes: the lack of a specific questionnaire would necessitate in fact the application of an exploratory factor analysis for unveiling the suitable number of latent structures and the items appropriate for measuring these structures.

Once the measurement section of the model has been defined, the estimation procedure is optimised by applying a confirmatory factor analysis and estimating the effects.

4.1 Exploratory factor analysis

The dual aim of the exploratory factor analysis is to detect the optimal number of latent factors and identify the variables that prove to be appropriate for measuring them. The analysis highlighted between five and eight factors and we chose seven factors.

The factorial loadings are presented in Table 2. The variables uncorrelated with the factors are marked with asterisks. All the other variables are clearly associated with only one factor, with the exception of that regarding consistency between occupation and studies (SODDCOER). On this ground, we identified the measurement section of the model (Table 3).

4.2 Confirmatory factor analysis

A typical configuration of confirmatory factorial analysis (CFA) is a model of structural equations that foresees all the possible links between latent structures. The purpose of CFA is to assess the quality of the measurement structures of the model identified in the previous step.

By applying the WLSMV (*Weighted Least Square Mean and Variance*) estimation logarithm, recommended by the *MPlus* software, the model suggested by the explorative factorial analysis fails to converge. Nevertheless, the resulting estimates are the starting values for the following models.

We performed another confirmatory factor analysis that accounted for the results of the exploratory analysis and the first implementation of the ECSI model (Table 4).

---

vector of means and the matrix of covariance contain the information required for estimation. Nevertheless, there are estimation procedures that use non-normal continuous data but require the entire data matrix. The statistical tests of fitting generally depend on sample size. Likewise, when the sample is limited in size, there is a risk of accepting the model even if the fitting is poor. Several indices have been proposed for measuring the proximity of the data to the model (Hox & Bechger, 1998).

Table 2. Factor loadings of 7-factor exploratory factor analysis

	A	B	C	D	E	F	G
GIUDIZIO	-0,648	0,251	0,028	-0,018	0,101	0,028	0,122
RAPDOC	0,842	-0,176	0,019	0,034	0,027	-0,013	-0,032
RAPCOL	0,829	-0,154	-0,036	0,020	0,018	0,006	-0,012
RAPNDOC	0,652	-0,201	-0,054	-0,036	-0,046	0,020	-0,010
RAPSTUD	0,432	-0,139	-0,015	-0,022	-0,076	0,017	-0,053
STRAULE	0,201	-0,749	-0,042	-0,036	-0,014	0,057	0,005
STRLAB	0,286	-0,665	-0,066	-0,041	0,003	0,104	-0,010
STRBIB	0,236	-0,513	0,042	0,009	-0,051	-0,030	-0,038
SERMENSE	0,189	-0,329	0,025	-0,102	-0,045	0,006	0,021
INT_STUD	0,038	0,008	-0,031	0,053	-0,040	-0,528	-0,033
ATTFORM	-0,003	0,019	-0,118	0,014	0,070	-0,563	-0,011
EFFESTER	-0,022	0,003	0,828	0,033	0,201	0,124	0,269
MOTISCR1	-0,010	-0,006	0,005	-0,587	0,218	0,252	0,002
MOTISCR2	0,065	-0,009	-0,014	0,937	-0,022	-0,028	-0,001
MOTISCR3	-0,068	0,013	0,017	-0,773	-0,104	-0,099	0,005
* MOTCOR1	0,056	-0,074	-0,014	-0,133	0,179	0,274	0,034
MOTCOR2	-0,045	0,088	0,043	0,469	-0,184	-0,140	-0,005
* MOTCOR3	0,077	-0,005	0,045	-0,220	0,014	0,073	-0,129
MOTCOR4	-0,012	-0,067	-0,039	-0,433	-0,004	-0,072	0,063
* MOTCOR6	0,057	-0,031	-0,050	-0,308	0,288	0,145	-0,114
SODDCOER	0,056	-0,025	-0,531	-0,054	-0,181	-0,151	-0,583
SODDICUL	0,039	-0,022	-0,184	-0,033	-0,128	-0,178	-0,728
SODDPROF	0,042	-0,013	-0,102	-0,039	-0,087	-0,090	-0,791
SODDTOT	0,070	0,017	-0,081	0,031	-0,106	0,090	-0,716
COMPETEN	-0,066	0,010	0,932	-0,006	0,183	0,051	0,244
OKUNIV	0,050	-0,103	-0,142	0,056	-0,840	-0,069	-0,230
OKCOR	0,122	-0,050	-0,190	0,067	-0,881	0,022	-0,179
* CERNWLAV	-0,039	0,008	0,185	-0,018	0,234	-0,246	0,393
IPREISC	0,114	-0,052	-0,154	0,110	-0,739	0,005	-0,193

Table 3. Measurement section suggested by the exploratory factor analysis

A	Quality of relationships with professors, assistants, non-teaching staff, fellow students
B	Evaluation of classrooms, libraries, laboratories, canteens
C	Level of usage of skills acquired at university
D	Reason for attending university and choosing particular subjects
E	Opinion concerning the hypothesis of attending university again and studying the same subjects
F	Intention to continue studies and educational activities
G	Consistency between occupation and studies, consistency between occupation and own cultural interests, acquired skills, overall job satisfaction

Table 4. Measurement of confirmatory factor analysis according to the exploratory one

IMAG:	MOTISCR1 MOTISCR2*-0.68 MOTISCR3*0.264
EXPE:	MOTCOR2 MOTCOR4 MOTCOR6
QUAHW:	STRAULE STRBIB*1.06 STRLAB*0.817 SERMENSE*0.568
QUAUW:	RAPDOC RAPCOL*0.98 RAPNDOC*0.815 RAPSTUD*0.554
VALU:	SODDCOER SODDICUL*0.884 SODDPROF*0.876 SODDTOT*.755
SATI:	OKUNIV OKCOR*1.006 IPREISC*.913
LOYA:	INT_STUD ATTFORM*1.163

The adaptation indices denote a model that is capable of describing the subject matter in a more satisfactory manner (TFI = 0.984 e RMSEA = 0.028). This makes it possible to rely on the measurement section detected and concentrate solely on the structural section, estimating the causal links between the latent components to verify the strength of the expected relationships (Figure 2).

4.3 Structural equation models

The CFA model was re-configured into a *complete* ECSI-SEM model, where the regression equations between the latent components are specified. The term ‘complete’ describes the particular configuration assumed by the causal links between the latent components, because of its typical temporal and sequential features. That is to say, in the complete model each factor on the left side of the graph is assumed capable of exerting both a direct and an indirect influence on factors on its right. For instance, all the theoretic causes of consumption experience (IMAG, EXPE, QUAHW, QUAUW) are assumed as explanatory factors of both the actual components of the satisfaction (VALU e SATI) and the loyalty (LOYA).

Moreover, since it is plausible to assume that the consumption experience can activate a cognitive process whereby a value is attributed to a service purchased before the arousal of actual feelings of satisfaction, it is logical to consider VALU as a possible cause of SATI and LOYA.

The implementation of a complete model required new links within the model originally proposed (see Figure 3 – added links in bold print). Due to the extreme complexity of the structural section, the complete ECSI-SEM model did not converge. The estimates produced by the software used have nonetheless been useful guidelines for improving the model since they suggest which causal links had to be eliminated.

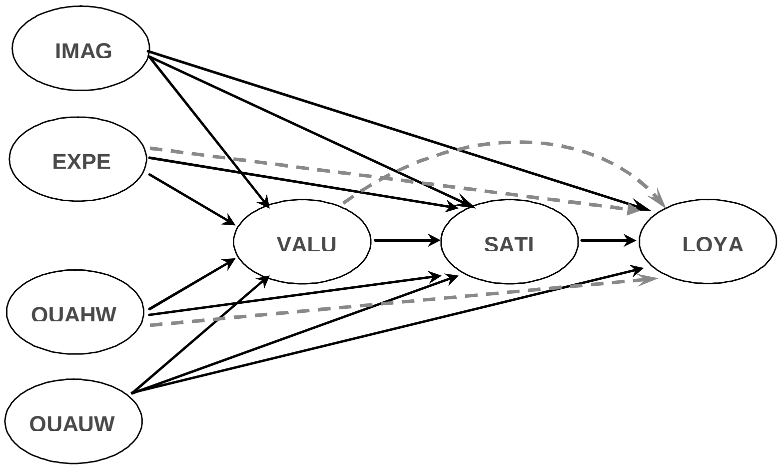


Figure 3. Complete ECSI-SEM model: structural section highlighting the added links

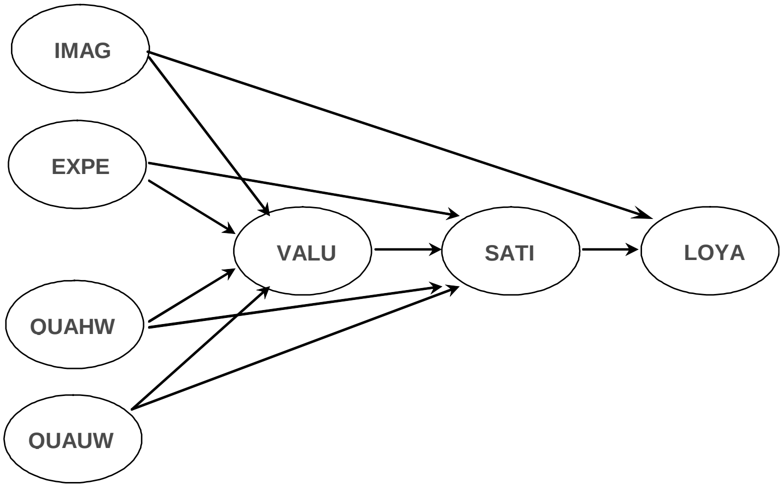


Figure 4. Structure of the first ECSI-SEM convergence model

Simplifications applied to the structural section for subsequent steps have generated the model shown in Figure 4. Even in this case, the normal adaptation indices indicate that the model effectively fits the data (TFI = 0.982; RMSEA = 0.029).

Being aware that this procedure is tricky and open to criticism, but wishing to define a more satisfactory model from an interpretative point of view, the non-significant links have been gradually eliminated.

Just the essential results obtained during this step will be discussed here. It is sufficient to notice that the adaptation indices always emphasise sensible modifications tending towards the value that represents perfect adaptation.

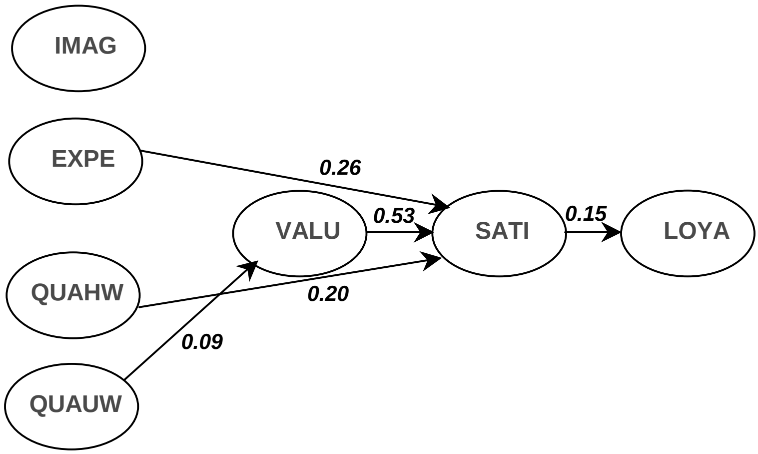


Figure 5. Structure of the final ECSI-SEM mode**.**

Figure 5 shows the final ECSI-SEM model for which no improvement was appreciable. The values of the estimated coefficients are shown on the arrows in the graph that represent the causal links differing from zero: note that all the effects are oriented in the expected direction. This graph does not show the correlation relationships that nevertheless existed between the latent components. Once again, the good adaptation of the model is substantiated by the values given by the TLI (0.985) and RMSEA (0.027) indices.

5. Conclusions

The interest on quality assessment of educational activities in Italian universities is based on law (Laws 168/89 and 537/93). It foresees internal controls on the efficiency of processes and effectiveness of results of university management. Especially in recent years, numerous projects for the assessment of university education in terms of efficiency and effectiveness have been proposed and realised.

The success achieved by university graduates in the labour market may be considered the mirror image of the quality of educational processes. This success may be measured in terms of either objective indicators, like salary and career, or subjective ones, such as satisfaction for the occupation.

Our work evaluated the ECSI-SEM models as a tool for the analysis of university education quality. The results of our analysis not only confirm the validity of the application of ECSI-SEM models in this subject matter but also stimulate interest towards the implementation of more detailed analyses, which, with the collection of appropriate data, could help the development of theoretical and methodological aspects.

A suitable questionnaire could be studied to collect opinions on satisfaction. As far as the methodological aspects are concerned, an evaluation of the effects caused by the presence of abnormal observations could be performed by applying the forward search algorithm and implementing new models for groups of graduates who attended the same programme, in order to test the existence of specific causal links of satisfaction.

References

BOLLEN K.A. (1989) *Structural Equation with Latent Variables*, Wiley, New York.

CHIANDOTTO B., BERTACCINI B. (2003) *Profilo e sbocchi occupazionali dei laureati e diplomati dell'Ateneo fiorentino nell'anno 1999*, Gruppo VALMON - Università degli Studi di Firenze.

CHIANDOTTO B., BACCI S., BERTACCINI B. (2004) *Profilo e sbocchi occupazionali dei laureati e diplomati dell'Ateneo fiorentino nell'anno 2000*, Gruppo VALMON - Università degli Studi di Firenze.

CORBETTA P. (2002) *Metodi di analisi multivariata per le scienze sociali. I Modelli di Equazioni Strutturali*, Il Mulino, Bologna.

ECSI Technical Committee (1998) *European Customer Satisfaction Index: Foundation and Structure for Harmonized National Pilot Projects*. Report prepared for the ECSI Steering Committee, October.

FORNELL C. (1992) A National Customer Satisfaction Barometer, the Swedish Experience, *Journal of Marketing*, **56**: 6-21.

FORNELL C., JOHNSON M.D., ANDERSON E.W., CHA J., BRYANT B.E. (1996) The American Customer Satisfaction Index, Nature, Purpose and Findings, *Journal of Marketing*, **60**: 7-18.

HOX J.J., BECHGER T.M. (1998) An Introduction to Structural Equation Modeling, *Family Science Review*, **11**: 354-373.

JOHNSON M.D., GUSTAFSSON A., ANDREASSEN T.W., LERVIK L., CHA J. (2001) The Evolution and Future of National Customer Satisfaction Index Models, *Journal of Economic Psychology*, **22**: 217-245.

JÖRESKOG K.G. (1973) A General Method for Estimating a Linear Structural Equation System. In: GOLDBERGER A.S., DUNCAN O.D. (eds) *Structural Equation Models in the Social Sciences*, New York, Academic Press: 85-112.

JÖRESKOG K.G. (1990) New Developments in LISREL. Analysis of Ordinal Variables Using Polychoric Correlations and Weighted Least Squares, *Quality and Quantity*, **24**: 387-404.

JÖRESKOG K.G., SÖRBOM D. (1993) *New Features in PRELIS2*, Scientific Software International Chicago, IL.

MONTGOMERY D. (1997) *Introduction to Statistical Quality Control*, Wiley, New York.

MÜTHEN B.O., MÜTHEN L.K. (2003) *MPlus 3.0 User Guide* (http://www.statmodel.com/verhistory.shtml).

WOLD H. (1975) Path Models with Latent Variables, the NIPALS Approach. In: BLALOCK H.M., AGANBEGIAN A., BORODKIN F.M., BOUDON R., CAPPECCHI V. (eds) *Quantitative Sociology. International Perspectives on Mathematical and Statistical Modeling*, Academic Press, New York: 307-353.

Evaluating the University Educational Process. A Robust Approach to the Drop-out Problem

Matilde Bini, Bruno Bertaccini

Statistics Department "G. Parenti, University of Florence, Italy

Summary. The use of robust procedures in regression model estimation identifies outlier data that can inform on specific subpopulations. The aim of this study is to analyse the problem of first year dropouts at the University of Florence. A set of administrative data, collected at the moment of enrolment, combined with the information gathered through a specific survey of the students enrolled in the 2001-2002 academic year at the same athenaeum, was used for the purpose. In order to identify the most important variables affecting the students' dropout, the data were first fitted with generalized linear models estimated with classical methods. The same models were then estimated with robust methods that allowed the detection of groups of outliers. These in turn were analysed to determine the personal or contextual characteristics. These results may be relevant for the implementation of academic policy changes.
Keywords: Dropout rate; Outliers; Forward search method.

1. Introduction

The evaluation of the higher education system – and particularly of the university one – and the use of statistical methods for measuring its performance have become important issues given the cultural, social and economic relevance of this educational level. The statistical analyses carried out on the Italian university system highlight, among the other aspects, its weaknesses (Bini, 1999; Bertaccini, 2000; Chiandotto & Bertaccini, 2003).

We argue that the average levels of single indicators do not emphasize the 'net' effects of factors that quantify the indicators, because of the existing interactions among these and grouping variables. Hence, it is necessary to implement appropriate analytical models for representing the relationships with the phenomenon considered.

The linear regression model satisfies this requirement. The applications described in the following are the basis for a complete and complex analysis performed using the *Forward Search* algorithm (Atkinson & Riani, 2000). This method may be used to make correct decisions both for the allocation of resources and verify the planned objectives of the educational programmes. In fact, this method is able also to reveal the presence of unusual characteristics of the phenomenon under study.

The procedure starts by fitting the model with a number of observations sufficient for estimating parameters and continues the fitting of the model to increasing subsets. The units are ordered according to their proximity to the fitted model. If the model agrees with the data, the robust and least squared procedures yield similar parameters and error estimates. However, these may change considerably with the *Forward Search*. The monitoring of the changes and of some statistics used to make inference in regression models allows reaping information useful not only for detecting outliers, but also, and above all, to comprehend their importance in inference making.

We present an application of detection of observations with characteristics that can explain the low degree of withdrawal from university programmes where the drop-out problem is particularly significant.

In Section 2 we introduce the estimation problems caused by the presence of outliers, in Section 3 we show the properties of the least median squares as the robust approach to the regression model; Section 4 is devoted to the presentation of the forward algorithm applied to classic linear models and to generalized linear models. Concluding remarks are presented in the Section 5.

2. The problem of outliers

In various fields of research, the regression model is a common statistical tool. The properties of the Ordinary Least Squares (OLS) estimators justify its popularity but not the mistreatment that occasionally occurs with their use, when insufficient attention is given to both verification of the specification theories and the presence of anomalies in the data at hand.

It is to be remembered that the estimate of the p parameters in a regression model depends on p statistics computed on the whole dataset; if any of these differ from the bulk of the data, the fitting process can conceal these differences or, otherwise, be strongly influenced by them.

The outliers can derive from mistakes performed during the recording steps, or from unusual phenomena, or can identify units accidentally included in the sample but belonging to other populations.

The response variable is not the only factor that can undergo irregularity. Outliers may be related with explanatory variables because of the larger frequency with which atypical data may be collected.

If the regression parameters were known, there would be no difficulty in detecting the outliers as units enhancing the highest terms of error. However, difficulty arises when the parameters of the model have to be estimated with observations that may contain abnormal units.

To sidestep their presence, *robust* methods of estimation (so called because they can produce estimates that are not influenced by contaminated data) have been proposed.

These methods identify as *outliers* the units that reveal very high residuals[1]. Methods for investigating several outliers simultaneously entail the use of robust techniques for organising observations according to their residuals. One that deserves special mention is the *Forward Search* algorithm (Section 3).

3. The *Forward Search* for classical linear models

The ordinary least squares method yields estimates of β coefficients that minimise the residual sum of squares. The distance $(y_i - \hat{y}_i)^2$ gives higher weights to units with higher residuals. Therefore, if there are few observations with very large residuals, ε_i, the estimates are strongly affected, particularly those concerned with high leverage points. Because of this sensitivity, estimation methods that yield good estimates in the presence of contaminated data are termed "*robust*" (Box & Andersen, 1955).

Donoho & Huber (1983) introduced the *breakdown point* (Rousseuw, 1987; Atkinson & Riani, 2000) that is the smallest fraction of contamination that can induce a certain estimator $T(Z')$ to assume values arbitrarily distant from $T(Z)$, where $Z = (X, y)$ stands for the matrix $n \times (p + 1)$ of the original data and Z' for all the possible contaminated samples.

Contaminated are the samples obtained by substituting any m number of the original observations with arbitrary values. No reference is made in this definition to the probability distribution of the data.

Among the various robust approaches, special mention must be made to the studies by Huber and to the *Least Median of Squares* (LMS) estimation method (Rousseuw, 1984). This latter estimator has a 50% breakdown point, i.e. at least half of the observations must be outliers in order to have repercussions on the estimates. Even if the contamination rate is lower, the method

---

[1] An alternative approach is the *diagnostic analysis*, which implies the computation of statistics that detect abnormalities and the data most liable to exert an influence (Cook & Weisberg, 1982; Atkinson, 1985). The main drawback of these procedures is that, as the number of potential outliers increases, there is a combined explosion of the possible subsets under investigation. In robust regression, the model is adapted with techniques that do justice to the bulk of the data and examine the units that mostly differ from the predicted values. Nevertheless, robust approaches and diagnostic analyses often produce the same results.

gives unbiased estimates of the regression hyperplane, provided n is large. This is the maximum *breakdown point* a regression model can tolerate[2].

The *Forward Search* algorithm (Atkinson & Riani, 2000) combines the diagnostic ability in identifying groups of *outliers* with the properties of the robust estimation methods, particularly the LMS. The basic steps of this algorithm are (Section 3.1): choice of the best starting-set free of outliers, the addition of observations, and the monitoring of the statistics that detect outliers.

3.1 The choice of the initial subset

The best initial subset is detected with the least median of squares approximation (Rousseuw, 1984), which guarantees a starting-set free of outliers.

The basic steps of the algorithm are the following. Let $Z = (X, y)$ be a $n \times (p+1)$ matrix.

If n is not too large and $p<<n$, the choice of the initial subset derives from the enumeration of the $\binom{n}{p}$ distinct p-tuples $S^{(p)}_{i_1,\dots,i_p} \equiv \{z_{i_1},\dots,z_{i_p}\}$, where $z^T_{i_j}$ is the i_j-th row of Z $(j = 1, \dots, p)$ and $1 \le i_j \ne i_{j^*} \le n$.

Let $i^T = [i_1, \dots, i_p]$ and $e_{i,S^{(p)}_\iota}$ be the least squared residual for the i-th unit since the regression model is fitted only with the observations in $S^{(p)}_\iota$. Hence, the starting-set is the set $S^{(p)}_*$ of p-tuples that satisfies

$$e^2_{[med],S^{(p)}_*} = \min_\iota \left[e^2_{[med],S^{(p)}_\iota} \right], \tag{1}$$

where $e^2_{[k],S^{(p)}_\iota}$ is the k-th ordered square residual among $e^2_{i,S^{(p)}_\iota}$, $(i = 1, \dots, n)$ and *med* is the integer part of $(n + p + 1)/2$.

If $\binom{n}{p}$ is large, the subset is chosen by using (1) and enumerating 3,000 p-tuples of matrix Z.

Therefore, the *Forward Search*, according to the size of the data set and to the capacity of computers, finds the minimum of (1) among the 3,000 p-tuples

---

[2] The very robust behaviour of the LMS estimator is in contrast with the Ordinary Least Squares (OLS). In OLS, a sensitive variation in the estimations can take place when just one outlier assumes arbitrary large values; since the *1/n* fraction tends towards zero as the sample size increases, the breakdown point will be zero.

from a sample of size n. If the number of p-tuples is less than 3,000, the *Forward Search* will take account of all the subsets.

3.2 The addition of observations

Let $S_*^{(m)}$ be a subset of size $m \geq p$; the *Forward Search* procedure moves towards the subset $S_*^{(m+1)}$ by selecting units that have the first $m+1$ ordered residuals $e^2_{[k],S(m)}$. The procedure ends when all the observations are included in the subset; that is, when $S_*^{(m)} = S^{(n)}$. Therefore, the *Forward Search* estimator $\hat{\beta}_{FS}$ is the set of the $n-p+1$ ordinary least squared estimators obtained at each step of the procedure; that is

$$\hat{\beta}_{FS} = \left(\hat{\beta}^*_p, \hat{\beta}^*_{p+1}, \ldots, \hat{\beta}^*_{n-1}, \hat{\beta}^*_n \equiv \hat{\beta}_n\right).$$

The change of the dimension from m to m+1 implies in general the entrance of only one unit into the previous subset. Nevertheless, two or more units could enter $S_*^{(m+1)}$ while one or more could leave.

This last situation may occur, though less frequently, when an observation that belongs to a group of outliers emerges during a study. In fact, in the subsequent step, the remaining units of the group of outliers reveal a less typical behaviour and a few can enter the subset at the same step.

This new approach combines the features of the least median squares with the efficiency of the OLS estimators. The robustness of the approach consists in the continuous inclusion of units in the subset free of outliers, rather than in the choice of a particular estimator having a high breakdown point. In other words, this method is not particularly affected by the technique used to select the initial subset of units, since it generates a subset free of outliers, or with masked outliers that may be removed at following steps of the search.

3.3 The monitoring of the procedure

The estimate of σ^2 changes during the procedure since, at each step of the search procedure, m units with the smallest residuals enter the subset ($m = p+1, \ldots, n$). Therefore, even without outliers, we have $s^2_{S(m)} \leq s^2_{S(n)}$ given $m < n$. Usually, the curve drawn by $s^2_{S(m)}$ shows an initial slight increase, as it happens when the data and the adopted model agree, and a steep increase if significant outliers are present.

A very important graphical representation is the one that allows checking the behaviour of all *n* residuals at each step of the *Forward Search*. Large values of residuals from units not included in the data point out the presence of outliers.

Since $s^2_{S_*^{(m)}}$ is strongly dependent on *m*, all residuals are standardized with the mean of the squared residuals s^2. The plot that shows the trajectories of leverage values is useful for detecting outliers. At any step, a unit enters the subset $S_*^{(m)}$, and corresponding leverage $h_{i,S_*^{(m)}} = x_i^T \left(X^T_{S_*^{(m)}} X_{S_*^{(m)}} \right)^{-1} x_i$ values are plotted.

At start, the search of the subset $S_*^{(p)}$ includes only *p* observations, each one with a leverage value equals to one. Thereafter, these values decrease. The outliers that enter the $S_*^{(m)}$ subset at the final steps of the analysis may show higher leverage values than the others, though it is possible that units that make up the initial set show the highest values for the whole analysis.

The *Forward Search* algorithm can be extended to Generalized Linear Models (GLMs) that are an extension of linear regression models where the response variable is not normally distributed, so that the expected values are modelled with a link function (Agresti, 2002). In particular, if models for binomial data are adopted, as in this work, the search algorithm is the same as the linear regression one, except that squared residuals of deviance d_i^2 are used rather than least square residuals. In this case, the procedure starts with a random selection of subsets in *p* dimension and chooses the subset with the smallest value of the median of the deviance[3].

The afore-mentioned situations highlight that *Forward Search* algorithm can be effective to the sample by combining its diagnostic capability in identifying groups of outliers together with the properties expressed by robust methods of estimation. The results obtained when the algorithm is applied are twofold: anomalous units can be eliminated to produce a model that is more stable and conform to reality, and, on the other hand, the information resulting from the analysis of the structural composition of the identified outliers can be used to perform further investigation into the phenomenon under study.

In this respect, the application was successful in evaluating both the effectiveness of university titles in relation to employment and the discontinuation

[3] This is the general rule except in the case of models for binary data. In fact, except for the response variable, the number of zeros is not equal to number of ones, the Least Median of Squares method of estimation will include in the initial subset only observations with frequent responses. Hence, in order to keep both types of response in equilibrium during the procedure, the search algorithm must be modified.

of university studies, and proper for supplying indications that may be helpful for improving the quality of the educational process.

4. An analysis of the dropout rate in Florence

There has been a significant evolution in the Italian public administration over the recent years. In particular, the new regulations acknowledge the Ministry of Education, University and Research as the responsible entity for defining the targets and general strategies for developing the tertiary educational system and its assessment, while the universities are granted ample self-governance, even though part of the funds are subject to the fulfilment of specific requisites[4].

The decentralization and self-government, and the restriction on funding, imply that the universities, as responsible for the results achieved, must necessarily perform procedures of intense and exhaustive self-assessment in terms of both efficiency and internal and *external effectiveness*.

One of the most important indicators of internal efficiency is the dropout rate, which continues to override 50% in many universities, even after the recent reform of the study cycles, with serious repercussions on programming university policies and even on the society as a whole.

Florence University data show that during the 2002/03 academic year almost 30% of the enrolled students dropped out of the course they chose in the previous year. Because of the severe consequences this phenomenon has on educational programming, Florence University carried out a *CATI (Computer Aided Telephone Interviewing)* survey on the possible causes of dropout[5] of the students enrolled in the 2002/2003 academic year and who resulted to be in one of the following conditions:

- *Passed* (P) to another degree course in Florence University;
- *Transferred* (T) to another university;
- *Renounced* (R) explicitly to the studies;
- *Implicit* (I) if he/she is not enrolled in 2002/03 and does not belong to any of the previous categories;
- *Suspended* (S), e.g., for national military service; these students could also be included in the Implicit category.

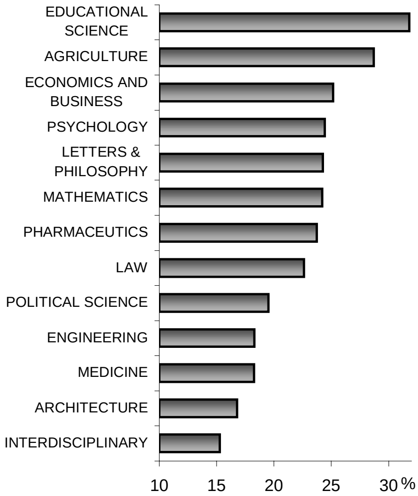
The complete database includes a student's profile (gender, age, residence), high school curriculum (type and final marks), faculty and study programme upon first registration, employment status and position regarding compulsory

4 Financial, managerial and organisational self-government of every Italian university was introduced under Laws no. 168/89, 537/1993, 59/1997 and 127/1997.

5 The survey was realised thanks to funds of *CAMPUS*ONE and *OUTCOMES* projects. The data collected permitted the correction (due to errors or delayed input by the Student secretariat) of some data in the university files.

Table 1. Distribution of the dropout rate of the sample by faculty

Degree programme	*Response Y*	
	Enrolled	**% (Y=1)**
AGRICULTURE	*358*	28,7
ARCHITECTURE	*1,065*	16,8
ECONOMICS AND BUSINESS	*1,029*	25,2
PHARMACEUTICS	*197*	23,8
LAW	*777*	22,6
ENGINEERING	*889*	18,3
LETTERS & PHILOSOPHY	*1,518*	24,3
MEDICINE	*717*	18,3
PSYCHOLOGY	*1,429*	24,4
EDUCATIONAL SCIENCE	*462*	31,8
MATHEMATICS	*649*	24,2
POLITICAL SCIENCE	*450*	19,5
INTERDISCIPLINARY	*513*	15,3
TOTAL	***10,053***	22,3



military service at the time of enrolment, as well as the correct form of registration and drop-out motives of transfer to another programme within Florence University or to another university.

The analysis was performed from the following perspectives: dropouts can be considered a phenomenon depending exclusively on the organisation of the study programme or on the central organisational policies of the university. The transfer from one course to another may be an important rupture in a student's career since it involves a waste of time and resources both for the student and, indirectly, for the study programme. However, the number of students who change programme within the same university is attributable to the sudden awareness of the incompatibility between their personal aptitudes and the subject matter of the programme.

The response variable is:

- *Y=1*, if the student has dropped out of the initial enrolment programme and did not pass to any other programme at Florence University the year after (the options in the questionnaire were *transferred*, *renounced*, *implicit* or *suspended*);
- *Y=0*, if the student is still registered in the original programme, or has transferred or submitted a request to transfer to another study programme within Florence University.

Preliminary analyses indicated the association of some covariates with the response variable[6] (Table 1). Only the faculty and the study programme re-

6 The contingent of students reduced to 9007 because of the omissions in the administration files corresponding with the variables related to the phenomenon under study; nevertheless, this did not alter the distribution of drop-out rates shown in Table 1.

Table 2. List of the explanatory variables correlated with response variable

Factor	Description	Levels	
DEGREE	Degree programme selected on enrolment	1 = AGRICULTURE	
		2 = ARCHITECTURE	
		3 = ECONOMICS AND BUSINESS	
		6 = PHARMACEUTICS	
		7 = LAW	
		8 = ENGINEERING	
		9 = LETTERS & PHILOSOPHY	
		10 = MEDICINE	
		11 = EDUCATIONAL SCIENCE	
		12 = POLITICAL SCIENCE	
		13 = PSYCHOLOGY	
		14 = MATHEMATICS	
		15 = INTERDISCIPLINARY	
COURSE	Course programme	104 levels	
GENDER	Gender	1 = Male;	2 = Female
COUNTY	Area of residence	1 = Florence - Hinterland	
		2 = Other towns in provinces of Florence & Prato	
		3 = Other provinces in Tuscany	
		4 = Other northern and central regions	
		5 = Southern regions and Islands	
HSCHOOL	Kind of high school	1 = Classics;	2 = Scientific
		3 = Technical;	4 = Others
HSSCORE	Final marks	1 = 60 – 56	2 = 55 - 51
		3 = 50 – 46	4 = 45 - 41
		5 = 40 – 36	
AGEENROLL	Age at enrolment	1 = Under 20	2 = 20
		3 = 21 – 25	4 = over 25
WORK	Occup. status at enrolment	1 = Employed	2 = Unemployed

sulted to be related to the quality of the organisation of the education, whereas the others related to the profile of the students (Table 2).

All the variables correlated with the phenomenon under study were considered as possible predictors[7] in a logistic model for binary data, automated with a *backward elimination* option. The model is merely exploratory, since it is used to detect the subsets of explicative variables with the highest powers of discrimination. In Table 3, we illustrate the main results.

For the best possible assessment of the impact of education organisation as a possible cause of drop-out, the *Forward Search* algorithm was applied to the "global data"[8] in order to develop a logistic model for binomial data. In this

---

7 The variables associated with the original faculty were deleted from the list of possible explanatory variables. In addition, the numeric-type variables, such as average marks at exams and upon degree, were categorised. The loss of information connected with this transformation caused negative consequences that are nevertheless insignificant for our purpose, but are advantageous from a computational and interpretative point of view.

8 The term global data means that all observations showing the same levels of the investigated factors are grouped together.

case, the unusual data will be groups of subjects with a common profile. Analysis of the structural composition of these groups might reveal, for example, contingencies sufficiently significant to justify the moderate dropout rates within contexts where the dropout phenomenon is particularly serious.

Table 3. Main results of the logistic model for data grouped according to the levels of the factors, identified at the start of the analysis

```
Summary of glm(formula = y ~ County + Degree + AgeEnroll + HSscore,
  Family = binomial(link = "logit"), data = abnd03, weights = enr)

  Deviance Residuals:
        Min           1Q      Median          3Q         Max
  -2.8730178  -0.8601809   0.0091846   0.7433140   2.9480225

  Coefficients:   Estimate Std. Error z value  Pr(>|z|)
  (Intercept)   -1.298283   0.155187 -8.3659 < 2.2e-16 ***
  County2        0.256007   0.081815  3.1291 0.0017534 **
  County3        0.342038   0.073441  4.6573 3.203e-06 ***
  County4        0.654752   0.108170  6.0530 1.422e-09 ***
  County5        0.708523   0.098155  7.2184 5.260e-13 ***
  Degree2       -0.593890   0.167855 -3.5381 0.0004030 ***
  Degree4       -0.189403   0.275305 -0.6880 0.4914699
  Degree5       -0.115360   0.171284 -0.6735 0.5006290
  Degree6       -0.164708   0.170176 -0.9679 0.3331091
  Degree7       -0.177111   0.155681 -1.1376 0.2552667
  Degree9       -0.603253   0.179289 -3.3647 0.0007663 ***
  Degree10       0.143121   0.174118  0.8220 0.4110916
  Degree11      -0.342498   0.200154 -1.7112 0.0870500 .
  Degree49      -0.295448   0.156798 -1.8843 0.0595299 .
  Degree52      -0.045730   0.163929 -0.2790 0.7802749
  Degree56      -0.274399   0.187682 -1.4620 0.1437303
  Degree58      -0.606601   0.204628 -2.9644 0.0030326 **
  AgeEnroll2     0.552379   0.078484  7.0381 1.949e-12 ***
  AgeEnroll3     0.958635   0.079691 12.0293 < 2.2e-16 ***
  AgeEnroll4     1.262348   0.114303 11.0438 < 2.2e-16 ***
  HSscore2      -0.374010   0.063411 -5.8982 3.676e-09 ***
  HSscore3      -0.710940   0.097891 -7.2626 3.797e-13 ***
  HSscore4      -0.983368   0.110800 -8.8752 < 2.2e-16 ***

  Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of Fisher Scoring iterations: 3
Null deviance: 1126.89  on 453  degrees of freedom
Residual deviance:  495.73  on 431  degrees of freedom
AIC: 1688.75
  Analysis of Deviance Table
  Terms added sequentially (first to last)
             Df Deviance Resid. Df Resid. Dev  P(>|Chi|)
  NULL                         453   1126.889
  County      4   61.712       449   1065.178 1.2666e-12
  Degree     12   81.135       437    984.042 2.5057e-12
  AgeEnroll   3  383.222       434    600.820 9.5323e-83
  HSscore     3  105.091       431    495.730 1.2492e-22
```

The subjects are grouped according to the covariates detected during the preliminary stage (*faculty, residence, age at enrolment* and *final marks of high school*). It is well known that if the binomial denominators are not large enough, the usual goodness of fit statistics do not meet with any known distribution. Moreover, it is likely that very small groups, which yield very little information, might disturb the search algorithm. Hence, we decided to remove groups containing less than five individuals.

The model for binomial data adapted to the new dataset is shown in Table 3. The groups thus created are 454. The deviance of this model demonstrates that adaptation is more than sufficient (the residual deviance is just slightly higher than the levels of free residuals). The effects of the factors and the single levels are evident, with the exception of those referring to certain faculties that do not show significant effects on the response due to the high levels of the standard errors. Once the working model has been defined, the *Forward Search* algorithm is applied.

An important result is the graph of the statistical test for the goodness of fit of the link function (Figure 1). The test values demonstrate a decrease beyond the limits of significance[9] in the final part of the search algorithm, due to the presence of groups that differ considerably from the bulk of the data.

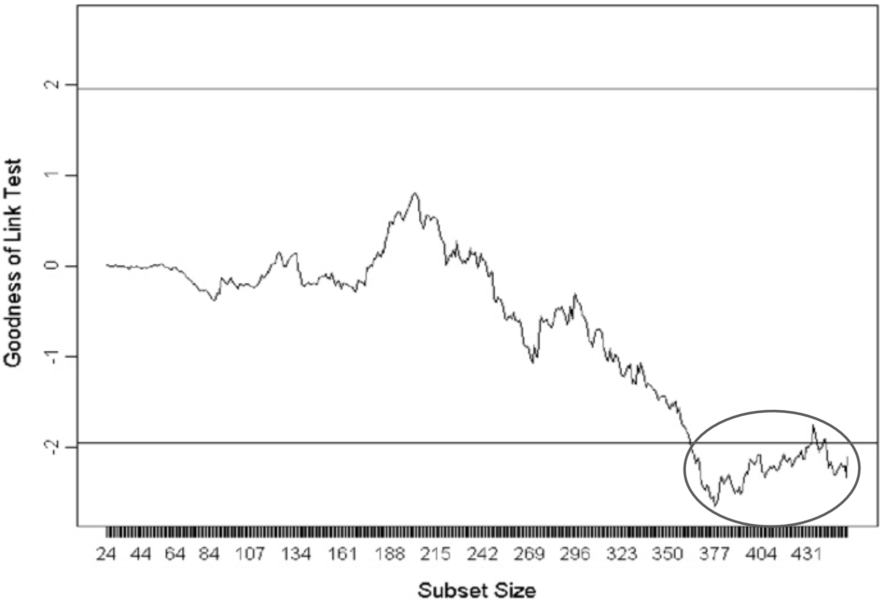


Figure 1. Forward Search: goodness of fit test of the link function

[9] The test has been structured so that values statistically equal to zero (1 - α = 95%) tend towards the goodness of the function chosen.

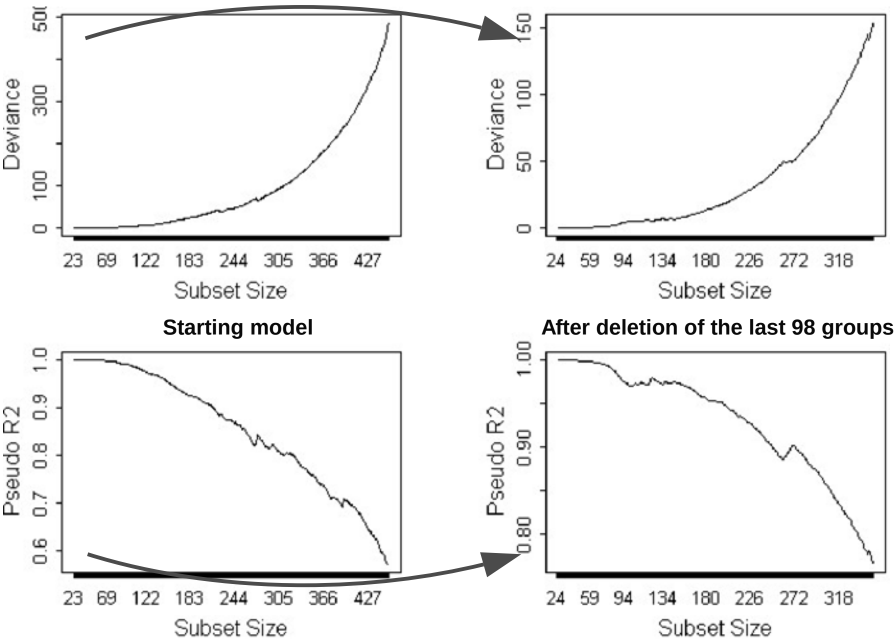


Figure 2. Forward Search: model deviance before and after the deletion of groups of outliers

Therefore, the last 98 groups are worthy of further investigation. The presence of observations with abnormal trends can be detected also from the graphs referring to explained deviation and R^2 index (Figure 2). The last 98 groups cause an exponential increase in the residual deviation (from 150 to 500) and a great decrease in the Pseudo R^2 values (from 0.80 to 0.60).

It is interesting to see how the results would have been different after deletion of the groups at the final stages of the search. The deletion contributed evidently to the adaptability and stability of the goodness of fit test (Figure 3).

As an example, we can examine the composition of group 270, the last one that entered in the analysis. It is composed of 12 subjects attending Medicine, 11 of which are enrolled in the Nursing programme. The dropout rate is 83.3% of students (10 out of 12 students) and 81.8% with reference to the programme (9 out of 11); this latter value is very high in comparison with the mean of the overall course enrolments (27.4%, see Table 4).

This may depend on the profile of the group: these are subjects much older at the enrolment and with lower high school marks than the others. Moreover, they lived in Tuscany but not in the provinces of Florence and Prato, whereas the percentage of students residing in these provinces for the entire programme duration was 27.4%. This suggests that dropout is attributable more to personal characteristics of the students than to the organisation of the educational programme.

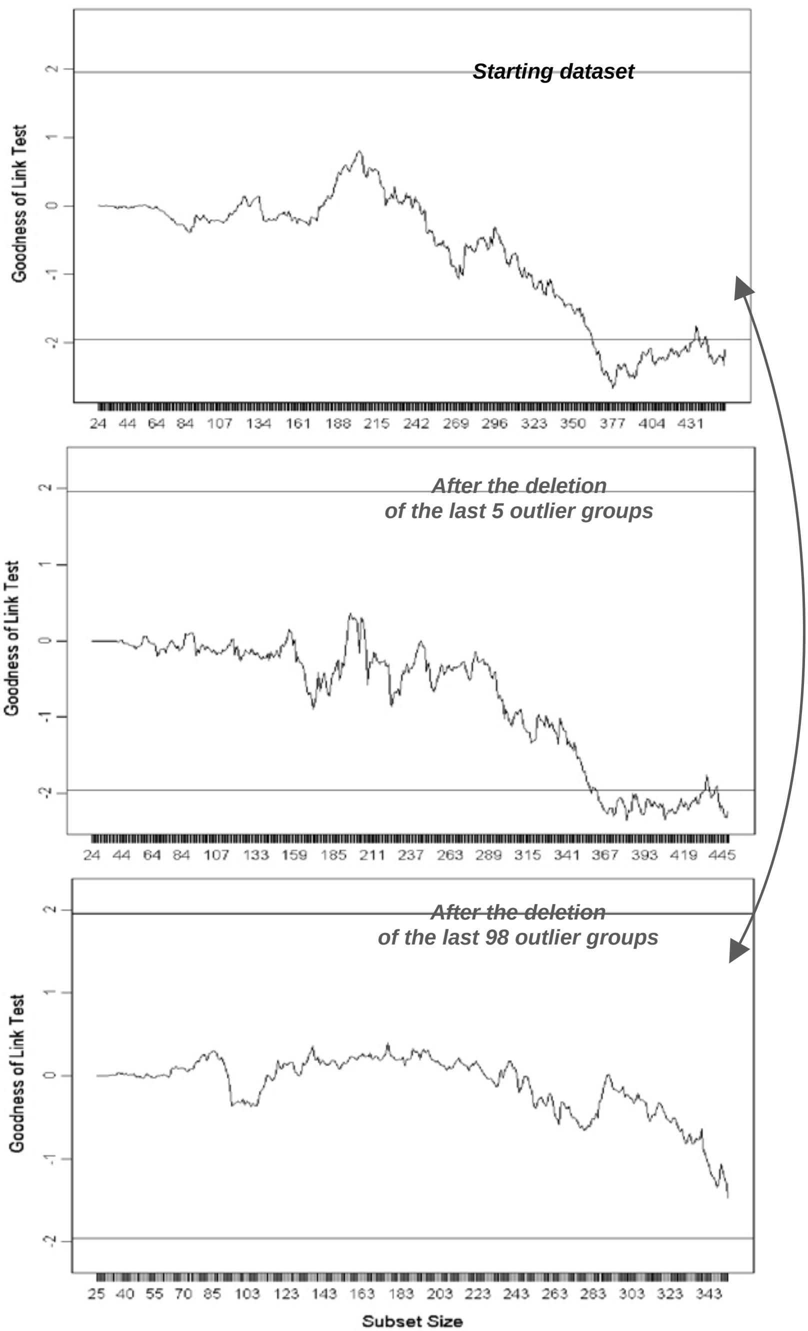


Figure 3. Forward Search: goodness of fit test of the link function before and after the deletion of groups of outliers

Table 4. Analysis of the composition of cluster 270

Cluster composition and characteristics

Course	Obs	Variable	Mean
Nursing	11	AgeEnrollment	32.54
		HighSchoolScore (36-60)	37.90
(92% of the cluster)		*Y (drop rate)*	***81,8%***

Note:
Freshmen enrolled from Tuscany but out of the counties of Florence and Prato

Course characteristics

Course	Obs	Variable	Mean
Nursing	197	AgeEnrollment	21.73
		HighSchoolScore (36-60)	42.13
		Y (drop rate)	***27,4%***

Note:
31% of freshmen enrolled from Tuscany but out of the counties of Florence and Prato

5. Concluding remarks

The analyses performed, together with the evidence of a certain level of performance in university education, are useful tools for programming and organising facilities and educational activities.

Nonetheless, in order to carry out a more thorough analysis of the complex system of relationships and factors that influence the drop-out rate, it will be necessary to use specific analytical models. The robust diagnostic tools applied to regression analysis are capable not only of supplying a reply to this necessity, like the regression models based on traditional estimation methods, but can also identifying the units or groups of units with particular characteristics. These observations may be a source of information that is useful for defining new educational programmes, for improving their quality and, as a consequence, for reducing the rate of drop-out at university.

References

AGRESTI A. (2002) *Categorical Data Analysis, 2nd Ed.*, Wiley, New York.

ATKINSON A. C. (1985) *Plots, Transformations and Regression*, Oxford University Press, Oxford.

ATKINSON A.C., RIANI M. (2000) *Robust Diagnostic Regression Analysis*, Springer, New York.

BERTACCINI B. (2000) *Misure di efficacia esterna dell'istruzione universitaria: indicatori statistici e analisi robusta* (B.A. dissertation), University of Florence.

BINI M. (1999) *Valutazione della Efficacia dell'Istruzione Universitaria rispetto al Mercato del Lavoro*. Rdr 03/99. Osservatorio per la Valutazione del Sistema Universitario - Ministero dell'Università e della Ricerca Scientifica e Tecnologica.

CHATTERJEE S., HADI A. S. (1988) *Sensitivity Analysis in Linear Regression*, Wiley, New York.

CHIANDOTTO B., BERTACCINI B. (2003) *Profilo e sbocchi occupazionali dei laureati e diplomati dell'Ateneo fiorentino nell'anno 1999*, University of Florence.

COOK R. D., WEISBERG S. (1982) *Residual and Influence in Regression*, Chapman and Hall, London.

DONOHO D.L., HUBER P.J. (1983) The notation of breakdown point, In: BICKEL P.J., DOKSUM K., HODGES J.L.Jr (eds) *A festschrift for Erich Lehmann*, Wadsworth Inc., Belmont, CA: 157-184.

ROUSSEEUW P.J. (1984) Least Median of Square Regression, *Journal of the American Statistical Association*, ***85***: 633-639.

ROUSSEEUW P.J., LEROY A.M. (1987) *Robust Regression and Outlier Detection*, Wiley, New York.

A Structural Model of the Employment Pathways of the University of Foggia Graduates[1]

Corrado Crocetta*, Francesco d'Ovidio**

* *Department of Economic, Mathematic and Statistic Sciences, University of Foggia, Italy*
** *Department of Statistic Sciences, University of Bari, Italy*

Summary. The purpose of this study is to investigate the strategies used by graduates of the University of Foggia to enter the labour market. Using both quantitative and qualitative variables, quantified by means of optimal scaling, a structural equation model has been created to analyse the relations between latent variables tied to university education, and graduates' expectations and methods of job searching. Furthermore, we study if the correlation structure between these latent variables is constant observing separately female and male graduates.
Keywords: Graduates; Labour market; University of Foggia; Factor analysis; Structural equation models; Optimal Scaling; CATPCA; LISREL.

1. Introduction

In the following, we intend to analyse the strategies used by graduates to enter the labour market. Because of difficulties in the availability of the data, our analysis considers only the labour offer side.

Our objective is to construct an econometric model capable of detecting the variables that influence the placement of graduates and to study the relations between these variables and the *labour situation* (employed/unemployed) of graduates.

A great deal of the information collected with surveys can be considered *proxy* of latent variables which are particularly useful for describing a fact not

[1] In this joint work, C. Crocetta was responsible for the final editing of Sections 1, 2, 4, 6 and 7, whereas F. d'Ovidio was responsible for Sections 3 and 5. The authors wish to thank anonymous referees for their precious suggestions.

directly measurable, like the behaviour of graduates looking for employment (Crocetta & d'Ovidio, 2004).

If we have to analyse both quantitative and qualitative variables together, we can apply the Optimal Scaling procedure known as CATPCA (CATegorical Principal Component Analysis). Such a procedure may be used to quantify, in the factorial space, the observed variables (De Leeuw, 1984; Meulman & Heiser, 1999).

For this, we performed a series of analysis to verify:

1. which latent variables are to be considered,
2. the relations that exist between them,
3. whether it is possible to construct a model for explaining the variability of employment rates of graduates,
4. whether there are other variables, not considered in the model, which can explain the employment,
5. whether, by considering separately female and male graduates, the structure of the correlation between the latent variables changes significantly.

2. A survey on graduates of the University of Foggia

The data for this analysis came from the archives of the Student office of the University of Foggia and was integrated with telephone interviews of a sample of graduates[2]. The questionnaire is divided into four parts, which are dedicated to graduates who are already working, those who are looking for employment, and those with a work experience in progress, and, finally, to the satisfaction regarding the services and the preparation achieved.

We have found that, as the time upon graduation passes, the number of fixed term contract grows (Table 1). This situation regards all those who graduated in Agriculture, 92.5% of graduates in Economics and 83.3% of graduates in Law. However, by considering the graduates in the period 1997-1999, doctors in Medicine have the highest rate of stable position (83.3%), followed by their colleagues in Economics (75.6%) and Law (74.7%), while just 67.7% of graduates in Agriculture during the same period have a stable job.

The attainment of a stable job is not so difficult even for those who graduated less than three years before (59.3% of them are in this condition). In this case, too, graduates in Medicine (70.0%) have a small advantage over those in Economics (61.2%), in Agriculture (58.8%) and in Law (56.1%).

[2] The first objective was to contact all the 2,924 graduates of the Athenaeum of Foggia starting from when it was set up (1994). Up to seven attempts were made to contact them at different times, before they were considered unreachable. A number of 2,133 interviews was performed, which is the 72.3% of the eligible population.

Table 1. Percentage distribution of the employed graduates from the University of Foggia, according to the year of graduation, faculty, and work contract

Contract	Faculty				
	Economics	Agriculture	Law	Medicine	Total
	Years 1994-96				
Stable employment (full or part-time)	92.5	100.0	83.3	-	88.6
Temporary job or job training scheme	5.0	-	9.5	-	6.8
Occasional employment and the like	2.5	-	7.1	-	4.5
	Years 1997-99				
Stable employment (full or part-time)	75.6	67.7	74.7	83.3	74.7
Temporary job or job training scheme	11.9	29.0	12.4	16.7	13.5
Occasional employment and the like	12.5	3.2	12.9	-	11.9
	Years 2000-02s				
Stable employment (full or part-time)	61.2	58.8	56.1	70.0	59.3
Temporary job or job training scheme	27.2	41.2	30.2	30.0	28.9
Occasional employment and the like	11.6	-	13.8	-	11.8
	All employed graduates				
Stable employment (full or part-time)	69.2	68.5	68.0	75.0	68.7
Temporary job or job training scheme	19.7	29.6	19.3	25.0	20.1
Occasional employment and the like	11.1	1.9	12.7	-	11.2

The proportion of employed graduates with a fixed term contract or on a job training scheme tends to decrease as time passes. A period of precariousness seems, however, inevitable, especially for graduates in Agriculture (29.6%) and Medicine (25.0%). Graduates in these two faculties, unlike their colleagues of Economics and Law, tend not to accept seasonal and temporary work. Presumably, the number of graduates of the two former faculties is not so large, and this may prevent the competition that often makes graduates accept temporary or unsuitable jobs.

3. Analysis of the categorical components of the model

We carried out a descriptive analysis for screening the variables for the model. After that, we estimated a logit model, whose criterion variable is the binary position, employed *vs.* unemployed, of graduates. The variables whose coefficients were significant at 5% level are listed in Table 2.

Some variables are quantitative (such as the 0-100 score of the suitability of the university training received) or on an ordinal scale, but many others are nominal.

Table 2. Variables selected for the logit model

Quantitative variables	Nominal variables
▪ University graduation final grade.	▪ Faculty.
▪ Secondary school graduation final degree.	▪ Type of Secondary School degree.
▪ Age at university graduation.	▪ Pre-graduation work experience.
▪ Number of years between second. school graduation and university enrolment.	▪ Field of economic activity currently employed in or searching for.
▪ Number of years after the end of course before graduating. (*off-programme*)	▪ Current or desired professional position
▪ Number of months between university graduation and first employment.	▪ Professional or teaching qualification
▪ Overall score assigned to the adequacy of university education with respect to employment obtained or desired.	▪ Employment search methods. ▪ Knowledge of post-graduation prospects (orderable nominal variable).

We wanted to apply a LISREL model. Because the structural equation model, given the hypothesised normality of the latent variables, does not allow using categorical variables[3], it was necessary to quantify such variables with an Optimal Scaling (OS) method.

Given a population of n individuals described by a set of m categorical variables $\mathbf{x}_1 \ldots \mathbf{x}_j \ldots \mathbf{x}_m$, the OS procedure transforms the categories into real values $\boldsymbol{\omega}_j$. OS methods minimize a loss function regarding the categories of interest.

First, a scalar g_{ijh} has been defined with value 1 or 0 according to whether the i^{th} individual possesses the h^{th} category of the $\mathbf{x}_j$ variable. The vector $\mathbf{g}_{jh}$ is given by this scalar attached to the units in category h of x_j.

With all categories of $\mathbf{x}_j$, the column vectors $\mathbf{g}_{jh}$ originate the indicator-matrix $\mathbf{G}_j$ (of dimensions $n \times k_j$). Extending this procedure to all the m categorical variables, we obtain the disjunctive complete indicator-matrix, $\mathbf{G} = [\mathbf{G}_1 \ldots \mathbf{G}_j \ldots \mathbf{G}_m]$, of the order $n \times K$, where $K = \sum_j k_j$.

In this way, each categorical variable is a product of an indicator-matrix by a vector $\boldsymbol{\omega}_j = [\omega_{j1} \ldots \omega_{jh} \ldots \omega_{jk_j}]'$ of scaling parameters that, once estimated ($\hat{\omega}_{jh}$), originate quantitative variables:

$$^{os}\mathbf{x}_j = \mathbf{G}_j \hat{\boldsymbol{\omega}}_j \qquad \text{or, equivalently,} \qquad ^{os}\mathbf{x}_j = \sum_{h=1}^{k_j} \mathbf{g}_{jh} \hat{\omega}_{jh} \qquad (j=1, 2, \ldots, p),$$

where the superscript "OS" indicates the optimally scaled variable. Extending this procedure to all the units of the population and all variables, we obtain the matrix of optimally scaled individual scores, $^{os}\mathbf{X} = (^{os}\mathbf{X}_1, {}^{os}\mathbf{X}_2, \ldots, {}^{os}\mathbf{X}_m)$.

---

[3] The estimates of the LISREL model parameters with the method of the maximum likelihood are asymptotically biased, because of the violation of the normality hypothesis of the latent variables and, implicitly, of the observed variables (Browne, 1984). This problem may be overcome by using non-parametric loss functions, such as WLS, GLS and ULS (see Lovaglio, 2000).

The vectors $\boldsymbol{\omega}_j$ may be estimated by optimizing a target function with identification constraints. It is worthwhile estimating simultaneously the quantities of the categorical variables, and the parameters of the model[4], by directly optimizing the target function with ALSOS (Alternative Least Squares Optimal Scaling) methods (De Leeuw *et al.*, 1976; Young *et al.*, 1978; Vittadini, 1999).

Among the available ALSOS procedures, we chose CATPCA (CATegorical Principal Component Analyses), a non-parametric algorithm using the main components of the transformed variables in a factorial *p*-dimensional space ($p \leq m$) (De Leeuw & Meulman, 1986; Meulman & Heiser, 1999).

In neither the simple case of no weighting for cases or variables and no supplementary nor multiple variables, the optimisation procedure estimates the ω_j scaling parameters (iteratively) by minimising the function

$$\sigma(\mathbf{Y};\boldsymbol{\Omega}) = n^{-1}\sum_{j}^{m} \mathrm{tr}\left[(\mathbf{Y}-\mathbf{G}_j\boldsymbol{\Omega}_j)'\,\mathbf{M}_j\,(\mathbf{Y}-\mathbf{G}_j\boldsymbol{\Omega}_j)\right],$$

where matrix $\mathbf{M}_j$ is diagonal (of $n \times n$ order) with elements 0 if the i^{th} observation is missing and 1 in the other cases; whereas $\mathbf{Y}$ (of $n \times p$ order) represents the *p*-dimensional standardised factor scores, with the following constraints of standardisation or centring (given $\mathbf{u}$, unit-vector of order n, and $\mathbf{M}=\sum_j \mathbf{M}_j$):

$$\mathbf{Y}'\mathbf{M}\,\mathbf{Y} = n\,\mathbf{I}_p, \qquad \mathbf{u}'\mathbf{M}\,\mathbf{Y} = \mathbf{0}. \tag{1}$$

The algorithm begins with an estimate of $\mathbf{Y}$ which satisfies those constraints (unless otherwise specified, standardised and centred random numbers); the initial factor loadings $\mathbf{a}_j$ are calculated as cross-product between $\hat{\mathbf{Y}}$ and the categorical codes centred and re-scaled: ${}^{c}\mathbf{x}_j = [\mathbf{I}_n\text{-}\mathbf{M}_j\mathbf{uu}'/(\mathbf{u}'\mathbf{M}_j\mathbf{u})]\ \mathbf{x}_j$, with j= 1, 2, ..., m (De Leeuw *et al.*, 1976; Meulman & Heiser, 1999).

The first step of the iteration consists of calculating, given $\mathbf{D}_j = \mathrm{diag}(\mathbf{G}'_j\mathbf{G}_j)$,

$$\hat{\boldsymbol{\Omega}}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\,\hat{\mathbf{Y}}\,. \tag{2}$$

After a first quantifications of categories[5] we standardised the data with $\hat{\boldsymbol{\omega}}_j^{\perp} = \hat{\boldsymbol{\omega}}_j\sqrt{n\ /(\hat{\boldsymbol{\omega}}_j'\,\mathbf{D}_j\ \ \hat{\boldsymbol{\omega}}_j)}$ to compute factor weights $\mathbf{a}_j = (\boldsymbol{\Omega}_j'\ \mathbf{D}_j\ \hat{\boldsymbol{\omega}}_j^{\perp})/n$. Through the standardised matrix of the scaling estimates $\hat{\boldsymbol{\Omega}}_j^{\perp} = \hat{\boldsymbol{\omega}}_j^{\perp}\mathbf{a}_j'$, the matrix $\hat{\mathbf{Y}} = [\mathbf{I}_n\text{-}\mathbf{Muu}'/(\mathbf{u}'\mathbf{Mu})]\ (\sum \mathbf{M}_j\mathbf{G}_j\hat{\boldsymbol{\Omega}}_j^{\perp})$ could be now calculated. The process keeps resuming the algorithm with a *singular value decomposition* of $\hat{\mathbf{Y}}$ starting again from [2]. After a certain number of iterations, the final estimates $\hat{\boldsymbol{\omega}}_j$ of the categories are obtained.

---

4 The scaling approach is not separable from the aims of the research, and quantification must be obtained under specific statistical models (Bradley *et al.*, 1962; Kruskal, 1965; de Leeuw *et al.*, 1976).

5 If variables are categorical, the factorial weights $\hat{\boldsymbol{\omega}}_j = \hat{\boldsymbol{\Omega}}_j\,\mathbf{a}_j$ are directly used; if variables are ordinal, the weights $\hat{\boldsymbol{\omega}}_j$ are obtained through a monotone regression of the weighted $\hat{\boldsymbol{\Omega}}_j\,\mathbf{a}_j$ with the diagonal elements of $\mathbf{D}_j$, whereas, if they are numeric, a weighted linear regression is used.

Table 3. Percent distribution of the University of Foggia graduates according to work condition and some other characteristic

	Labour condition Unem-ployed	Employed		Labour condition Unem-ployed	Employed
University Faculty			*Field of activity employed in/searching for*		
Medicine and Surgery	80.4	19.6	n.a.	85.0	15.0
Law	59.4	40.6	Public Administration	64.9	35.1
Economics	42.2	57.8	Industry	64.1	35.9
Agriculture	39.6	60.4	Commerce	46.8	53.2
Type of secondary school diploma			Agriculture	29.3	70.7
Languages	90.9	9.1	Services	26.0	74.0
Others	71.4	28.6	Other field	27.4	72.6
Classical	60.4	39.6	*Current or desired professional position*		
Scientific	55.2	44.8	n.a.	98.1	1.9
Teacher training	51.6	48.4	Consultant.	61.5	38.5
Technical commercial	50.2	49.8	Entrepreneur	50.0	50.0
Professional	48.7	51.3	Employee/Manager	40.9	59.1
Technical Geometer	33.3	66.7	Self-employed	35.0	65.0
Knowledge of post-graduate prospects			Teacher/Professor	11.3	88.7
Yes, quite well	49.3	50.7	Other Position	14.5	85.5
Yes, in a generic way	59.9	40.1	*Employment search method*		
No	48.4	51.6	n.a.	92.0	8.0
Pre-graduation employment			Newspaper/Internet	86.3	13.7
Never worked before	56.5	43.5	Specialized Agencies	80.9	19.1
Worked before	48.0	52.0	Local employm.agency	79.7	20.3
Professional/teaching qualification			Curriculum sent	39.4	60.6
Not licensed	60.0	40.0	Interviews/exams	36.8	63.2
Licensed	31.3	68.7	Personal Contacts	29.0	71.0
			Direct calls	-	100.0
			Suggestions	9.8	90.2
Total	***53.0***	***47.0***	Other search-methods	13.6	86.4

In Table 3, the categories are shown in decreasing order of employment rate.

CATPCA defines just four components with eigenvalues higher than one (Table 4). Altogether, they explain 52% of the overall variability. For a better definition and identification of the factors, we performed a Varimax rotation of the factorial axes. The first factor accounted for 16.1% of the variability, whereas the forth accounts for 9.6%.

Let us now identify the four latent variables (Table 5). The *first factor* is directly connected to the age at graduation and to time between high school graduation and university enrolment, whereas it presents a negative correlation with the University and high school final grade. We named this factor "*regularity and proficiency of educational career*".

Table 4. Variability explained by the main components and rotated factors*

Components	Weights of unrotated components			Weights of rotated factors		
	Eigenvalues	% of variance	% cumulated variance	Eigenvalues	% of variance	% cumulated variance
1	2.54	18.1	18.1	2.26	16.1	16.1
2	1.93	13.8	31.9	1.86	13.3	29.4
3	1.68	12.0	43.9	1.82	13.0	42.4
4	1.13	8.1	52.0	1.34	9.6	52.0

* 14 components have a non-zero eigenvalue. Extraction Method: Principal Components Analysis. Rotation with Varimax Method.

Table 5. Coefficients obtained from factor analysis of variables optimally scaled by means of CATPCA for the graduates of the University of Foggia

Variables	Communality	Components			
		1	2	3	4
Time between graduation and first employment	0.66	*-0.17*	*0.27*	0.75	*-0.05*
Time between last exam and graduation	0.63	0.63	*0.14*	-0.46	*0.08*
Professional or teaching qualification	0.63	*-0.05*	*-0.16*	0.78	*-0.03*
Age at graduation	0.62	0.74	*0.03*	*-0.03*	*0.26*
Graduation final grade	0.59	-0.76	*0.08*	*0.12*	*0.02*
Current or sought professional position	0.54	*0.00*	0.74	*-0.05*	*-0.03*
Employment search strategies	0.52	*0.15*	0.50	0.48	*-0.15*
High school final grade	0.51	-0.70	*0.08*	*-0.10*	*0.05*
Faculty	0.50	*-0.27*	0.59	*-0.17*	*0.22*
Knowledge of post-graduation prospects	0.46	*0.07*	*0.11*	*0.06*	-0.67
Current or sought field of activity	0.44	*0.03*	0.61	*0.25*	*-0.03*
Pre-graduation work experience	0.42	*0.15*	*0.13*	*-0.07*	0.62
Type of high school diploma	0.38	*0.00*	0.41	*-0.12*	0.44
Time between high school and university enrolment	0.35	*0.30*	*-0.10*	*0.28*	0.41

The *second factor* is directly connected to the field of economic activity, to the professional position achieved, to faculty and type of high school diploma and to job search strategy. Because it is strongly influenced by type of studies and working experience of the graduates, we named it "*training-professional project*".

The *third factor* is correlated to the time spent between graduation and first job, to possession of professional or teaching qualification and to job search strategies, and opposite to time between high school graduation and university enrolment. All these variables have in common the *job finding experiences* of the graduates.

The *fourth factor* is mainly related to the work before graduation, to the type of high school diploma and to the time between high school graduation and university enrolment. This factor has also a negative correlation with the knowledge of the job prospects after graduation[6]. For this reason, we name it "*work plans and experiences*".

4. The structural equation model

LISREL is a structural equation model. It is popular in social sciences to study the cause-effect relations within a system. Generally, structural equation models are constructed with very simple relations.

By applying Path Analysis graphs (Wright, 1934), it is possible to represent the model with flow diagrams in which the surveyed variables are represented with right angles, whereas the latent variables and the erratic components are contained in elliptic shapes. These geometric figures can be connected with arrows that indicate the existence of a relation[7].

There is a distinction between measuring models, which are useful for identifying and measuring the latent variables through the observed variables, and structural models, which explain causal relations between the latent variables. The latter may be *exogenous*, if variables are explicative, or *endogenous* if they can be interpreted also as response variables. The LISREL model (Jöreskog, 1973, 1977; Wiley, 1973; Bollen, 1989) is defined as

$$\boldsymbol{\eta} = \mathbf{B}\,\boldsymbol{\eta} + \boldsymbol{\Gamma}\,\boldsymbol{\xi} + \boldsymbol{\zeta}\,,$$

with measuring equations given by

$$\mathbf{x} = \boldsymbol{\Lambda}_x\,\boldsymbol{\xi} + \boldsymbol{\delta} \quad \text{e} \quad \mathbf{y} = \boldsymbol{\Lambda}_y\,\boldsymbol{\eta} + \boldsymbol{\varepsilon}\,.$$

In this model, the quantities $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are, respectively, the cause and effect latent variables; the observed quantities $\mathbf{x}$ and $\mathbf{y}$ are variables linearly connected to $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ through the matrices of factorial weights $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$; $\boldsymbol{\Gamma}$ is the matrix of the coefficients of the cause variable in the structural relation, $\boldsymbol{\zeta}$ is the vector of random errors in the structural relation between $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, whereas $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are the vectors of measurement errors of $\mathbf{x}$ and $\mathbf{y}$ respectively; $\boldsymbol{\zeta}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are not correlated to one another, nor with $\boldsymbol{\xi}$, $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$.

---

[6] Crocetta & d'Ovidio (2003) stated that working during the university studies helps job finding prospects. However, if it concerns a job started before university enrolment, it is likely that it is maintained.

[7] For the relations of dependence, the previous character is the one represented in the graphic element from which the arrow goes, whereas the following one is the one indicated in the box the arrow points. The relations of interdependence are represented with arcs of circumference that have arrows at both ends.

Then we have:

$$E(\boldsymbol{\zeta}) = E(\boldsymbol{\varepsilon}) = E(\boldsymbol{\delta}) = E(\boldsymbol{\xi}) = E(\boldsymbol{\eta}) = \mathbf{0}\,,$$
$$\mathrm{Cov}(\boldsymbol{\zeta}) = \boldsymbol{\Psi}\,, \quad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}_{\varepsilon}\,, \quad \mathrm{Cov}(\boldsymbol{\delta}) = \boldsymbol{\Theta}_{\delta}\,, \quad \mathrm{Cov}(\boldsymbol{\xi}) = \boldsymbol{\Phi}\,,$$

where $\boldsymbol{\Phi}$ is the matrix $k\times k$ of co-variance of latent factors and $\boldsymbol{\Theta}$ are diagonal matrices of only variances.

We can estimate the coefficients and the matrices of variances and co-variances with various techniques (Jöreskog, 1973; Jöreskog & Goldberger, 1975; Browne, 1974). We chose GLS estimators because they are robust to non-normality of the distribution of the latent variables (Browne, 1984)[8]

The LISREL models can be used to analyse the data coming from several groups thus giving the opportunity of making comparisons with control groups or between groups undergoing different treatments.

It is possible to impose constraints on some or all the parameters considered. If we want to compare two groups, it is necessary to estimate each group separately for there to be no bonds; whereas if the data has to be analysed simultaneously to have efficient estimates, crossed constraints must be imposed between groups (Bollen, 1989; Kline, 1998; Civardi & Zavarrone, 2000, 2002).

We have to check whether the matrices of co-variances and correlation of the observed variables are equal for each group. To verify the equality of the matrices of correlation of $\mathbf{x}$, it is necessary to set $\boldsymbol{\Theta}_{i;\delta} = \mathbf{0}$ and $\boldsymbol{\Lambda}_{i;x}$, as diagonal matrices of the standard deviations of $\mathbf{x}$ (where i=1, …, m denotes the group) and $\mathbf{0}$ is a null matrix. Testing the hypothesis of equality between correlation matrices is like checking that $\boldsymbol{\Psi}_1 = \boldsymbol{\Psi}_2$, where $\boldsymbol{\Psi}_i$ is the correlation matrix between the latent factors of group i.

If the hypothesis of invariance of the model is refused without any restriction, more constraints may be imposed to verify the causes of the lack of equality. First, the hypothesis of invariance of the initial factor weights can be tested for the measurement model in each of the groups. If this hypothesis is not acceptable, the invariance of the covariance of the unique factors and factor weights can be tested. The third hypothesis foresees the test of invariance of the co-variances of the unique factors and of the variances of the common factors and factor loadings.

If the hypotheses are less rigid, we can use the first additional hypothesis of the structural model. This foresees covariance matrices of the unique invariant factors symmetrical with some similar elements. Then the case with other constraints, in which the covariance matrices of the unique factors are invariant and symmetric to some elements set to zero, can be tested.

---

[8] The quantification of each categorical variable through optimal scaling is referred to a limited number of modalities. So, it is not recommended to assume the hypothesis of normality of the latent variables.

5. A structural model for evaluating how graduates work

LISREL results represent just the starting point of the analysis because to obtain a model with convergent estimates it has been necessary to do small changes. The latent variables used in our analysis are the factors determined by means of factor analysis, with the exception of the variable *irregular studies*, whose importance has been reduced by the relations between latent variables. The starting variables are those obtained with the optimal scaling quantification procedure. The resulting model (see Figure 1) is laid out as a network of relations and it is complex; so we will consider the main correlations between the observed variables surveyed and the latent variables.

The most correlated (observed) variables are age at graduation and time passed between graduation and employment (r=0.95); this variable is also connected to the business field of employment (r=0.49) and study irregularity (r=0.41). It is noticeable the relation between the graduation grade and school leaving grade (r=0.38), and the correlation between type of course chosen and time between graduation and first job (r=0.35). The other correlations considered are lower than 0.33 in absolute terms; they are significant and have to be kept in the model to help the convergence or to improve adaptability.

The standardised regression weights show the direction and intensity of relations between latent and observed variables. Those relations slightly differ from those that come from the explorative analysis, because of the causal relations hypothesised between latent variables and the structure of the factors themselves.

The first factor was called education curriculum because it was directly correlated to graduation and secondary school final grades, and inversely to age at graduation and time between graduating and the last exam. The structural model keeps the same relations and adds a slight connection to the evaluation based on the adequacy of university education and study irregularity. This latent variable is influenced by the graduates' work plans and influences, in its turn, the post-graduate activities.

The latent variable *training and professional path* is mainly measured by the faculty chosen (regression coefficient=0.73), whereas the relations between the types of diploma achieved, the job achieved or sought, and the business field are less strong. The *work plan* factor has a positive relation with the knowledge of post graduation prospects, and negative (-0.69) with work before graduation. It influences (0.32) the non-observed variable, education curriculum. The *postgraduate activity* factor[9] influences three observed variables: years between graduation and employment (0.95), professional or teaching qualification (0.42) and job finding strategies (0.28).

---

9 Differently from exploratory factor analysis, there is no business field related with the educational and professional pathway, and time between last exam and graduation.

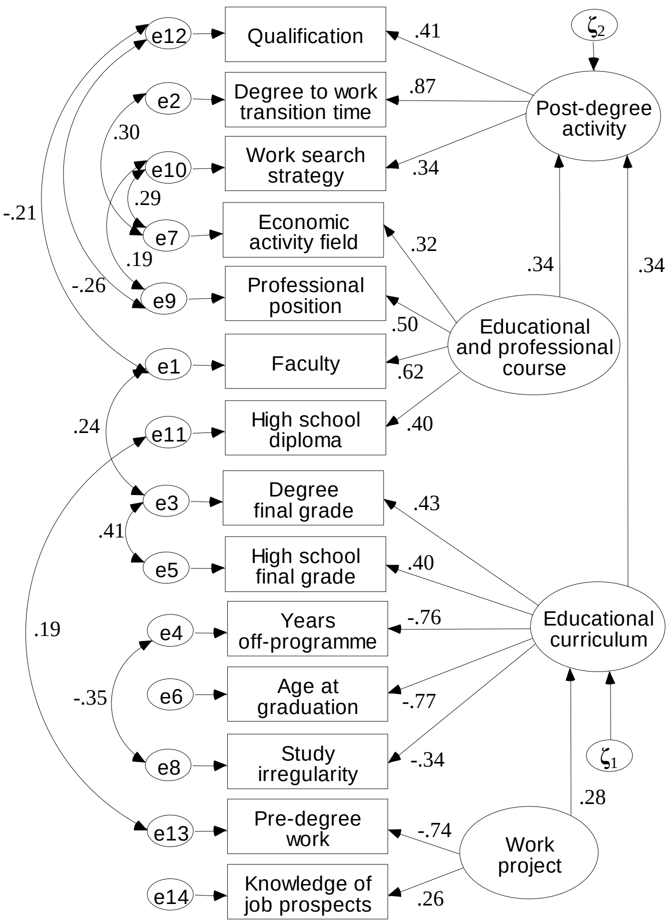


Figure 1. LISREL Model to describe the working way of the graduates of the University of Foggia (Italy)

Table 6. Fit Indexes of the LISREL model used for the analysis of working strategies of graduates from the University of Foggia

Fit Indexes	Models		
	Estimated	Saturated	Independence
ECVI *(Expected Cross-Validation Index)*	0.33	0.10	1.19
GFI *(Goodness of Fit Index)*	0.96	1.00	0.83
AGFI *(Adjusted Goodness of Fit Index)*	0.93	-	0.81
RMR (*Root Mean Square Residual*)	0.07	0.00	0.22
RMSEA *(Root Mean Square Error of Approximation)*	0.06	-	0.11
HOELTER critical N (α=0,05)	*296*	-	*98*

This model represents a first step of research in this field. Indeed, to obtain reliable estimates on employment rates or on the probability of entering the labour market, methods that are more sophisticated have to be used with information that is more detailed.

The obtained model fitted well the data, as is shown by the statistics reported in Table 6. The ECVI index is much closer the minimum value regarding the saturated model (0.10) than the independence one (1.19), indicating that the discrepancy index is rather low. In addition, the GFI index suggests that this model is very close to maximum adaptability[10].

The AGFI (adjusted goodness of fit index) verifies the adaptability of the model looking at the degrees of freedom available to test the model. In our case the value 0.93 is very close to 1, which indicates perfect adaptability.

The RMR (root mean square residual) index is given by the square root of the mean square deviations between the sample variance and its estimate obtained under the assumption that the model is correct. Obviously the lower this index is, the better is its adaptability. In our case, the value 0.07 is much closer to the value of the saturated model (0) than to that of the independence model (0.22).

The RMSEA index is 0.06. In general, a RMSEA lower than 0.08 indicates a good matching of data to the model.

The *N* statistic, for a significance level of 5%, is over the critical threshold (N=200) recommended by Hoelter (1983), whereas for the independence model this statistic is much lower than the suggested level.

Overall, the processed model seems to represent adequately the relations that exist within the data.

6. A structural model of invariance between genders

A variable that may influence the occupational possibility of graduates in Southern Italy is *gender*. Within the groups of graduates analysed here, 54.3% of males and 41.8% of female graduates actually work[11].

Gender in itself is not a determinant of employment, but it is connected to a series of socio-economical obstacles that make it more difficult for women to

[10] The *goodness of fit* (GFI) index is given by the complement to one of the relation between the discrepancy function minimum between model and the sample, in the hypothesis that the variability of the groups is null. This index varies between 0 and 1 where the value 1 indicates the perfect adaptability.

[11] In a segmentation analysis of graduates' placement (Crocetta & d'Ovidio, 2003), this variable appeared in the third branch of the classification tree, thus describing a situation of prevalence of male employment. However, this may depend on the interaction of gender with other structural variables.

enter the labour market than men. Gender is, therefore, a proxy of the lower possibility a woman be assigned roles of responsibility which require either total commitment, or frequent movements, or changes of residence, and implies work discontinuity caused by absences for personal reasons (pregnancy, children's illnesses, etc.)

Gender also influences the choice of educational courses: there are some faculties, such as Literature and Philosophy, with a very low male rate. Another aspect of university choice connected to gender is the distance from home: Antonucci *et al.* (2002) showed a particular inclination for women to choose the university close to their residence.

For these reasons, we decided to check whether the LISREL model described above remains the same with gender. Application of the same structure with males and females immediately produced convergence of the model.

The first hypothesis tested shows invariance of the correlation structure between models estimated separately for male (41.7% of the sample) and female graduates (58.3%): there is no constraint apart from the basic ones ($\boldsymbol{\Theta}_{i;\delta} = \mathbf{0}$; $\boldsymbol{\Lambda}_{i;x} = \mathbf{I}_p$; $\boldsymbol{\Phi}$ matrix with the elements of the main diagonal set to 1). If we test this hypothesis with the minimum value of the discrepancy function, which is distributed as a χ^2, we obtain $C_{MIN} = 40.8$ with 10 degrees of freedom, which corresponds to $p<0.0001$: the hypothesis of *invariance of the structures of correlation* (i.e. of the measuring model) has to be rejected.

The hypothesis of invariance of the *initial factorial weights* is then verified: C_{MIN}=3.3 with 3 degrees of freedom (p=0,134), therefore the hypothesis of invariance of the structural model can be accepted.

Looking at the fit indices (Table 7), it is evident that the adaptation compared to the saturated model does not perform significantly worse, even if the number of groups has more or less halved (the number of female graduates is 1,215, whereas the number the male ones is just 869).

Table 7. Fit Indexes of the LISREL model of invariance of the structural weights (compared to saturated model and model of independence) used for the analysis of working strategies of male and female graduates from the University of Foggia

Fit Indexes	Models		
	Invariance	Saturated	Independence
ECVI *(Expected Cross-Validation Index)*	0.39	0.20	1.23
GFI *(Goodness of Fit Index)*	0.95	1.00	0.83
AGFI *(Adjusted Goodness of Fit Index)*	0.93	1.00	0.80
RMR *(Root Mean Square Residual)*	0.09	0.00	0.22
RMSEA *(Root Mean Square Error of Approximation)*	0.04	-	0.08
HOELTER critical N (α=0,05)	*515*	-	*179*

Table 8. Standardised regression weights of the variables of the LISREL model for the employment analysis of male and female graduates of the University of Foggia

Observed variables	Stand.weights M	Stand.weights F	Observed variables	Stand.weights M	Stand.weights F
Work project			*Educat./professional course*		
Knowledge of post-degree prospects	0.28	0.27	Faculty	0.63	0.61
			Prof. position current/sought	0.54	0.51
Pre-degree employment	-0.70	-0.71	High school diploma	0.36	0.40
Educational Curriculum			Economic field curr./sought	0.30	0.32
University degree final grade	0.39	0.41	*Post-degree activity*		
High School final grade	0.34	0.35	Number of years between degree and first employment	0.89	0.85
Study irregularity	-0.20	-0.22	Professional/teaching qualification	0.45	0.47
Age at graduation	-0.62	-0.78			
Years off-programme	-0.82	-0.87	Work search strategies	0.35	0.33
Latent components of the variable ***Educational curriculum***			Latent components of the variable ***Post-graduation activity***		
Work project	0.34	0.21	*Educational curriculum*	0.43	0.42
			Educat./professional course	0.34	0.29

* The latent variables in bold type are endogenous.

Although the relations between latent and observed variables changed somewhat (Table 8), the factorial structure is invariant.

The two groups of graduates seem to differ in few aspects, such as the *age at graduation* and, to a lesser extent, the *number of off-programme years*, the *type of high school diploma* and the *number of years after completing the course before graduation*.

The relations between latent variables differ very little between genders, as the structural invariance study indicated.

The analysis confirms that male graduates' approach to the labour market is similar to the one of female graduates, but there are differences in female behaviour towards university training, because women study with more regularity.

7. Conclusions

We selected some variables that represent the pathways through which graduates access the labour market.

The obtained estimates allowed us to measure the external effectiveness of University of Foggia education and to study in non-monetary terms the effects of graduates' choices to enter the labour market[12].

Since the relation between education and employment is complex[13], we considered it worthwhile to underline the result based on the individual student (learning level, ability to finding employment, amount of human capital), that is, the final target of the university teaching in the current social and economical system (Gori, Crema & Vidoni, 2003).

It is thus particularly useful to have a model capable of examining quantitative, ordinal and nominal variables altogether through an appropriate scaling procedure.

The analysed models, both the one applied to the whole sample and those predicted to check the relation-to-gender invariance, showed a good level of matching with the observed data. This allows us to trust the reliability of our estimates and of the hypotheses put forward.

The analysis highlighted that the most influential latent variable on *placement* and success in the labour market is *postgraduate activity*, which depends on the *educational and professional pathway* chosen, on *curriculum* and indirectly on the work planning.

However, we showed that there is no difference between male and female graduates. This does not mean that choices and constraints are the same for the two genders, but that the relations between choices, constraints and latent variables are of the same order, and that the mental structure regarding the decisions is similar for both males and females.

---

[12] The comparison between the professional skills of graduates and market demand means "the size and durability of the competences of the trainees, adaptability to situations they have to face, the propensity to learn from experience, the propensity to evolve from technical to managerial work" (Fabbris, 2003).

[13] The knowledge determined by the training process and the consequent increase of capacity of finding work are "experience goods", whose effect can only be evaluated expost at different time intervals (Gori & Vittadini, 1999). The results of the training process, as well as the resources, can be measured in monetary or physical quantities (e.g, hours of lessons, number of graduates, etc.), in order to construct productivity indices for processes, structures, results, extending corporate techniques typical of industrial processes to the university case (Bini, 1999). However, a greater quantity of lessons does not imply a better learning and qualification for the labour world (Vittadini, 2001).

References

ANTONUCCI L., CROCETTA C., MECCARIELLO I. (2002) *La popolazione studentesca e l’Università di Foggia, Indagine sulle aspettative degli studenti dopo la riforma*, Uniongrafica Corcelli, Bari.

BINI M. (1999) *Valutazione dell’efficacia dell’istruzione universitaria rispetto al mercato del lavoro,* Rapporto di Ricerca 3/99, Comitato Nazionale per la Valutazione del Sistema Universitario. MIUR, Roma.

BOLLEN K.A. (1989) *Structural Equations with Latent Variables,* Wiley & Sons, New York-Toronto.

BRADLEY R.A., KATTY S.K., COONS I.J. (1962) Optimal scaling for ordered categories, *Psychometrika*, **27**: 355-374.

BROWNE M.W. (1974) Generalized least-squares estimators in the analysis of covariate structures, *South African Statistical Journal*, **8**: 1-24.

BROWNE M.W. (1984) Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures, *British Journal of Mathematical and Statistical Psychology*, **37**: 62-83.

CIVARDI M., ZAVARRONE E. (2000) Il ricorso a modelli invarianti per valutare la qualità della didattica. In: CIVARDI M., FABBRIS L. (eds) *Valutazione della didattica con sistemi computer assisted*, CLEUP, Padova: 179-189.

CIVARDI M., ZAVARRONE E. (2002) Modelli strutturali multigruppo per l’analisi dei tempi di permanenza nel sistema universitario. In: PUGGIONI G. (ed) *Modelli e metodi per l’analisi di rischi sociali e sanitari*, CLEUP, Padova: 1-20.

CROCETTA C., D’OVIDIO F. (2003) La valutazione dell’inserimento lavorativo dei laureati all’Università di Foggia attraverso un’analisi di segmentazione. In: CIVARDI M. (ed) *Transizione Università-Lavoro: la definizione delle competenze*, CLEUP, Padova: 111-132.

CROCETTA C., D’OVIDIO F. (2004) La soddisfazione nella transizione Università-lavoro - Un modello ad equazioni strutturali. In: AURELI CUTILLO E. (ed) *Strategie metodologiche per lo studio della transizione Università-Lavoro*, CLEUP, Padova: 159-187.

DE LEEUW J. (1984) *Canonical Analysis of Categorical Data*, *2nd ed.* DSWO Press, Leiden (NL).

DE LEEUW J., MEULMAN J.J. (1986) Principal component analysis and restricted multidimensional scaling. In: GAUL W., SCHADER M. (eds.) *Classification as a Tool of Research*, Amsterdam (NL): 83-96.

DE LEEUW J., YOUNG F.W., TAKANE Y. (1976) Additive structure in qualitative data: an alternative least squares method with optimal scaling features, *Psychometrika*, 41: 471-504.

FABBRIS L. (2003) *Il monitoraggio dell’inserimento occupazionale e professionale dei laureati e dei diplomati dell’Università di Padova*, Quaderno Pharos n.6, Osservatorio sul Mercato Locale del lavoro, CLEUP, Padova.

GORI E., CREMA F., VIDONI D. (2003) Quali prospettive dalla ricerca sulla qualità e l’efficacia della scuola per la costruzione di sistemi di accountability dell’istruzione, *Non Profit*, **1**, Maggioli, Rimini.

GORI E., VITTADINI G., eds (1999) *Qualità e valutazione nei servizi di pubblica utilità*, ETAS, Milano.

HOELTER J.W. (1983) The analysis of covariance structures, *Sociological Methods and Research*, 11: 325-344.

KLINE R. B. (1998) *Principles and Practice of Structural Equation Modeling*, The Guilford Press, New York.

KRUSKAL J.B. (1965) Analysis of factorial experiments by estimating monotone transformations of the data, *Journal of Royal Statistical Society* (B), 27: 251-263.

JÖRESKOG K.G. (1973) A general method for estimating a linear structural equation system. In: GOLDBERGER A.S., DUNCAN O.D. (eds) *Structural Equation Models in the Social Sciences*, Academic Press, New York: 85-112.

JÖRESKOG K.G. (1977) Structural equation models in the social sciences. In: KRISHNAIAH P.R. (ed) *Application of Statistics*, North Holland, Amsterdam: 265-287.

JÖRESKOG K.G., GOLDBERGER A.S. (1975) Estimation of a model with multiple indicators and multiple causes of a single latent variable, *Journal of the American Statistical Association*, 70: 631-639.

LOVAGLIO P.G. (2000) *Modelli con variabili latenti e indicatori di tipo misto*, Doctoral Thesis, Università degli Studi di Trento.

MEULMAN J.J., HEISER W.J. (1999) *Categories 10.0*, SPSS Inc., Chicago.

VITTADINI G. (1999) Analysis of qualitative variables in structural models with unique solutions. In: VICHI M., OPITZ O. (eds) *Classification and Data Analysis: Theory and Application*, Springer Verlag, Heidelberg: 203-210.

WILEY D.E. (1973) The identification problem for structural equation models with unmeasured variables. In: GOLDBERGER A.S., DUNCAN O.D. (eds) *Structural Equation Models in the Social Sciences*, Academic Press, New York: 69-83.

WRIGHT S. (1934) The methods of path coefficients, *Annals of Mathematical Statistics*, **5**: 161-215.

YOUNG F.W., TAKANE Y., DE LEEUW J. (1978) The principal component of mixed measurement level multivariate data: an alternating least squares method with optimal scaling features, *Psychometrika*, **43**: 279-281.

Measurement of University External Effectiveness Based on the Use of the Acquired Skills

Bruno Chiandotto, Silvia Bacci[1]

Statistics Department "G. Parenti", University of Florence, Italy

Summary. In this paper, we analyse the skills used at work, 18 to 30 months from the completion of studies, by the students who graduated at the University of Florence in the year 2000. The aim is pursued by detecting the determinants of the phenomenon with particular attention to the possible differences between study programmes. We performed two analyses: in the first, we identified homogeneous groups of degree programmes and applied a proportional odds (logistic) model for each group and a partial proportional odds model for the whole university. The second analysis was an ordered logistic model with random intercept having two levels of aggregation with the degree types as second-level units.
Keywords: University effectiveness; Skills; Ordered logistic regression, Proportional odds models; Partial proportional odds models; Multilevel models; Cluster analysis.

1. The evaluation of external effectiveness

One of the ways of assessing the quality of the education offered by a university is to evaluate its performance in terms of internal and external efficiency and effectiveness. External effectiveness is the 'capacity' of a university programme to satisfy labour market needs as indicated by the first-employment rate (Chiandotto & Bacci, 2004), the length of time lapse between graduation and employment, the actual usefulness of the qualifications for the work undertaken, the degree to which graduates use at work the skills they have acquired at university, and so on.

---

[1] The idea, the structure and the setting out of the contribution are due to both the authors; the data processing and the estimate of the models have been done by S. Bacci.

In this research, we focused on the use of skills that graduates achieved at university, with particular reference to the capacity of study programmes of creating the competences required by the labour market. For this purpose, after an appropriate aggregation of study programmes, we estimated a logistic regression model for every aggregation of graduates as well as for the entire contingent of the employed graduates. In this latter case, the groups of programmes were given the role of explicative variable. Thereafter, the results obtained were compared with those from the application of a multilevel logistic regression model on the same set of data; in this application, the programmes represented second-level units.

In Section 2, we discuss the phenomenon under study and its possible interactions with both the individual variables and different types of jobs. In Section 3, we analyse the resulting groups, while in Sections 4 and 5 we discuss the predictions and the results of the fitting of a regression model according to study programme and for the whole Florence University, for finding the net effect of the determinants of the achieved skill use. In Sections 6 and 7, we comment on the estimates obtained with a multilevel regression model. At the end, conclusions about the advantages and drawbacks of using a multilevel model for hierarchic models are drawn.

2. The skill use of the University of Florence graduates

In the year 2000, 5245 students graduated from Florence University. Out of these, 4846 had a Master and 399 a Bachelor degree. Of the 3856 graduates we interviewed, 2882 (approximately 75%) resulted to be working at the time of the interview: 1867 (64.8%) were using intensively the skills they had acquired at university, 730 (25.3%) were using it to a fair extent, while the other 285 (9.9%) did not use it at all.

The response distributions are far from being homogeneous with either the study programme[2] or the faculty (Table 1). For instance, the graduates in the Humanities who said their occupation entailed considerable use of the skills acquired at university went from 40.0% (Philosophy) to 65.0% (Modern foreign language and literature), while the graduates in Science ranged from 45.8% (Natural science) to 81.3% (Physics).

A comparison between the faculties revealed great differences. Taking into account the “considerable use of the skills”, there was an 86.7% of graduates from Medicine and an 81.5% of those from Pharmacy.

---

[2] Because of the small numbers of graduates, Political economics has been merged with Economics, Tropical and subtropical agriculture and relative Sciences to Agronomy; moreover, the two courses in Foreign language and literature under Letters and Philosophy and Formation Science have been considered as a single study programme.

Table 1. Use of the university skills at work, by study programme and faculty

Faculty/programme	**High**	*%*	**Fair**	*%*	**Not at all**	*%*	***Graduates***
AGRICULTURE	**44**	***60.3***	**17**	***23.3***	**12**	***16.4***	***73***
Agronomy	11	*55.0*	7	*35.0*	2	*10.0*	*20*
Forestry	17	*63.0*	6	*22.2*	4	*14.8*	*27*
Forestry and environmental st.	11	*61.1*	3	*16.7*	4	*22.2*	*18*
Scientific and technical agronomy	5	*62.5*	1	*12.5*	2	*25.0*	*8*
ARCHITECTURE	**378**	***68.0***	**155**	***27.9***	**23**	***4.1***	***556***
Architecture	378	*68.0*	155	*27.9*	23	*4.1*	*556*
ECONOMICS	**361**	***72.6***	**121**	***24.3***	**15**	***3.0***	***497***
Business economics	15	*83.3*	3	*16.7*	-	-	*18*
Economics	323	*72.3*	111	*24.8*	13	*2.9*	*447*
Actuarial and statistical science	17	*70.8*	5	*20.8*	2	*8.3*	*24*
Statistical and economic science	6	*75.0*	2	*25.0*	-	-	*8*
PHARMACY	**53**	***81.5***	**10**	***15.4***	**2**	***3.1***	***65***
Chemistry and pharmac. techn.	25	*89.3*	3	*10.7*	-	-	*28*
Phaarmacy	28	*75.7*	7	*18.9*	2	*5.4*	*37*
LAW	**120**	***63.2***	**43**	***22.6***	**27**	***14.2***	***190***
Law	120	*63.2*	43	*22.6*	27	*14.2*	*190*
ENGINEERING	**224**	***71.8***	**72**	***23.1***	**16**	***5.1***	***312***
Civil engineering	68	*81.9*	15	*18.1*	-	-	*83*
Electronic engineering	57	*62.0*	27	*29.3*	8	*8.7*	*92*
Mechanical engineering	48	*69.6*	17	*24.6*	4	*5.8*	*69*
Computer engineering	11	*64.7*	6	*35.3*	-	-	*17*
Environmental engineering	31	*81.6*	3	*7.9*	4	*10.5*	*38*
Telecommunication engineering	9	*69.2*	4	*30.8*	-	-	*13*
LETTERS and PHILOSOPHY	**186**	***53.6***	**96**	***27.7***	**65**	***18.7***	***347***
Philosophy	14	*40.0*	11	*31.4*	10	*28.6*	*35*
Letters	97	*52.4*	48	*25.9*	40	*21.6*	*185*
Foreign language and literature	38	*55.1*	24	*34.8*	7	*10.1*	*69*
Modern foreign lang., literature	26	*65.0*	9	*22.5*	5	*12.5*	*40*
History	11	*61.1*	4	*22.2*	3	*16.7*	*18*
MEDICINE	**65**	***86.7***	**9**	***12.0***	**1**	***1.3***	***75***
Medicine and surgery	39	*84.8*	6	*13.0*	1	*2.2*	*46*
Dentistry and dental prosthesis	26	*89.7*	3	*10.3*	-	-	*29*
FORMATION SCIENCE	**208**	***67.3***	**64**	***20.7***	**37**	***12.0***	***309***
Foreign language and literature	10	*62.5*	2	*12.5*	4	*25.0*	*16*
European foreign lang. and liter.	8	*61.5*	3	*23.1*	2	*15.4*	*13*
Humanities	14	*60.9*	4	*17.4*	5	*21.7*	*23*
Pedagogy	23	*57.5*	11	*27.5*	6	*15.0*	*40*
Psychology	28	*62.2*	12	*26.7*	5	*11.1*	*45*
Educational science	125	*72.7*	32	*18.6*	15	*8.7*	*172*
POLITICAL SCIENCE	**85**	***36.0***	**100**	***42.4***	**51**	***21.6***	***236***
Political science	85	*36.0*	100	*42.4*	51	*21.6*	*236*

Faculty/programme	**High**	***%***	**Fair**	***%***	**Not at all**	***%***	***Graduates***
SCIENCE	**143**	***64.4***	**43**	***19.4***	**36**	***16.2***	***222***
Chemistry	27	*69.2*	6	*15.4*	6	*15.4*	*39*
Physics	13	*81.3*	2	*12.5*	1	*6.3*	*16*
Mathematics	12	*48.0*	10	*40.0*	3	*12.0*	*25*
Biology	48	*64.9*	12	*16.2*	14	*18.9*	*74*
Information science	7	*58.3*	5	*41.7*	-	-	*12*
Geology	25	*78.1*	4	*12.5*	3	*9.4*	*32*
Natural science	11	*45.8*	4	*16.7*	9	*37.5*	*24*
Total	**1,867**	***64.8***	**730**	***25.3***	**285**	***9.9***	***2,882***

On the other hand, the majority of those who said that they were not able to make any use of the acquired skills had graduated from Political Science (21.6%), Letters and Philosophy (18.7%), Agriculture (16.4%) and Sciences (16.2%). Overall, the qualifications that corresponded to the widest use of skills are the ‘technical’ ones, while the general, non-technical, qualifications (such as Political Science) encountered greater difficulties in job searching.

The high levels of Cramer’s V and Chi-square statistics (0.236 and 299.329, respectively; $p<0.0001$) point out a significant difference between the type of studies undertaken and use of the skills: that is why we examined the influence exerted by the specific study programmes.

There are factors other than the type of degree that can influence the use of the acquired skills at work (Chiandotto *et al.*, 2004). The highest percentages of graduates who use the acquired skills are employed in small firms, have taken at least one post-graduate programme and either have a managerial position or are self-employed.

Moreover, the skill use is positively correlated with both job satisfaction and the perceived usefulness of the qualifications for the work duties.

3. Response variable and cluster analysis

We applied an ordinal logistic regression model. The response variable was *the use at work of the skills acquired at university* on a three level scale: high (Y=1), fair (Y=2), not at all (Y=3).

Since our aim was to assess the effect exerted by the study programme on the response variable, the programmes have been aggregated in order to estimate a model for every group detected. Another model was estimated for the entire contingent of graduates with the programmes as an esplicative variable[3].

[3] From a theoretic point of view, these analyses could have been performed by study programme rather than groups of programmes. We grouped the data because of the small number of graduates in several courses and because of the large number of programmes.

Due to the differences between study programmes within the same faculty (Table 1), we grouped them with a cluster analysis (Chiandotto, 1978). We adopted the hierarchic grouping method based on minimum variance (Ward's approach)[4]. The variables were selected according to their explicative capacity between programmes[5]: median age at graduation, median final marks on degree, median finals marks at secondary school, % of males, % of graduates who attended an internship to complete the studies, % of graduates who had work experience during university studies, and % of graduates coming from classical or scientific high schools.

The cluster analysis formed four groups of study programmes[6] (Table 2): Group 1 contains mainly programmes of the humanities and formative science and theoretical programmes (Biology and Natural science); Groups 2 and 3 include programmes of a technical nature (Group 2 refers roughly to economics and engineering while Group 3 to health fields); Group 4 corresponds almost entirely to Agriculture.

From the analysis of the aggregation variables, profiles that are more specific emerged (Table 3). In the study programmes of Group 1 there is a prevalence of female graduates (62.4%), who come from classical or scientific high schools (76.6%), had generally low final marks at secondary school (21.7% had marks ranging from 36/60 to 40/60) and took quite a long time to complete their university studies (12.3% were over 6 years over the prescribed time and 26.2% graduated after 30 years).

Group 2 was the only one with a prevalence of males (55.9%) and with the highest percentage of technical school diploma-holders (40.0%). The high school final marks of these graduates were high (25.5% achieved their secondary school diploma with at least 55 out of 60), whereas their graduation marks were the lowest in the university (54.8% graduated with less than 100/110 and only 14.2% received a 110/110 degree cum laude[7]) and the time taken to complete their studies was longer (25.6% took more than twice the programme duration).

The majority of graduates in Group 3 came from classical or scientific high schools (82.4%), with high final marks at school (38.6% had marks over 55/60) and with less previous work experience than the other groups (only

---

4 Ward's method was chosen because the other grouping methods (complete link, median and centroid methods) gave unsatisfactory results in terms of distribution of the study programmes among groups: there was a tendency to form a principal group containing programmes heterogeneous with one another and of several small groups.

5 The grouping variables have been standardized (zero mean and unit variance) because of different units of measurement.

6 We partitioned into four groups because a finer aggregation would have generated problems in the maximization of the likelihood function for the estimation of the model, especially for groups containing very few individuals.

7 In the other three groups, the same frequency did not exceed 21.6% (Group 3), in the first case and at least 34.6% (Group 3) in the second case.

45.1% vs. 70-80%). However, 34.6% of them attended an internship in order to complete their studies (compared with the 7-8% of those in groups 1 and 2). The graduates in this group took the shortest length of time to finish their studies: almost 25% of them graduated within the established time.

Group 4 showed a prevalence of females (69.8%) and subjects from technical schools (38.8%), with low final marks at school (26.4% had marks under 41/60). However, 93.8% of them had attended an internship and 58.9% graduated within less than one and a half times the normal time.

Table 2. Composition of the study programme groups

GROUP 1	GROUP 2	GROUP 3	GROUP 4
Architecture	Business economics	Chemistry	Agronomy
Philosophy	Economics	Chem.+pharm. techn.	Forestry
Letters	Law	Pharmacy	Educational science
Foreign language & lit.	Civil engineering	Physics	Scient. techn. agronomy
Modern foreign lang. lit.	Electronic engineer.	Medicine and surgery	Forestry and environ. st.
European foreign lang. l.	Computer engineer.	Dentistry dental pros.	
Humanities	Mechanical engin.	Environmental engin.	
Pedagogy	Mathematics	Telecommunic. eng.	
Psychology	Information science		
Biology	Geology		
Natural science	Political science		
Actuarial and statistics.			
Statistics and economics			
History			

Table 3. Profiles of the study programme groups (percentage values)

	GROUP 1	GROUP 2	GROUP 3	GROUP 4
Female	62.4	44.1	51.0	69.8
High school: *Lyceum*	76.6	60.0	82.4	61.2
Diploma mark >55/60	19.0	25.5	38.6	13.2
Graduates in legal time	6.2	2.3	24.2	13.2
Degree mark <100/110	11.0	54.8	21.6	8.5
Work during university	74.0	73.2	45.1	83.0
Internship	7.1	8.2	34.6	93.8
Graduates (n)	*1170*	*1221*	*246*	*245*

4. Independent variables

The models for investigating the use of skills were estimated on 2882 newly graduated subjects in the year 2000 at Florence University who were employed at the time of interview. The explanatory variables were selected according to the descriptive analyses (see Section 2) and to our prior knowledge.

The following covariates were included at least in one of the initial steps:

<u>*Quantitative explicative variables*</u>

1. *Graduation marks* (*gradmark*). It is the mark at graduation that ranges from 66 to 113 over 110 (113 corresponds to 110 cum laude). Since the variables referring to the graduation mark and to the mean of the marks achieved in the examinations are strongly correlated (r=0.84), only the first variable was included in the model to avoid multicollinearity.
2. *Age at graduation* (*agegrad*), ranging from 21.9 years to 67.6 years.
3. *Final secondary school marks* (*finmark*), ranging from 36/60 to 60/60.
4. *Number of hours worked per week* (*num_hrs*) ranging from 2 to 90; the highest frequency corresponded to 30, 36, 40, 45 and 50 hours per week.

<u>*Qualitative explicative variables*</u>

5. *Gender* (*gend*): it is a binary variable with categories ‘female’ (*female*) and a ‘male’ (*male*). Since the female graduates were 56.4%, ‘female’ was the reference category.
6. *Work experience during studies* (*work1*): this is a binary variable (*work1_no* if the graduate did not have work experience during his/her studies and *work1_yes* otherwise). Even though the majority of graduates had had experience of work during their studies, the ‘no’ category was chosen as reference in order to directly interpret the estimated coefficient.
7. *Attendance of an internship or training programme for completion of the studies* (*stage*). It is a binary variable with categories ‘no-internship’ (*stage_no*) and an ‘internship done’ (*stage_yes*). More than 83% of the graduates did not attend an internship.
8. *Type of high school diploma* (*typdip*). Its categories are: ‘classical’ (*classics*), ‘scientific’ (*scientific*), ‘technical’ (*technical*), ‘other’ (*othdip*). The ‘scientific’ category was the reference.
9. *Social class background* (*p_socgen*). Its categories are: ‘upper middle class’ (*upper* – adopted as reference), ‘white-collar middle class’ (*white-collar*), ‘lower middle class’ (*low_mid*), ‘working class’ (*working*).
10. *Knowledge of English* (*lang1_gb*). Its categories are: ‘non/poor’ (*engl_poor*), ‘sufficient’ (*engl_suff*), ‘good’ (*engl_good*), ‘excellent’ (*engl_excel*). The majority of graduates had a good knowledge of English.

11. *Knowledge of word processors* (*info3_wp*). Its categories are the same as those for *langl_gb*; the most frequent was that referring to a good knowledge of word processors (*wp_good*).
12. *Geographical area of work site* (*site_work*). Its categories are 'Florence and Province', 'North and abroad', 'Centre', 'South'; the 'Florence and Province' category was taken as reference. The work site was considered more interesting for analytical purposes than that of the area of residence of graduates, since the latter does not always coincide with the work site.
13. *Occupation at graduation* (*occgrad*). Its categories are: already working at the time of graduation and still the same work, working at the time of graduation and changed work, not working at the time of graduation (this last was taken as reference).
14. *Type of work* (*typ_work*) with categories, with categories: 'employee' (reference mode – *employ_work*) and self-employed' (*self_work*).
15. *Type of contract* (*typ_contr*) with categories 'tenure employment' (mode used as reference) from 'precarious'; this latter is comprehensive of all types of contracts limited in time (such as contracts for collaboration, training, etc.).
16. *Economical field of business* (*econbus*): with categories 'public' (*public*) and 'private' (*private*); the latter was taken as reference.
17. *Professional position of the graduate* (*prof_pos*): the original categories were 'manager', 'employee', 'free-lance', 'self-employed worker' and 'other'. When the model was being estimated, this variable turned out to be somewhat significant though with just some categories, so dichotomization was resorted to by distinguishing employees and workers on one side (mode taken as reference – *employ/worker*) and managers, free-lancers and self-employed workers (*man/free*) on the other, in order to separate 'lower' from 'higher' professional positions.
18. *Post-graduate qualification accomplished* (*post_grad*): it is a binary variable aimed at assessing the effect of post-graduate studies on the quality of job.
19. *Size of the firm*[8] (*size_firm*): a distinction was made between 'small/medium firm' (*small_firm* – max. 50 employees) and 'large firm' (*large_firm* – over 50 employees)
20. *Qualification necessity* (*necess*): a distinction was made between the qualifications required by law (*qualif_law*), qualification actually useful (even if not compulsory by law – *qualif_useful*) and qualification useless for the type of work performed (*qualify_useless*). The 'qualification required by law' was adopted for reference.
21. *Satisfaction in work performed* (*satisf*) with categories: 'great/very great' (*very_satisf*), 'sufficient' (*sufficient_satisfy*), 'little or none at all'

[8] Note that the term 'firm' is intended as workplace in general, and defines both private companies and public entities.

(*little_statisf*) and refers to job satisfaction. Most graduates were very satisfied with their work.

22. *Search for new work* (*search_work*) for which a distinction was made between those already employed but on the lookout for a new position and those already working but not in search of a new position.

5. Proportional odds logistic model and partial proportional odds logistic model

The nature of the variable response (ordinal polytomous with three categories) and the results of the *score* test for evaluating the hypothesis of *proportional odds* suggested the use of an ordinal logistic model with proportional odds for each one of the four groups of degree programmes, with the following general structure expressed in logit terms (McCullagh & Nelder, 1989: chapter 5)

$$\begin{cases} \operatorname{logit}(P_{1i}) = \log\left(\dfrac{P_{1i}}{1-P_{1i}}\right) = \log\left(\dfrac{P_{1i}}{P_{23i}}\right) = \alpha_1 + \sum_{j=1}^{h} \beta_j \cdot x_{ji} \\ \operatorname{logit}(P_{12i}) = \log\left(\dfrac{P_{12i}}{1-P_{12i}}\right) = \log\left(\dfrac{P_{12i}}{P_{3i}}\right) = \alpha_2 + \sum_{j=1}^{h} \beta_j \cdot x_{ji} \end{cases}$$

where: i indicates a generic graduate, $Pr(Y=1)=P_1$, $Pr(Y=2)=P_2$, $Pr(Y=3)=P_3$, $P_1+P_2=P_{12}$, $P_2+P_3=P_{23}$, α_1 and α_2 the intercepts of the two logit models, x_j is the j^{th} explicative variable and β_j is the corresponding regression coefficient; in this case, the covariate number h is equal to *22*.

A logistic model for ordinal data is a set of logistic models for binary data as the number of categories of the Y response variable minus one. In the same way as for the dichotomous case and in the presumption of proportional *odds*, the β_j coefficient of regression can be interpreted as the increase (if positive), or as the decrease (if negative), that the *logits* undergo with a unit increase of the x_j predictor. With a 3-mode Y, e^{β_j} indicates both the *odds* of P_1 compared to P_{23} and the *odds* of P_{12} compared to P_3.

We estimated the four models with the PROC LOGISTIC of SAS software, using the Fisher-scoring maximization algorithm; the explicative variables and any quadratic or interaction effect were selected with the *forward* procedure[9]. To assess the goodness of fit, we adopted the likelihood and score tests and computed Pearson’s statistics and deviance, and R-square[10]. The significance

---

[9] We applied also stepwise and backward selection procedures. In the case of Groups 1, 2 and 4 they produced identical results. Vice versa, in the model referring to Group 3 there were problems of fitting and convergence of the maximization algorithm.

[10] For non-linear models, the R^2 very seldom reaches high values; hence, a 20-25% quota of explained variance may be satisfactory.

of the coefficients was evaluated with Wald's tests[11].

Table 4 shows the list of selected covariates, the estimates of the intercepts and coefficients of regression; together with the relative standard errors and the results of Wald's test (significance was 10%). The reference category is labelled '0'.

The last column shows the odds ratio, i.e. the variation of P_1 with respect to P_{23} or, similarly, of P_{12} with respect to P_3, with a unit increase of an explicative variable.

Table 4. Proportional odds ordered logistic models, for every group

Effect	*Reference*	*Estimate*	*Standard Error*	*Value t*	*p-value*	*Odds ratio*
Group 1 (R^2=0,27)						
Intercept 1	-	1.9775	0.1885	110.0778	<.0001	-
Intercept 2	-	3.8092	0.2607	213.5712	<.0001	-
qualif _useful	*qualif_law*	-0.8117	0.2261	12.8918	0.0003	0.444
qualif _useless	*qualif_law*	-3.3951	0.3536	92.1649	<.0001	0.034
search_work_yes	*search_work_no*	-0.6415	0.2512	6.5232	0.0106	0.526
sufficient_satisfy	*very_ satisfy*	-0.6897	0.2279	9.1625	0.0025	0.502
little_ satisfy	*very _ satisfy*	-1.4907	0.3642	16.7531	<.0001	0.225
Group 2 (R^2=0,22)						
Intercept 1	-	-2.4672	1.2882	3.6680	0.0555	-
Intercept 2	-	-0.2560	1.2856	0.0397	0.8422	-
post_grad_no	*post_grad_yes*	-0.3506	0.1675	4.3810	0.0363	0.704
man/free	*employ/worker*	0.4196	0.2061	4.1440	0.0418	1.521
large_firm	*small_firm*	-0.4888	0.1797	7.3975	0.0065	0.613
qualif _useful	*qualif_law*	-1.1002	0.1932	32.4162	<.0001	0.333
qualif _useless	*qualif_law*	-3.0929	0.3504	77.9196	<.0001	0.045
sufficient_satisfy	*very_ satisfy*	-0.6967	0.1831	14.4867	0.0001	0.498
little_ satisfy	*very _ satisfy*	-1.3932	0.3213	18.8010	<.0001	0.248
gradmark	-	0.0463	0.0124	13.8852	0.0002	1.047
Group 3 (R^2=0,20)						
Intercept 1	-	8.3295	2.4155	11.8907	0.0006	-
Intercept 2	-	9.8271	2.4746	15.7702	<.0001	-
post_grad_no	*post_grad_yes*	-1.0962	0.5044	4.7233	0.0298	0.334
self_work	*employ_work*	2.7989	1.2444	5.0594	0.0245	16.427
little_ satisfy	*very _ satisfy*	-1.2445	0.5154	5.8305	0.0158	0.288
public	*private*	1.8586	0.5388	11.8975	0.0006	6.415
agegrad		-0.2624	0.0896	8.5806	0.0034	0.769

[11] The former one was applied to assess the overall significance of each discrete polytomous covariate; the latter was used for evaluating the significance of the individual estimated coefficients of regression.

Effect	*Reference*	*Estimate*	*Standard Error*	*Value t*	*p-value*	*Odds ratio*
Group 4 (R^2=0,26)						
Intercept 1	-	2.2181	0.5576	15.8256	<.0001	-
Intercept 2	-	3.3970	0.6229	29.7390	<.0001	-
post_grad_no	*post_grad_yes*	-1.3308	0.4809	7.6576	0.0057	0.264
man/free	*employ/worker*	1.0200	0.4964	4.2222	0.0399	2.773
qualif _useful	*qualif_law*	-0.8744	0.5325	2.6961	0.1006	0.417
qualif _useless	*qualif_law*	-3.7764	0.8838	18.2576	<.0001	0.023

The results highlight that some covariates are significant in three of the four groups: the usefulness of the qualifications with respect to the type of job (Groups 1, 2 and 4), the level of job satisfaction (Groups 1, 2 and 3), and the accomplishment of at least one post-graduate internship/training programme (Groups 2, 3 and 4). Even the probability of an intense use of skills is positively influenced by the usefulness of graduates' qualifications, job satisfaction and accomplishment of at least one post-graduate internship/training programme.

Each group model is characterized by the presence of specific covariates. In Group 1, the probability that graduates on the lookout for a new job use their skills to a great extent are half the probabilities of their colleagues who did not expect to change activity; therefore, a poor job satisfaction stimulates young graduates to search for new jobs.

In Group 2, the professional role of graduates, the size of the firm and the graduation marks determine the use of skills. Those employed in small firms, with higher responsibilities, with high marks at graduation, have greater probability of using their university skills to a higher level than those with lower graduation marks who work in clerical or waged positions in large companies[12].

In Group 3, the self-employed and those working in public corporations show higher odds ratios than those working as employees and in private concerns; moreover, as the age is more advanced at graduation, the lower the probability is to use the expertise acquired at the university.

In Group 4, the only additional variable with regard to attendance in post-graduate internships/training programmes and the usefulness of the relative qualification is the professional status of the graduate: managers and free-lancers show higher odds ratios than clerical and waged workers.

With regard to estimation of the model for the whole university, the results of the *score* test have suggested not to accept the hypothesis of proportional

---

[12] The effect of this covariate was very weak, so confirming the results of descriptive analyses that the final marks poorly discriminated among graduates because of a concentration of high values.

odds for the covariate represented by the groups of degree programmes. Hence, the model applied is a *partial proportional odds logistic model*, whose general structure is as follows:

$$\begin{cases} \operatorname{logit}(P_{1i}) = \log\left(\dfrac{P_{1i}}{1-P_{1i}}\right) = \log\left(\dfrac{P_{1i}}{P_{23i}}\right) = \alpha_1 + \sum_{j=1}^{h} \beta_j \cdot x_{ji} + \beta_{1,h+1} \cdot x_{h+1,i} \\ \operatorname{logit}(P_{12i}) = \log\left(\dfrac{P_{12i}}{1-P_{12i}}\right) = \log\left(\dfrac{P_{12i}}{P_{3i}}\right) = \alpha_2 + \sum_{j=1}^{h} \beta_j \cdot x_{ji} + \beta_{2,h+1} \cdot x_{h+1,i} \end{cases}$$

where x_{h+1} indicates the explicative variable (the *cluster* variable, in this case) for which the proportional odds hypothesis is not valid and for which $\beta_{1,h+1}$ and $\beta_{2,h+1}$ are the x_{h+1} coefficients of the first and second logit, respectively. In other words, the partial proportional odds model takes on proportional odds for only certain explicative variables and not for others.

Consequently, the two logits that compare $Y = 1$ with $(Y = 2) \cup (Y = 3)$ in the first case, and $(Y = 1) \cup (Y = 2)$ with $Y = 3$ in the second, are distinguishable not only because of the intercept but also of the value assumed by the regression coefficients for which no proportional odds had been assumed. Since Y assumes 3 categories, $e^{\beta_{1,h+1}}$ indicates the P_1 variation with respect to P_{23} with a, unitary increase x_{h+1}, while $e^{\beta_{2,h+1}}$ indicates the same variation that P_{12} undergoes with respect to P_3.

Once the significant variables have been selected and each of them submitted to the hypothesis of proportional odds, the final model for the whole university is estimated with the PROC GENMOD application of the SAS software, using Newton Raphson's maximization algorithm.

Table 5. Partial proportional odds ordered logistic models, for every group

Effect	*Reference*	*Estimate* P_{12}/P_3	*Estimate* P_1/P_{23}	*S.E.* P_{12}/P_3	*S.E.* P_1/P_{23}	*Odds ratio* P_{12}/P_3	*Odds ratio* P_1/P_{23}
Intercept	-	-0.9734		1.0523		-	
post_grad_no	*Post_grad_yes*	-0.2461		0.1075		0.782	
man/free	*employ/worker*	0.4661		0.1232		1.594	
large_firm	*small_firm*	-0.3375		0.1126		0.714	
qualif _useful	*qualif_law*	-3.2424		0.2072		0.353	
qualif _useless	*qualif_law*	-1.0422		0.1219		0.039	
sufficient_satisfy	*Very_ satisfy*	-1.3017		0.1984		0.510	
little_ satisfy	*very _ satisfy*	-0.6732		0.1155		0.272	
agegrad	-	0.0410		0.0091		1.042	
Group 1	Group 2	-0.7510	-0.3904	0.2415	0.2683	0.472	0.677
Group 1	Group 3	0.0710	-0.4736	0.4187	0.4879	1.074	0.623
Group 1	Group 4	0.3125	-0.4561	0.3511	0.4335	1.367	0.634
Group 2	Group 3	0.8220	-0.0833	0.4156	0.4782	2.275	0.920
Group 2	Group 4	1.0635	0.1249	0.3531	0.4228	2.896	1.133
Group 3	Group 4	0.2415	0.0935	0.4893	0.5854	1.273	1.098

In Table 5, the selected covariates are listed with an indication of the effect compared to the category assumed as reference, followed by the estimated regression coefficients, the relative standard errors and the odds ratios[13]. Due to the non applicability of the hypothesis of proportional odds for the *cluster* variable, the estimates are given for both logits, that is, for P_{12}/P_3 and P_1/P_{23}.

The determinants of the use of skills are the type of study programme, the attendance of post-graduate internships/training programmes, the professional position, the size of the company, the use of the qualifications achieved, the satisfaction for the job, and the marks obtained upon graduation. The net effect exerted by the determinants of the proportional odds is similar to the effects revealed with by the analysis of the groups, even if covariates that were significant for single groups of study programmes (age at graduation, field of work, type of work, and search for a new job) were ignored in the general model.

The probability that a graduate belonging to Group 1 uses his/her skills to a high extent is approximately 33-37% less than that of colleagues belonging to the other groups; in contrast, the odds ratio for a graduate in Group 3 is 8.0% higher than that of graduates belonging to the other groups. The graduates belonging to Groups 2 have more of an advantage over Group 4 (P_1/P_{23} = 1,133) in job-hunting that conforms to the expertise acquired at university. However, both have higher odds ratios than their colleagues in Group 1 but lower than those in Group 3[14].

6. Two-level random intercept model: a comparison

The hierarchical structure of the data led to the adoption of a multilevel ordinal logistic model of regression in which the first level units were 2,882 employed graduates, whereas the second level units were 38 study programmes.

The explicative variables of the previous analysis were the first-level covariates, whereas the grouping variables of the study programmes served as

---

13 The *p*-values do not exceed 2.2%. Though the *cluster* variable (*p*-value 0.0019) showed overall significance, some of the comparisons between groups were not significant. This problem could have been avoided with a further aggregation of the groups, but we preferred to present the results so to analyse the four original groups.

14 The interpretation of the odds ratios that compare $(Y=1)\cup(Y=2)$ with $Y=3$, that is, P_{12}/P_3, is not so clear. The focus on one of the two odds ratios depends on the relevance given to the $Y=2$ intermediate response category: in the case P_1/P_{23} it is cumulated with the $Y=3$ mode, and has therefore a negative meaning, while in the other case P_{12}/P_3 it is added with the $Y=1$ mode, and is consequently positive. In our analysis, a negative meaning was considered more appropriate for $Y=2$, since this category represents a clear stand taken by the subject interviewed with regard to the questions asked, whereas the aim of our research is to pinpoint the study programmes characterized by their capacity to create skills useful for the labour market.

second-level covariates. We chose the model by estimating[15] the null model (that is, without explanatory variables) in order to assess the significance of the second-level residual variance. Through a stepwise procedure, the significant first and second-level covariates were selected as well as any quadratic terms and effects of interaction. A random intercept model with two levels of aggregation *(proportional odds*, see: Fielding *et al.*, 2003; Goldstein, 2003; Hedeker, 2004) was obtained.

The general structure for the k^{th} generic programme is the following:

$$\begin{cases} \log\left(\dfrac{P_{1ik}}{P_{23ik}}\right) = \log\left(\dfrac{P(Y_{ik} \le 1)}{P(Y_{ik} > 1)}\right) = \alpha_{1k} + \sum_{j=1}^{k} \beta_j \cdot x_{jik} \\ \log\left(\dfrac{P_{12ik}}{P_{3ik}}\right) = \log\left(\dfrac{P(Y_{ik} \le 2)}{P(Y_{ik} > 2)}\right) = \alpha_{2k} + \sum_{j=1}^{k} \beta_j \cdot x_{jik} \end{cases}$$

Since the α_{1k} and α_{2k} intercepts vary according to the degree programme,

$$\alpha_{1k} = \gamma_1 + U_{0k} \text{ and } \alpha_{2k} = \gamma_2 + U_{0k},$$

the model takes on the following form:

$$\begin{cases} \operatorname{logit}(P_{1ik}) = \log\left(\dfrac{P_{1ik}}{1-P_{1ik}}\right) = \log\left(\dfrac{P_{1ik}}{P_{23ik}}\right) = \gamma_1 + \sum_{j=1}^{h} \beta_j \cdot x_{jik} + U_{0k} \\ \operatorname{logit}(P_{12ik}) = \log\left(\dfrac{P_{12ik}}{1-P_{12ik}}\right) = \log\left(\dfrac{P_{12ik}}{P_{3ik}}\right) = \gamma_2 + \sum_{j=1}^{h} \beta_j \cdot x_{jik} + U_{0k} \end{cases}$$

where U_{0k} indicates the residual second-level component corresponding to the k-th group; it is assumed that the second-level residuals are normally distributed with zero mean and constant variance.

Because of the failure in the convergence of the log-likelihood algorithm, in some cases, and due to the non-significance of the estimates obtained in others, the estimated model has no random coefficients. The main consequence is that the net effect exerted by the explanatory variables for each study programme cannot be distinguished. On the other hand, our analysis allows the estimation of coefficients and the consequent detection of groups of covariates for every group of degree programmes.

The estimated model (Table 6) showed significant variability in the second-level residuals, and the adoption of a multilevel analysis is so justified. No second-level covariate resulted to be significant, probably because of the aggregation process of these variables.

With regard to the first-level covariates selected, five out of seven were also in the partial proportional odds logistic model and showed similar net effects.

---

[15] The model was estimated via the PROC NLMIXED system of SAS software, by means of maximization of the log-likelihood function using the Dual Quasi-Newton algorithm and the Gaussian adaptive quadrature method.

Table 6. Two-level random intercept ordered logistic model (proportional odds)

Effect	*Reference*	*Estimate*	*Standard Error*	*Value t*	*p-value*	*Odds ratio*
FIXED EFFECTS						
Intercept 1	-	-0.0863	0.9022	0.10	0.9243	--
Threshold	-	-2.1287	0.0804	26.47	<0.0001	--
male	*Female*	-0.1962	0.0948	2.07	0.0453	0.8218
self_work	*employ_work*	-0.5353	0.1552	3.45	0.0014	0.5855
post_grad_no	*post_grad_yes*	-0.2008	0.0888	-2.26	0.0296	0.8181
man/free	*employ/worker*	0.7452	0.1460	-5.11	<0.0001	2.1069
qualif _useful	*qualif_law*	-1.0146	0.0994	10.20	<0.0001	0.3625
qualif _useless	*qualif_law*	-3.1033	0.1726	17.98	<0.0001	0.0449
sufficient_satisfy	*very_satisfy*	-0.7012	0.0948	7.40	<0.0001	0.4960
little_ satisfy	*very _satisfy*	-1.3428	0.1540	8.72	<0.0001	0.2611
gradmark	-	0.0172	0.0086	-2.01	0.0521	1.0174
RANDOM EFFECTS						
St.Dev. (U_{0k})	-	0.2904	0.06473	4.49	<0.0001	--

However, in contrast to what emerged in the previous analysis, the size of the firm was no longer significant; nonetheless, both the gender of the graduate and the self-employment activity became part of the model.

The probability of using skills to a wider extent resulted to be 18% greater among male graduates than female; the same probability was 42% less in the case of self-employed subjects as compared to employed colleagues. This relationship contradicts what emerged for the same variable in the logistic model referred to Group 3 (Table 4).

7. Conclusions

The two approaches adopted for detecting the determinants of the use of skills acquired by young graduates at university and for calculating their relative net effect led to similar conclusions. The higher the graduation mark, the greater the graduate's chance of finding a suitable position, and the more the graduate's qualification was useful for job-hunting, the more he/she may be satisfied with his/her job. Moreover, if he/she has accomplished at least one post-graduate internship/training programme and taken up a position entailing a certain degree of responsibilities (manager or free-lance) it is likely that he/she will be assigned tasks that require a satisfactory use of his/her skills.

Over and above these undoubtedly important conclusions, the analysis of groups permits more exhaustive considerations than those from the multilevel analysis. In the group of non-technical graduates, the search for a new job may be explained by the particularly poor usage of the university skills. The size of the firm and the professional position occupied are critical for the graduates in Engineering, Economics and Law who can choose between self-employment and employment. Lastly, physicians, dentists and chemists, who can be either self-employed or employed in public or private fields, is consistent with the type of work (self-employed or employed) and with the field of work (public or private).

The study programmes in the humanities and formative sciences are associated with a poor use of the skills acquired at university, those related to health are right the opposite, and technical study programmes are in the middle. We may point out a lack of coherence in the classification of some study programmes: the statistical programmes are classified in the first general group and the Educational science together the health group.

To conclude, even if the multilevel approach has a theoretical structure that takes into consideration the hierarchy of the data on graduates, it does not contribute significantly to the analysis of the phenomenon. Contrariwise, the mathematical complexity of the model and the optimisation algorithms, make it extremely difficult to estimate the random coefficients, which are essential for a comprehension of the effects of the second-level units (degree programmes) on the criterion variable.

References

CHIANDOTTO B., BACCI S. (2004) Un modello multilivello per l'analisi della condizione occupazionale dei laureati. In: CROCETTA C. (ed) *Modelli statistici per l'analisi della transizione Università-lavoro*, CLEUP, Padova: 211-234.

CHIANDOTTO B., BACCI S., BERTACCINI B. (2004) *I laureati e diplomati dell'Ateneo Fiorentino dell'anno 2000: profilo e sbocchi professionali*, Università degli Studi di Firenze, Firenze.

CHIANDOTTO B. (1978) L'analisi dei gruppi. Una metodologia per lo studio del comportamento elettorale (parte I), *Quaderni dell'Osservatorio Elettorale*, **4**.

FIELDING A., YANG M., GOLDSTEIN H. (2003) Multilevel ordinal models for examination grades, *Statistical modelling*, **3(2)**: 127 - 153.

GOLDSTEIN H. (2003) *Multilevel Statistical Models*, Arnold Publishers, London.

HEDEKER D. (2004) *Multilevel Models for Ordinal and Nominal Variables*, (http://tigger.uic.edu/~hedeker/ml.html)

MCCULLAGH P., NELDER J.A. (1989) *Generalized Linear Models*, Chapman and Hall, London.

Graduates’ Job Quality Dimensions According to a Delphi-Shang Experiment

Luigi Fabbris, Maria Cristiana Martini[1]

Statistics Department, Padua University, Italy

Summary. In this paper, we discuss the salient information drawn from a Delphi experiment (Shang version) on some Italian job market issues realised by an email interview of a panel of experts. We define firstly the job quality dimensions of newly hired graduates and then compare it with the possible situation of graduates at the end of their careers. The dimensions are compared with a multivariate statistical analysis on the relationships between the satisfaction perceived by the Paduan graduates for their own job, and some personal and job characteristics. Such an evaluation may suggest new criteria for a future survey on “external effectiveness” of university education based on graduates’ reports. The dimensions of initial and end-of-career quality of graduates are correlated to the experts’ concepts through a semantic differential analysis.
Keywords: Job quality; External effectiveness of teaching; Delphi method; Shang method; Semantic differential analysis; Multiple regression analysis; University of Padua.

1. Aims and methodology of Delphi experiment

«This way, it will be required to have a university degree to work as a street-sweeper». This sentence was typical in the Sixties when, interrupting the long-lasting elite access to university, masses of youngsters entered the Italian university. The apparently sceptical motto meant that Italy was starting a process of liberalisation of the tertiary educational level, and individual educational

[1] The two authors jointly designed and realised the work described in this paper. However, M.C. Martini has edited Sections 3, 4 and 5 of the paper and L. Fabbris the others.

choices were going to be determined by the job market more than by family wealth.

So many things have changed since then. The birth rate almost halved, the educational level increased rapidly, and about three quarters of young people have a high school degree. After the Italian educational reform, determined by Decree no. 509 of 1999, the enrolment in university steeply increased, so that in the academic year 2004/05, the ratio between the number of university re-freshers and the number of people in the age of 19 was 60%.

It is likely that in a few years about half the young people entering the Italian job market will possess a university degree. Right now, the proportion of graduates searching for a job exceeds the proportion of technicians, intermediate, executive and professional workers in the Italian private and public economic sectors. Cappa & Fabbris (2004) showed that several graduates already accepted general clerical and other positions that before were rejected as initial jobs.

Hence, the quality of initial jobs is the key issue to define an acceptable job for a graduate.

A job is of good quality if it satisfies both the worker's and the employer's requirements. Job quality is a multidimensional concept that varies in time, space and among clusters of people.

In the following, we will represent the job quality at the insertion stage and privilege the viewpoint of graduates. In fact, we compare the experts' perceptions collected through a Delphi survey – Shang version, with the graduates' ones collected by means of a specific survey.

We selected the 12 experts among people who covered top positions in various cultural and professional Italian bodies specialised in labour market issues (academy, employers' and workers' associations, operators for guidance and job matching). Three sets of questions progressively focusing on job quality aspects were administered to the experts by email.

The first set of questions focused on the selection of quality dimensions. The experts were asked to define the three most important dimensions of job quality at graduate's recruitment-stage. The second set of questions related to the assignment of percentages of importance (with a 100 sum) to the five selected dimensions (see Section 2). The third to the importance of job quality dimensions at the top-of-career stage (Section 3). Then, the experts were made aware of the mean parameters of the obtained distribution and asked, if they wanted, to change their importance grades. This procedure can be considered a Delphi method, Shang version (Ford, 1975).

After the request of the last set of opinions, the experts were asked, through 20 dichotomous items, to state their knowledge and their perception of the characteristics of job market for the new graduates. We applied a semantic differential method on the obtained answers (Section 4).

The distribution of the obtained answers was compared with the analysis of data independently obtained from employed graduates of the University of

Padua surveyed six months and eighteen months after graduation. The data related to satisfaction with various aspects of their occupation.

The two informative approaches are radically different: the former is based on conjectures of 'third parties' about the relevance of selected variables for graduates' work; the latter is based on perception of employed graduates for aspects of their job referred more or less to the same time reference.

The comparison between the two viewpoints will highlight discrepancies between conjectured and practiced job dimensions and may suggest contents for a new survey on graduates' satisfaction with their job. The possible suggestions for a survey on early graduates' jobs are described in Section 5.

2. Dimensions of graduates' job quality

The recruited experts highlighted – in agreement with Touraine (1955) – the following dimensions:

- *Work environment and worker's logistics.* This dimension regards apparently heterogeneous features: easiness to reach the work site, quality (comfort, aesthetic) of physical spaces, adequacy of the available information, tools, methodology and technology[2] for work, quality of relational environment (interpersonal relationships with colleagues, hierarchical structure and customers, possibility of lifelong learning and working together with experienced colleagues and managers), interaction with other companies. Our experts weighed the importance of work environment and individual logistics 15.3%, with a certain variability of opinions (coefficient of variation 33%, see Table 1).
- *Social prestige.* This dimension is the mirror image of the job's relevance, as perceived by graduate's relatives, friends and other social microcosmos. The dimension is considered by experts less important than others as for initial jobs (weight: 12.7%), since initial jobs are transient and most of them are obscure to people who do not belong to work organisations similar to the graduate's one.
- *Internal acknowledgment of competences.* This dimension may be decomposed in two sub-dimensions: value given to economic features (income) and professional features (type of job and qualification, use of high-level competences). Experts rated this dimension 20.4%, mainly as a function of the relevance of professional acknowledgment of graduate's work.
- *Chances of professional enhancement.* As the previous dimension, it is structured as the union of economic and professional sub-dimensions, but in a dynamic, instead than cross-sectional, perspective. Specifically, those

[2] For the ambivalent role of technology as a parameter of job quality for intellectual workers see Lutz (1992).

Table 1. Average score and standard deviations of graduates' job quality dimensions and compositional aspects drawn from a Delphi survey

Dimensions	*% weight*	*Aspects*	*Average score*	*Standard deviation*
Purposive autonomy	22.9			7.5
		Flexibility of working times & modes	7.1	2.0
		Possibility to propose projects	8.5	2.1
Expectation of professional enhancement	29.6			8.4
		Increase professional position	8.7	0.6
		Increase income	7.6	1.1
		Improve competences	8.5	1.3
Internal competence acknowledgement	20.4			7.5
		Economic aspect	7.1	1.6
		Professional position	7.7	1.4
		Competence utilisation	7.7	1.1
Social prestige	12.1			4.0
Work environment and worker logistics	15.0			6.7
		Locations' comfort and aesthetic	7.1	1.0
		Technology, work methodology	7.6	1.0
		Relational environment	8.4	0.8
		Proximity to home, access easiness	5.8	2.6

graduates who feel that their professional potential is initially sacrificed accept a job according to the perception that *in nuce* chances of bettering their initial work position are important[3]. The rate assigned by experts to this dimension is 28.7%, absolutely higher than the relevance of job contract (tenure) and functional qualification. Moreover, the variability among experts is limited (coefficient of variation 22%).

- *Purposive autonomy*. This is the dimension of creativity, freedom of designing one's own work, spirit of responsibility, non-affliction even for a dependent work position. Autonomy relates both to the subjective choice of operational times and modes, to the possibility to propose innovative projects, and to the possibility of realising ever changing functions and applying various kinds of competences. Even this dimension is so fundamental for graduates jobs' qualification, that experts assign to it a 22.9% weight.

The five dimensions could be post-classified according to Herzberg (1959) essential classification criterion:

(a) *Hygienic factors*, i.e. factual features of work, such as physical environment, work organisation, internal social and cultural environment, income and incentives, perspectives of the economic division;

[3] Some psychologists (Nielsen, 1993) call this dimension "empowerment". Evidently, it is like a set of chances to be disbursed during the professional life.

(b) *Motivational factors,* which have roots into human superior needs: self-realization, social acknowledgement, and freedom in own work realization. The motivational factors are the intrinsic-to-job factors that may improve the worker’s feeling of belonging to a social community and promote his/her psychological growth[4].

The classification of needs of employed graduates may be juxtaposed to the Maslowian human needs hierarchy (Maslow, 1954). Interpreted in relation to work (Mc Gregor, 1960; Bonazzi, 2002), the needs may relate to:

(i) *physiology* if they concern short term worker’s economic safety,
(ii) *long term stability and professional safety,*
(iii) *socialization*, concerned with the worker’s environment,
(iv) *social esteem*, concerned with the workers’ expected appreciation for their activity and professional status;
(v) *self-accomplishment*, concerned with the enrichment of their psychological dimensions so to become “what they shall be able to become”.

Both Herzberg and Maslow hypothesize that each layer of human needs is to be satisfied in turn, from bottom to top. *I.e.*, a worker whose needs are below a certain level is not pulsed to satisfy a higher need.

Instead, the quantification of the importance of all needs for employed graduates, upon which our work relies, implicitly states that newly employed graduates are concerned, in different proportions, with all job quality dimensions[5].

If we borrow from the philosopher Campanella the image that social levels are like a terraced hill with a large basis and a progressively harder spiral path to the top (the so-called “Sun city” of Utopia), we can state that workers’ needs start from survival levels and move towards higher levels of self-fulfilment. Any worker aims to self-accomplish him/herself, once low and intermediate needs are satisfied. Graduates’ expectations start close to the hilltop. In fact, they expect to be offered high standard jobs, consistent with their educational level and strengthening their social esteem and early self-accomplishment.

The strictly motivational components qualify graduates’ jobs for 64.3%, those exclusively environmental for just 15.3%, those inherent to income and professionalism, which may be put half way between the two poles of Herz-

---

4 Herzberg (1966) argues that only the motivational factors may generate satisfaction, while income and the quality of the work environment may, at most, minimize work dissatisfaction. Vroom (1964) discussed the relevance of the individual motivation on job search and working activities.

5 Maslow’s scale describes the historical evolution of human needs. The dimensions of graduates’ jobs are cross-sectional, i.e. inherent to a particular time point of graduates’ lives. The graduates too, when the distance between their own aspirations and the real possibilities in the job market becomes overwhelming, will start satisfying the lower levels of the scale.

berg's scale, for the remaining 20.4%. Hence, even if the acknowledgement of competences were added to the environmental ones, the self-accomplishing dimensions would dominate the job quality rates.

Finally, our experts have put the job quality dimensions in a order that is similar to the Maslowian scale. What is different from that scale is the experts' indifference for social prestige stemming from the professional role of the newly hired graduate.

3. The employed graduate as a *homo socialis*

The representation of the job quality is different at the top stage of career and at the beginning (Table 2). Expectations of graduates tend to grow together with successes, may be conditioned to the work environment and evolution of social and professional models.

The most relevant differences between starting and final stages of career concern the importance of work environment quality and the possibilities of growth, both of which are less and less relevant as time goes on (the relative weight is 4% lower at the top of career for both dimensions), and that of external esteem, which increases of about 5%.

The possible changes the experts imagined are rather intuitive: the graduate's motivation for growth, together with the need for professional autonomy and creative freedom, remain salient. Altogether, these dimensions represent almost half (47.6%, but they were 51.6% at the start) of the job quality dimensions. The possibility of professional growth remains the most important dimension through the working life, so emphasizing that good jobs are continuously evolving jobs.

The external esteem for a graduate's job and hierarchical position increases its importance during the professional career. In fact, it may be a velleity at the very beginning, but it acquires relevance as the position consolidates. The ac-

Table 2. Percentage rates, and corresponding standard deviations, of 5 graduates' job quality dimensions, at the recruitment stage (after adjustments) and at the top of professional career, according to Delphi experts

Dimension	*Recruitment stage*		*Top of career*	
	Rate %	*Stand. dev.*	*Rate %*	*Stand. dev.*
Purposive autonomy	22.9	6.9	23.1	3.2
Expectation of professional enhancement	28.7	6.4	24.5	4.1
Internal competence acknowledgement	20.4	5.8	23.6	4.0
Social prestige	12.7	3.3	17.3	3.6
Work environment and worker logistics	15.3	5.0	11.5	4.1
	100.0		*100.0*	

knowledgement of the professional value of a job becomes a pathway for individual's social empowerment. A certain time after recruitment, a job becomes a vehicle to realize the ambitions of a graduate as *homo socialis,* a social person.

The quality of job environment, which is at the lowest rank in the beginning, loses further importance if referred to final stages of career. This is not surprising, since the quality of the work environment, being a "hygienic factor" *à la Herzberg*[6], is given for granted for the (high level) market segment we are dealing with. Of course, it is even less relevant if referred to the top managerial positions, which are likely at the edge of professional career.

Income, one of the internal acknowledgement features, assumes a different relevance at the last career stage. The employed graduates, focusing on their professional life, and comparing their income with the peers' ones[7], may perceive income as a gold standard for power and self-esteem.

Hence, the long-term representation is the most similar to Maslowian pyramidal order. The top dimension is that of self-realization through work, followed by social esteem and then by extrinsic quality needs qualified by work and relational environments.

It should be mentioned that the variance of rates assigned by experts to the dimensions is larger if referred to the initial working stages than to the job-consolidated paths. This suggests that experts' judgements are better defined if they relate to the final career stages (which the experts are directly experiencing) than to the volatile freshmen's conditions.

4. Semantic differential on expert's perception of the job market

In order to understand the basics of the opinions on job quality given by the consulted experts, we applied the method of semantic differential on their answers to a set of 20 couples of antithetic dichotomous adjectives. The experts defined, for each of the paired adjectives, which was the one that better represented their perception of the graduates' labour market.

The percentages of endorsement of each adjective are presented in Figure 1. The consulted experts perceive the large majority of features of the job market

---

6 Herzberg *et al.* (1959) hypothesize that job satisfaction and dissatisfaction are caused by different factors: satisfaction may derive from motivating factors while dissatisfaction by *hygiene factors*, where *hygiene* means work environment, worker logistics and income.

7 In comparing their incomes with peers' ones, women may assign a different value than men. Ross & Mirowsky (1996) state that women look, in a larger proportion than men, for other-than-economic satisfaction sources, like the acknowledgement of a well-done job.

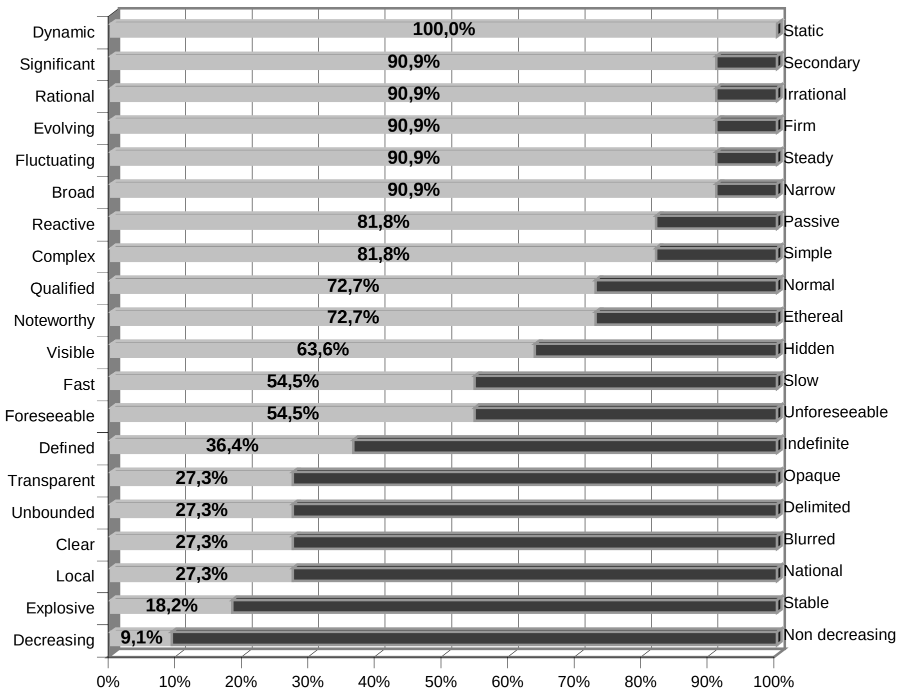


Figure 1. Relative frequencies of experts' endorsement of antithetic adjectives describing graduates' job market

in a similar way. They converge on perceiving that graduates face a very dynamic, rational, evolving, but fluctuating, job market. The analysts perceive differently its variations and the possibility given to them to understand the dynamics and forecast its endpoints.

The correspondence analysis (Greenacre, 1984; 1993) through the SAS-CORRESP procedure (SAS Institute, 1999, 2000) highlights the multiple relationships among experts' opinions. Main results[8] of correspondence analysis are presented in Figure 2, where the importance levels assigned by experts to the five dimensions of job quality at the recruitment stage are projected, too.

Along the main axis the vision of a dynamic and rational labour market, i.e. regulated by well-defined laws, is juxtaposed to a stagnant, unqualified and almost random market. The second dimension relates to the possibility of identifying the extension of the market, and the vision of a delimited and transparent market is juxtaposed to that of a vague, explosive and unbounded ground.

---

8 The two main dimensions account for almost 50% of the overall inertia. After Benzécri (1979) correction, the first dimension explains 51.5% and the second 28.4% of the inertia.

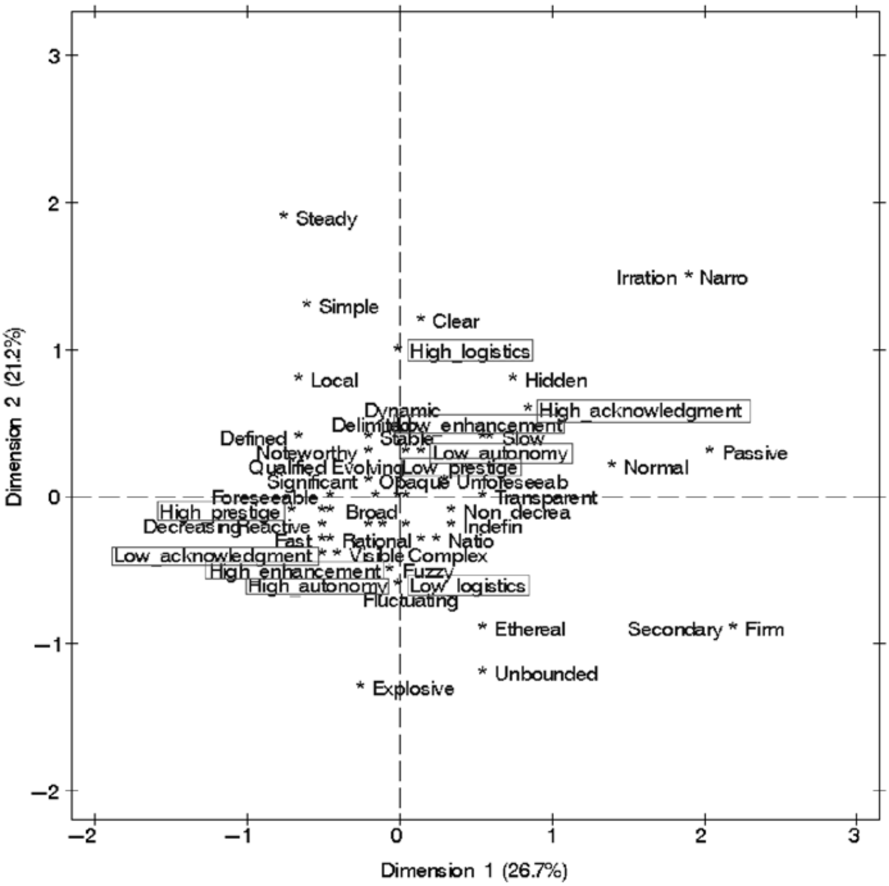


Figure 2. Correspondence analysis on bivariate occurrences of adjectives describing the graduates’ job market and projection of aspects rated by experts over the factorial axes

Both factorial dimensions are related to the possibility of forecasting how the job market evolves. The first dimension refers to the possibility to master the ruling principles of the market, the second to its complexity and extension.

These factors stem from the ratings given by the experts to the dimensions of job quality. The perception of difficulties in grasping the principles that govern the graduates’ job market is related to low and medium qualification level occupations, characterized by low levels of complexity, available mainly at the local level, in a certain way, to the professions less pertinent to the competences possessed by graduates.

On the opposite side, an occupational scenario characterized by qualified and wide-angle jobs, where the employees can exploit their competences and obtain decisional autonomy and growth chances, is perceived by experts as a rational situation they can master and represent on request.

5. Job quality satisfaction of employed graduates

The University of Padua graduates have rated the dimensions of their job market, too. Of course, the surveyed dimensions are those they could perceive at the very beginning of their professional career: the internal-to-company acknowledgement and the characteristics of the working environment. The missing ones are the growth chances, the decisional autonomy and the social (external) acknowledgement of job quality.

Our main hypothesis is that the newly recruited graduates reduce the complexity of job quality to the satisfaction rate for their job. The plausibility of the equation *Job quality = Satisfaction* for job is indirectly shown by the fact that graduates often associate high job satisfaction levels to underpaid or even to unpaid jobs, and show appreciation for potential job returns which undergo the contingent situation.

The determinants of job satisfaction are detected at two time points: six months and eighteen months after graduation. Before the first time point, the graduates experience several contacts with companies and attempt to start an autonomous activity. Between the two time points the working experience begins to consolidate, therefore the comparison between the two configurations may highlight possible trends of consolidated and prestigious career positions.

A multivariate regression analysis was performed on graduate's job satisfaction rate, on a 1 to 10 scale, as a criterion variable. The possible predictors may be classified in three categories:

- *characteristics of the graduates and of their career as university students*: attended course; gender; work situation at graduation; degree grade; time delay for degree achievement;
- *aspects of internal-to-company acknowledgement*: type of job contract; professional position; income; consistency between university title and job; value given to competences at work; use of competences acquired by attending either basic or professional courses; use of forma mentis for professional work; use of job-specific and on-the-job acquired competences; missing professional competences; excessively or not-enough job-specific education; adequacy of acquired competences for work;
- *aspects of working environment*: size and economic sector of the company; distance between home and workplace; number of coordinated employees; company habits to work in team; prevailing educational level of work-team; difficulties and advantages of working in team.

A sample of 1695 employed graduates and another of 1911 have been interviewed, respectively, six and eighteen months after graduation. Since graduates may be classified according to the attended faculty, we adopted in our first attempt a multilevel model with random intercept (Snijders & Bosker, 1999; Biggeri *et al*., 2001).

Table 3. Parameter and significance level estimates from stepwise regression analysis on the criterion variable "satisfaction with job six months after graduation"

	β	*s.e.*	*p-value*
Intercept	3.13	0.32	<0.001
Veterinary degree	-0.92	0.33	0.006
Autonomous job	0.50	0.16	0.002
Administration (general) employee	0.19	0.07	0.011
Unpaid work	0.99	0.38	0.009
Income lower than 750 Euro	-0.38	0.22	0.082
Company with 20 to 249 employed people	-0.15	0.08	0.073
Working in social services	-0.45	0.17	0.009
University degree needed for job	0.27	0.07	<0.001
Consistency between job and university degree	0.19	0.05	<0.001
Use of *forma mentis* for work	0.15	0.05	0.002
Competences missing for working purposes	0.14	0.07	0.052
Use of basic knowledge acquired at university	0.03	0.01	0.001
Relevance of competences at work	0.73	0.06	<0.001
Part-time job	-0.37	0.16	0.024
Difficulty in working alone	-0.81	0.27	0.003
Difficulty in team working	-0.87	0.21	<0.001
Advantages in team working	0.27	0.07	<0.001

The intercept variance was not statistically significant[9].

Therefore, we abandoned the multilevel approach and applied a multiple regression with stepwise selection of predictors (Draper, 1998). We used the SAS package (Sas Institute Inc., 1999; 2000) for regression estimation with parameters 0.10 significance level for selecting a marginal covariate, and 0.15 for discarding a selected covariate.

The estimated model is described in Table 3. The model explained 28% of the overall variance. It is rather evident that graduates' features do not explain satisfaction for job if the equation includes the indicators of job environment and working habits. The only exceptions are the graduates in Veterinary, who are much less satisfied than expected.

Various aspects of the internal-to-company acknowledgement and working environment are, instead, relevant for job satisfaction explanation. The internal acknowledgement pertains to economic gain, professional position and competence exploitation aspects. Since the survey was held just a few months

---

9 For the sake of precision, the intercept variance was very low for the nil model and went to zero after predictor selection. Hence, the application of the multilevel model gives parameter estimates analogous to those of the multiple regression analysis.

after recruitment, it is not possible to disentangle the professional career growth. Hence, at the very beginning, satisfaction stems from chances given to the newcomers to realize professionally qualified activities and to put in evidence their own professional skills.

Explanatory indicators of consistency between graduates' type of job and study contents, use of *forma mentis* and basic knowledge acquired at university, relevance of the degree and competences for their job, define the level of acknowledgment of graduates within the company.

Anyway, some distance between the education acquired at university and the competences required within the company – both for working alone and in team – is outlined by graduates as a job dissatisfaction cause. These difficulties are related mainly to cross-occupational skills.

Income is a positive issue for work refreshers but it is less relevant than other features of the internal acknowledgement of graduates' activities. In fact, a low income may be associated to an unsatisfactory job, but at least at the initial stage, interesting work activities may be more relevant for job satisfaction than being paid for that. In some sense, graduates feel, at the beginning, that a good occupational project may be an investment, similar to a prolongation of the study period.

Work setting is relevant for graduates' job satisfaction: working in a medium size company or in social services, and having part-time jobs are as makeshifts compared with both small and large size companies and full time jobs. The dissatisfaction of working in small size companies and in the social services is generated by the fact that most graduates belong to the humanities. That of part-time jobs may be due to the perception of some women to be unwillingly submitted to family duties and then to be forced to slow down their career.

Instead, the importance of the relationships with colleagues is clear, both for settings that privilege working in team or alone. Good work relationships may better the newcomer's perception of the work setting, whilst a good job may be spoiled by difficulties with colleagues. This corroborates the intuition of the Delphi experts that home-to-work distance is less important than the internal relational setting to define graduates' job satisfaction levels.

Autonomous jobs are associated with high levels of satisfaction. Indirectly, this indicates that work autonomy improves job satisfaction. In fact, who is more free to design and organize his/her own work than an autonomous worker?

For the analysis of the data eighteen months after graduation, we attempted a multilevel model, but, since the intercept variance was irrelevant after predictors' selection, we applied a multivariate regression model with stepwise selection of predictors. Parameters of the analysis were the same as in the six-month application.

The results, presented in Table 4, are quite similar to those of the six-month analysis: the estimated model explains a 30% variance, but neither personal

Table 4. Parameter and significance level estimates from stepwise regression analysis on the criterion variable "satisfaction with job eighteen months after graduation"

	β	*s.e.*	*p-value*
Intercept	3.13	0.22	<0.001
Bachelor obtained in Science Faculty	0.59	0.27	0.031
Autonomous job	0.30	0.13	0.028
Income above 1.200 Euro	0.15	0.08	0.074
Private company job	-0.23	0.06	<0.001
University degree needed for job	0.20	0.07	0.003
Use of *forma mentis* for work	0.14	0.05	0.004
Use of basic knowledge acquired at university	0.05	0.02	0.006
Use of professional knowledge at work	0.04	0.02	0.030
Relevance of competences at work	0.82	0.05	<0.001
Adequacy of formation for working purposes	0.05	0.02	0.013
Education unspecific for job purposes	0.12	0.07	0.064
Difficulty in working alone	-0.78	0.24	<0.001
Difficulty in working in team	-0.80	0.23	<0.001

characteristics of graduates, nor their university curriculum explain job satisfaction. The only exceptions, in this case, are the bachelors of the Science Faculty who show an above-the-average level of satisfaction.

Some time after recruitment, internal competence acknowledgement and socio-relational work setting determine job satisfaction. The importance of work setting lowers, in accordance with experts' perception of a diminishing relevance of the work conditions as the professional career consolidates.

Even eighteen months after graduation, it is evident that working in a private company has negative effects on satisfaction, while working autonomy induces positive effects.

Hence, broadly speaking, the regression analyses confirm the experts' intuitions that the first job quality is due mainly to the internal-to-company professional acknowledgement and autonomy left to graduates to design and organize their own job activities. The only aspects of the work setting that are relevant for job quality are the human and relational ones.

Our analyses highlight that the type of degree (bachelor, master) and the attended faculty do not significantly determine graduate's job satisfaction, even at the initial stages of work, with the only exception of one degree for each application. This implies that the existing differences in job satisfaction among graduates do not depend on their title, but on their individual professional paths. The graduates, who get acknowledgements for what they are able to do, and operate in a stimulating and reasonably competitive work setting, will be satisfied workers.

We can then state that job quality dimensions of graduates are rather universal and cross-disciplinary, and that the aspects presented in Sections 2 and 3 are common to all graduates with just slight differences between the technical and scientific programmes and the humanistic ones.

6. The design of questions on graduates' working life

We have been able to find the relevant dimensions of graduates' job quality with the help of experts consulted through a Delphi experiment. We estimated the importance of each dimension at the initial and consolidated stages of graduates' professional career and corroborated our estimates with the determinants of their job satisfaction. Now, we outline the possible questions to collect data on these aspects.

The most important dimension is growth potential at any stage of graduate's professional life. It is the cornerstone of any decision, from recruitment to subsequent activity changes. Although company dynamics are unforeseeable, the perspective tension of professional growth should be, and can be, perceived since the first contact between graduate and the company.

The indicators of professional growth trend are:

(i) the acknowledgment of competences by company's management and the consequent climbing of internal hierarchy. Graduates perceive that their competences may be valued if the relational environment is receptive and the structure is oriented to continuous learning. At least at the beginning, one should survey the consistency between the formative investment and the professional activities of graduates, i.e. if job-specific skills and *forma mentis* moulded at university as basic items. Some time after recruitment, one should be checked to see if graduates are assigned a responsibility position that gives them the possibility to manage projects and people.

(ii) The economic acknowledgment of professionalism. Even if income may appear initially secondary to other issues, in the end, social life and comparison with peers raises its importance. Hence, satisfaction for economic benefits and, possibly, (broad-class) income represent questions related to complementary aspects of the recognized professional value of a graduate worker.

One may argue if a newcomer is able or not to foresee his or her economic and professional future in a company. However, since job quality is the mirror image of the complex relations between the graduates and their job activities (within the framework of work setting), the workers are the best witnesses of their own professional destiny and the appropriate sensors of their jobs' multifacet evolutionary trends.

Another important dimension of job quality is the chance given to graduates to design and manage, in an autonomous and responsible way, their own work activities. The questions related to autonomy may deal with:

(i) jobs realized either as self-employed (entrepreneur, professional, handi-craftsman), or as partners in a cooperative or a professional office,

(ii) job flexibility level of an employee in terms of chance to manage times and modes of work. The cases of managers/executives, researchers, persons belonging to industrial technical offices, and other people who can manage themselves, may be put in evidence.

In a broader sense, work autonomy may be conceived as the chance to express creativity at work. Since the perception of professional freedom is a distillate of the empathic relations between the graduates and their work structure, this subjective aspect may be highly relevant for job quality measurement.

Another job quality dimension is the responsiveness of the work setting. A first aspect to be surveyed is the distance between home and workplace. Experts consider this aspect of little importance; nevertheless, Italian (young) graduates consider it a sensible issue in job bargaining. It may be asked if the workplace is either within the same commune of residence, or within the same province, or the same district, or outside the district, or abroad.

The sustainability of the work setting, as a daily vital environment and the quality of methods and work-tools available may be surveyed by asking graduates to rate any aspect of it. A question on setting sustainability and another on technical (or technological) level of the instruments may suffice to detect the physical aspects of work setting. To figure out the quality of relational dimensions, questions should be posed on relations with the hierarchy, colleagues and customers.

The last dimension identified by experts is the social (external) recognition of the graduate's job relevance. It is the least important dimension at the starting stage, but it becomes more and more relevant as the career progresses and the graduate settles down. The social acknowledgement may be detected by asking the prestige associated by relatives and close friends to the job position.

At the end, the level of overall satisfaction for job should be measured, anyway. Quantitative (for instance, 1 to 10) scales may allow comparisons in time and space more than just ordinal ones.

Our experts figured a graduate worker as a member of professional and social elites. However, graduate's job quality dimensions appear applicable for any other worker. Hence, in future experiments, it could be verified if there are differences in the expected intensity level of quality dimensions. For instance, is it the need of self-completion of the same order of magnitude for all categories of workers, or do expectations vary according to the (beginning, or target) hierarchical position? Do expectations vary according to the technical role? Do expectations vary according to interactions between expertise, technology and professional position?

In our experiment, we took for granted that job quality dimensions are similar in all companies. Nevertheless, small and medium companies require competences and satisfy workers in a way that is different from the larger ones[10]. How these differences were related to job dimensions, they may be an issue for further research.

Our experiment was not fully successful in detecting the relationship between the importance of job dimensions and experts' viewpoints. This may be due to a substantial homogeneity of viewpoints. It could be interesting to test if the thin semantic differences were related to the simply dichotomous scales used, or to the small sample size of experts.

References

BENZECRI J.-P. (1979) Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum à [BIN-MULT], *Cahiers de l'Analyse des Données*, **4**: 377-378.

BIGGERI L., BINI M., GRILLI L. (2001) The transition from university to work: a multilevel approach to the analysis of the time to get the first job, *Journal of the Royal Statistical Society, Series A*, **162(2)**: 293-305.

BLAUNER R. (1964) *Alienation and Freedom*, University of Chicago Press, Chicago.

BONAZZI G. (1989, 2002) *Storia del pensiero organizzativo*, FrancoAngeli, Milano.

CAPPA V., FABBRIS L. (2004) L'ambiente di lavoro come metro di qualità di una professione. In: FABBRIS L. (ed) *Capitalizzazione delle competenze dei laureati dell'Università di Padova*, Quaderno PHAROS n. 9/2004, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova: 27-42.

DRAPER N.R. (1998) *Applied Regression Analysis*, John Wiley & Sons, New York.

FABBRIS L., VISENTIN E. (2005) Tessere di competenze per mosaici di attività in aziende di varia dimensione. In: D'OVIDIO F. (ed) *Professioni e competenze nel lavoro dei laureati*, Cleup, Padova: 21-44.

FORD D. (1975) Shang inquiry as an alternative to Delphi: Some experimental findings, *Technological Forecasting and Social Change*, **7(2)**: 139-164.

GREENACRE M.J. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press, London.

GREENACRE M.J. (1993) *Correspondence Analysis in Practice*, Academic Press, London.

HERZBERG F. (1959) *The Motivation to Work*, Wiley, New York.

HERZBERG F. (1966) *Work and the Nature of Man*, World Publishing Co., New York.

HERZBERG F., MAUSNER B., SNYDERMAN B.B. (1959) *The Motivation to Work*, Chapman & Hall, London.

[10] Blauner (1964) finds that alienating production processes, typical of large companies, threatens the empathy between workers and company. Fabbris & Visentin (2005) highlight that larger productive organisations exploit specialized competences better than small-size ones, and generate higher satisfaction levels for newly recruited graduates.

LUTZ B. (1992) The contradictions of post-Tayloristic rationalization and the uncertain future of industrial work. In: ALTMANN N., KÖHLER C., MEIL P. (eds) *Technology and Work in German Industry*, Routledge and Keagan, London: 26-45.

MASLOW A. (1954) *Motivation and Personality*, Harper & Brothers, New York.

McGREGOR D. (1960) *The Human Side of Enterprise*, McGraw-Hill, New York.

NIELSEN D. (1993) *Partnering with Employees. A Practical System for Building Empowered Relationships*, Jossey-Bass, San Francisco.

ROSS C.E., MIROWSKY J. (1996) Economic and Interpersonal Work Rewards: Subjective Utilities of Men's and Women's Compensation, *Social Forces*, **75**: 223-246.

SAS INSTITUTE INC. (1999) *SAS/STAT® User's Guide*, Version 8, SAS Institute Inc., Cary NC.

SAS INSTITUTE INC. (2000) *SAS/STAT® Software: Changes and Enhancements*, Release 8.1. SAS Institute Inc., Cary, NC.

SNIJDERS T., BOSKER R. (1999) *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*, Sage, London.

TOURAINE A. (1955) *L'évolution du travail ouvrier aux Usines Renault*, Editions du CNRS, Paris.

VROOM V.H. (1964) *Work and Motivation*, Wiley, New York.

Networks of "Weak" Ties of Padua University Graduates Searching for Employment[1]

Anna Boaretto*, Gilda Rota*, Ilaria Silvestri*, Luigi Fabbris**

* *Internship and Guidance Service, Padua University, Italy*
** *Statistics Department, Padua University, Italy*

Summary. In this paper, we apply a model for the analysis of job search effectiveness through the so-called "weak ties", i.e. the personal links through which the attainment of a work position may be favoured. We analyse weak ties with reference to the first job research by the University of Padua graduates through family, friendship and university links. To obtain a work position more coherent with a graduate's educational investment, we analyse, too, the information about job opportunities that colleagues communicate to employed graduates. By hypothesising that graduates use their personal links in different portions, we segment, by means of a multivariate method, the sample of graduates according to the faculty and, within faculty, according to the curricular and social characteristics of graduates.
Keywords: Weak ties; Job Search; Parental links; Educational links; Professional links; Graduates; Padua University.

1. "Weak" tie networks

To return to the felicitous formula coined by Granovetter (1973), "weak ties" are those informal or extra-institutional ties that an individual is able to make use to get a job. The term covers both the acquaintances to which the individual knows he or she can turn and those effectively called upon to get a job regardless of the outcome.

---

[1] The four authors designed and realised jointly this research work. However, A. Boaretto wrote Sections 3 and 4 of the paper, whereas G. Rota wrote Section 2, I. Silvestri Section 5, and L. Fabbris the other sections.

That an individual's social network of acquaintance is a vehicle of opportunities towards the world of work is certainly not a recent discovery, nor is the phenomenon isolated to one part of the world (Katz, 1958; Lin *et al.*, 1981; Smelser & Swedberg, 1994). Such ties have been elevated to the status of respectable job-hunting channel since labour economists and economic sociologists have highlighted their relevance as alternatives or complements to the traditional pathways of job-hunting (Flap & De Graaf, 1986; Fevre, 1989; Watanabe, 1994; Bian & Ang, 1997). It is now commonplace to find that informal ties can generate or support opportunities, and that they are "weak" only insofar as they are less apparent (or less openly admitted to) than formal or institutional ties.

Graduates who apply for a job appear to the prospective employer - or selector acting on the employer's behalf - as a set of knowledge-driven, malleable abilities and wiliness to work in the context for which they are applying. If graduates also manifest their range of personal ties, these ties become part of how they present themselves. A recommendation from a person, of consequence, is perceived as a node in the network of relationships the person or institution evokes. Thus, when employing a graduate, not only do employers know that they are linking their network of acquaintance to the candidate's, they are also aware of gaining credit in the eyes of the person who made the recommendation.

Weak ties are not considered as socially negative; rather they are perceived as providing opportunities for the individual and results for the system of production as a whole. Of course, recommendations, which aim to place those candidates who benefit from them at the top of any shortlist regardless of their abilities, are exposed to individual moral considerations.

In the following, ties are evaluated in relation to the potential of obtaining their first job for new graduates from Padua University, and to the potential of changing jobs for those already in work who want to improve their position. What we seek to know is:

i) if there are groups of graduates looking for work who rely heavily upon these ties and whether the use of weak ties is exclusive, that is, the only job-hunting channel used, of if such ties are used to complement, perhaps in the wake of failed attempts, searches which followed traditional channels;

ii) if such ties are more effective in job-finding than formal or institutional channels, or whether they are effective only when used to defend positions obtained with other means, or, again, whether they facilitate low-level jobs, or high-profile positions (Granovetter, 1974). In other words, we want to understand the very nature of ties because it is quite different if they help to gain a low-level job, which present no substantial risk for the employer, instead of a high-profile position, which require total empathy between the new hire and the company.

In Section 2, we examine a number of hypotheses concerning individuals who make use of informal ties in their job-hunting. We analyse the acquaintances of and with friends and relatives (Section 3), as well as of any faculty members the graduates may be in contact with for research or teaching reasons, and in particular with their dissertation supervisor (Section 4), and with co-workers (Section 5). The hypotheses under examination are:

- *whether social status is linked to knowledge of and acquaintances in the labour market*. We hypothesise that the more the social background is elevated, the more likely it is that networks of the capitalist system of production, intellectual bourgeoisie and professional politics are aware of the possibilities available in the market for prestigious professions, and are able to dangle in front of their potential employers the possibility of "linking up with" the graduates' network;
- *whether female graduates, traditional less sure of themselves when applying for work, are more inclined to rely on their network of informal ties in job-finding*. The hypothesis that women make more frequent use of weak ties is controversial (Granovetter & Follis, 1998), such as it is also moot whether they do so because of entrenched cultural stances, because of objective weakness in the humanistic and social science disciplines that women traditionally tend to choose, or simply because of cultural inertia;
- *whether the academic qualification in question leads to differences in behaviour*. First, we can distinguish between the behaviour of university graduates. There is ample proof in the literature that the frequency of suggestions is linked to the level of academic achievement (Corcoran *et al*., 1980; Lin *et al*., 1981; Flap & De Graaf, 1986; Marsden & Hulbert, 1988; ISTAT, 1997), although a number of researchers have stated that the relationship is U-shaped. A second distinction can be made between the different faculties or academic schools, humanities or science, from which the individuals graduate, in that employment expectations may differentiate the feeling of prior acceptance from the market;
- *whether the individual's academic progress and achievement influence the use of weak ties*. The graduates' university curriculum is represented by age at graduation and final mark. All other things being equal, the younger the age at graduation the more efficient university progress has been, just as the final graduation mark indicates overall commitment to academic achievement.

The data under examination concern 2,057 graduates and diploma-holders from Padua University who look for work six months after gaining their qualification (Fabbris, 2003).

2. Effectiveness of weak ties for job finding purposes

When hiring a worker, a company seeks to identify the most appropriate candidate among a reasonably sized pool. For the same reason, job-hunters want

Table 1. Percent distribution of Padua University graduates and diploma-holders without jobs upon gaining their qualification, by use of weak ties in the search for work six months after graduation, by qualification, and by faculty

Type of degree	In searching for employment			No action	Total
	Weak ties	Weak ties and others	Other ties		
Master	9.5	30.3	32.8	27.3	100.0
Bachelor	8.2	27.4	48.1	16.3	100.0
Univ. of Padua	9.4	30.0	34.4	26.2	100.0

an idea of the vacancies available in the kind of job they are looking for in order to choose the best one. Suggestions from relatives, acquaintances and teachers tend to fulfil the needs of both new graduates and companies, and typically yield more information than the graduates would obtain simply by sending their CV directly to companies.

Around 40% of graduates make use of informal channels in job-hunting, that is, at least 53% of those who have made at least one attempt to find work[2] (Table 1). Among those graduates who made use of at least one informal channel, around a quarter used only this means, while the others used suggestions to back up other search options.

Those graduates who were in work at graduation and who make use of informal channels in job-hunting come mainly from the school of Education Studies, while those who were entering the world of work for the first time graduated mainly from Agriculture, Letters and Philosophy, Education, Psychology, Law, Political Science, and Economics. Graduates in Statistical Studies, Medicine and Surgery, and Engineering tend to use, almost exclusively, formal search channels (Table 2).

Thus, graduates from the humanities and social sciences as well as from Agriculture use informal channels more frequently. The latter present a number of differences compared with the former group, because of the nature of the profession: in the absence of a suitable heading, those agronomists who have found work with the family firm reported having made use of family ties. Thus, with the probable exception of Agriculture graduates, graduates who consider themselves in a weak position in the job market tend to make use of suggestions from friends and relatives.

Considering the graduate set as a whole, informal channels tend to be used in almost equal measure by both male and female graduates (Table 2). The tendency to do so among female graduates in the hard sciences, Veterinary Medicine, Engineering, and Statistics is slightly higher than male graduates

---

[2] The following did not engage in job-hunting after graduation: those who continued with their studies, those who fulfilled their military service obligations, and those who starter a family.

Table 2. Percent distribution of Padua University graduates and diploma-holders not in work upon graduation, by use of weak ties in their job hunting[3] at six months from gaining their qualification, by gender and by faculty

		Actions in searching for employment			No action	Total
		Weak ties	Weak ties and else	Other ties		
Agricolture	M	11.8	38.8	29.4	20.0	100.0
	F	12.8	43.6	28.2	15.4	100.0
Economy	M	6.8	34.1	45.5	13.6	100.0
	F	6.7	42.2	31.1	20.0	100.0
Pharmacy	M	6.5	29.0	41.9	22.6	100.0
	F	10.2	16.7	45.4	27.8	100.0
Law	M	5.9	13.2	26.5	54.4	100.0
	F	7.1	24.4	22.8	45.7	100.0
Engineering	M	13.9	32.7	37.2	16.2	100.0
	F	20.5	38.5	35.9	5.1	100.0
Humanities	M	5.9	35.3	26.5	32.3	100.0
	F	4.4	38.2	42.6	14.8	100.0
Medicine	M	18.2	9.1	54.6	18.2	100.0
	F	10.6	27.7	53.2	8.5	100.0
Veterinari Med.	M	9.1	18.2	27.3	45.5	100.0
	F	27.3	4.6	22.7	45.5	100.0
Psychology	M	6.9	17.2	10.3	65.5	100.0
	F	9.3	15.0	23.6	52.1	100.0
Pedagogy	M	25	50.0	0.0	25.0	100.0
	F	7.3	47.7	36.7	8.3	100.0
Political Science	M	9.7	26.6	28.6	35.1	100.0
	F	2.4	43.2	33.1	21.3	100.0
Sciences	M	10.3	29.3	39.7	20.7	100.0
	F	3.4	39.0	44.1	13.6	100.0
Statistics	M	4.7	25.6	55.8	14.0	100.0
	F	9.1	47.3	34.6	9.1	100.0
University of Padua	M	11.2	29.2	34.5	25.1	100.0
	F	8.0	30.7	34.3	27.1	100.0

from the same faculties (around 10%). As these are scientific disciplines, it means that, although the world of work is certainly more open to women today than in the past, female graduates continue to feel more at risk than their male counterparts when applying for jobs, and reinforce their chances through recommendations.

[3] Weak ties in the first six months from graduation were used overall by 174 individuals; informal ties in association with another channel, 623; other channels by 623; no channel by 407, for a toal of 1980 graduates and diploma-holders.

Table 3. Percent distribution of Padua University graduates and diploma-holders not in work upon graduation, by use of weak ties in job-hunting at six months from graduation, by qualification and degree result

	Degree	n	Actions in searching for employment			No action	Total
			Weak ties	Weak ties and else	Other ties		
Master	Up to 89	132	12.1	28.9	41.6	17.4	100.0
	90-99	550	10.7	33.3	32.6	23.4	100.0
	100-110	866	8.8	29.0	31.4	30.8	100.0
	110 laude	244	7.6	28.7	32.5	31.2	100.0
	Total	1792	9.5	30.3	32.8	27.3	100.0
Bachelor	Up to 89	16	0.0	27.9	50.4	21.7	100.0
	90-99	75	6.9	29.1	46.4	17.5	100.0
	100-110	79	12.8	27.8	46.9	12.5	100.0
	110 laude	11	0.0	39.5	41.9	18.6	100.0
	Total	181	8.2	27.4	48.1	16.3	100.0

Table 4. Percent distribution of Padua University graduates and diploma-holders not in work at graduation, by use of weak ties in their job search at six months from graduation, by qualification and age at graduation

	Age degree	n	In searching for employment				Total
			Weak ties	Weak ties and else	Other ties	No action	
Master	< 24	57	4.6	30.8	40.2	24.5	100.0
	24-26	776	9.0	28.7	32.0	30.4	100.0
	26-28	950	9.7	29.7	31.2	29.3	100.0
	28-30	407	14.0	30.5	30.4	25.1	100.0
	> 30	312	7.2	28.8	32.5	31.6	100.0
Bachelor	< 24	146	11.9	22.4	42.3	23.4	100.0
	24-26	96	7.8	19.4	57.2	15.7	100.0
	26-28	36	1.9	41.2	41.3	15.5	100.0
	28-30	21	13.0	33.9	33.6	19.4	100.0
	> 30	12	26.1	11.7	38.4	23.7	100.0

It is an unquestionable fact that graduates who feel weaker when it comes to the job market make use of informal ties. A higher proportion of individuals who graduate after the age of 26 or who graduate with a low final degree result make use of weak ties (Tables 3 and 4).

Separate consideration needs to be given to suggestions from university teachers. The category that benefits most from this kind of recommendation is that of new graduates from the scientific faculties who graduate at a young age with a good result.

These reasons for this state of affairs are easily identifiable: teachers in scientific disciplines have more contact with the business community and government agencies than their counterparts in the humanities and these teachers tend to mention the names of their more brilliant students, those who will reflect well on the teachers themselves. Suggestions from faculty members stand apart, because companies perceive them as rigorous declarations of the quality of the graduates in question.

In any case, the use of these channels is usually a winning move, regardless of the faculty or degree course. It is slightly more effective for bachelor degree-holders who apply for jobs with lower expectations in terms of salary and position in the company (Table 5).

Table 5. Percent distribution of Padua University graduates and diploma-holders in work at graduation, by use of weak ties in their job search at six months from graduation, by qualification and work status at graduation

	Degree	In searching for employment			No action	Total
		Weak ties	Weak ties and else	Other ties		
Employed	Master	88.5	81.3	76.8	75.1	81.0
	Bachelor	100.0	94.0	93.5	81.2	91.3
	Total	90.7	82.7	86.1	75.8	82.6
Unemployed	Master	69.4	62.5	58.8	11.3	48.1
	Bachelor	95.9	70.3	92.7	25.5	75.8
	Total	71.7	63.3	63.7	12.2	50.9

Table 6. Percent distribution of Padua University graduates and diploma-holders who found employment working for others or under an "atypical" contract at six months from graduation, by use of weak ties in their job search, and by relevance of job to qualification

Coherent?	Weak ties	Weak ties & else	Other ties
Not at all	6.0	9.6	9.3
Little	12.3	12.9	15.9
Enough	50.8	50.0	48.2
A lot	30.8	27.5	26.6
Total	100.0	100.0	100.0

Table 7. Percent distribution of Padua University graduates and diploma-holders who found employment working for others or under an "atypical" contract at six months from graduation, by use of weak ties in their job search, and by whether the job they do could be performed by a high school graduate or by someone less academically-qualified

It could be done by	Weak ties	Weak ties and else	Other ties
a non-graduate	44.3	45.4	49.5
a high school graduate	14.9	17.3	16.8
only university graduate	40.8	37.3	33.7
Total	100.0	100.0	100.0

Individuals who make use of informal ties not only find it easier to obtain work, they also tend to find positions that are in line with their qualification (Table 6). A proportion of 40% of graduates who used only their acquaintances (compared with 33% who used only formal channels) declared that their work required an academic qualification and fully 82% (compared with 75% who used only formal channels) have jobs that reflect their academic qualification (Table 7).

Because of finding a job in line with their qualifications, graduates are more optimistic with respect to professional growth and salary progression, compared with co-workers who, having obtained their position on their own strengths, start out from slightly less advantageous positions (Table 8).

Graduates usually follow more than one direction in their search; very few of them adopt only one channel in their job search. However, it is misguided to think that more attempts will lead to a higher probability of success.

Indeed the figures show that graduates, who look for work using either their own means and strengths, or informal networking, find work more easily than graduates who use both methods. Individuals who use both traditional and underground channels do so because they feel at a disadvantage and usually obtain less satisfactory results in employment terms than graduates who select just one approach (Table 4).

It can be deduced, then, that the fact of having a wide acquaintance, of being in contact with a large number of people rather than isolated in closed "networks", however intense they may be, is a winning condition in the hunt for work. We can infer that social background does influence, in the first instance, the choice of degree or qualification, and then the entry into work.

Table 8. Percent distribution of Padua University graduates and diploma-holders who found employment working for others or under an "atypical" contract at six months from graduation, by whether they used weak ties in their job search and by their expectations for professional progression in two years' time

Professional expectations	Weak ties	Weak ties and others	Other ties
Few or no changes	4.6	11.2	8.5
Economic improvements	14.2	6.8	5.2
Career progression	13.3	21.0	28.9
Increase of the operating within	21.4	6.6	15.8
Professional institutionalization	5.1	1.4	3.9
Social acknowledgment	2.2	0.0	4.0
Other	39.2	52.9	33.6
Total	100.0	100.0	100.0

3. Parental ties

To obtain a more precise picture of graduates who use informal search channels, we processed the sample with binary segmentation analysis (Fabbris, 1997), taking as the criterion variable the *logit* of the conditional probability that graduates would make use of the knowledge of friends and relatives.

The analysis was carried out using the LAID-OUT package (Schievano, 2002), following the variant proposed by Fabbris & Martini (2002) which, at each step in the analysis, maximises the difference in the logits of the proportions in the sub-samples that are formed by the dichotomous split in the sample:

$$\left[logit\left(\hat{\pi}(\mathrm{Y} \mid \mathrm{X}_1)\right) - logit\left(\hat{\pi}(\mathrm{Y} \mid \mathrm{X}_0)\right)\right] ,$$

where $\pi(Y|\mathbf{X})$ denotes the value of dependent variable Y conditioned by the dichotomized predictor vector $\mathbf{X}$ and *logit(.)* indicates the natural logarithm of the term in brackets related to one on completion.

The predictors considered in the segmentation are faculty attended, qualification (Bachelor/Master), gender, age at graduation (6 classes), degree result (above/below average), previous employment status (was/was not in work at graduation).

For the segmentation analysis, we decided

a) To force the variable "in work at time of graduation" as the first predictor;

b) To select ternary segmentation analysis where it is clearly better than a couple of binary segmentations;

c) To limit the minimum group number to 50.

It emerges that, employment status being equal prior to graduation, the factors which influence the choice to use personal and family acquaintances are the faculty attended, gender, age at graduation, and final degree result.

Among those in work at graduation, students who graduated from the Faculty of Education were more inclined (63.3%) to use personal connections compared with graduates from other faculties. In all likelihood, this is due to teachers/educators who use the relationships consolidated during their own working life to improve their position within private companies/bodies, such as nurseries, playschools, or health centres.

Among those graduates not in work who looked for work, the majority who made use of weak ties came from the Faculty of Education (52.9%), followed by Agriculture, Psychology, Law, Classics, Political Science and Economics, which presented average levels of employment slightly lower than the others (Fabbris *et al.*, 2002).

Gender is a significant factor in the breakdown of graduates from Engineering, Veterinary medicine, Mathematics, Physics, and the Natural Sciences. Female graduates look for jobs by means of weak ties in 51.1% of cases, as opposed to 41.6% for their male counterparts. In faculties that are considered

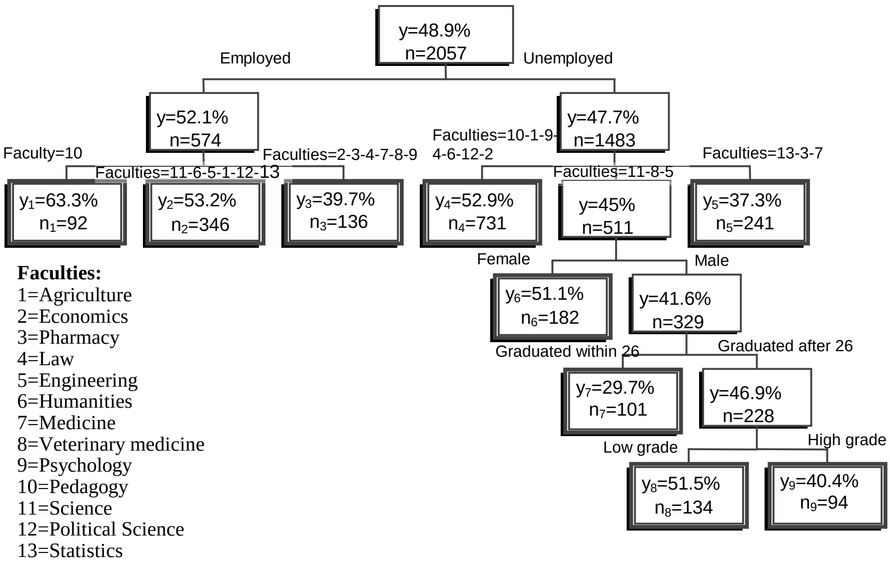


Figure 1. Segmentation tree for the “looks for work using personal and family acquaintances” variable

Table 9. Percentage in work, average level of satisfaction and percentage of interviewees who state that their job is in line with what they studied, by group derived from the segmentation analysis

Group	*n*	*% employed*	*Mean satisfaction*	*% very coherent with studies*
1	92	90.2	7.1	38.3
2	346	83.6	7.0	17.1
3	136	87.4	7.3	41.0
4	731	54.0	7.1	19.7
5	241	80.5	7.4	34.3
6	182	68.7	7.3	41.1
7	101	74.4	7.5	39.6
8	134	76.8	7.4	43.7
9	94	66.6	7.6	37.2
Total	2817	60.7	7.3	28.4

“strong” in employment terms, female graduates seek further guarantees. This may depend on the persistent masculine overtones attached to the professional qualities female graduates may believe they possess, and thus on the reduced chances they believe they have when it comes to applying for jobs, because of the fact that they have studied in faculties with male student prevalence.

Male graduates approach job-hunting in ways that are influenced by their final degree result and their age at graduation. Graduates who took longer to complete their degree and achieved lower results are more inclined to use personal and family connections (51.5%) compared with those who graduated in less time (17% fewer graduates in this category make use of weak ties) or with a high final result (11% fewer).

In such cases, weak ties bolster academic records that are less than brilliant. Nevertheless, the employment rate in this apparently less-gifted category stands at 76% and the percentage of graduates who declare that their jobs are in line with what they studied is 43.7%, among the highest percentage returned in the segmentation analysis (Table 9).

4. University ties

Among graduates who sought work within six months of graduating, 12.6% made use of contacts with teachers met during their studies at Padua University.

Data from the segmentation analysis (Figure 2 and Table 10) show that those students who had worked during their degree course tended less to make use of connections with teachers, and usually graduated at 30 years of age or later (2.8%), regardless of which faculty they had attended.

It is not surprising that graduates over 30 already in work have little familiarity with teaching staff, given that their aim is usually to use their new qualification to progress at work, but that, as they attended courses less often, they did not have the opportunity to build up a network of university contacts. Among graduates, this category declares the lowest satisfaction level (6.7%) and the lowest level of congruity between their jobs and the degree studied (15.7% compared with an average of 28.4%).

Among students already in work, those who used most their connections with teaching staff were early graduates from the faculties of Agriculture, Economics, and Engineering (20.2%).

Among those not in work at graduation, students from Political Science or Law were the least likely to make use of connections with their teachers (3.6%), whereas graduates with a high final graduation mark from Agriculture, Economics, Engineering, Veterinary medicine, the hard and natural sciences were more likely to make use of those connections (27.7%). For this last group, the congruity index between jobs and degree courses was particularly high (41.1%) and satisfaction in the job was above average (7.5%).

The factors, singly or in combination, which may lead students in scientific disciplines to cultivate a privileged relationship with their teachers are the following:

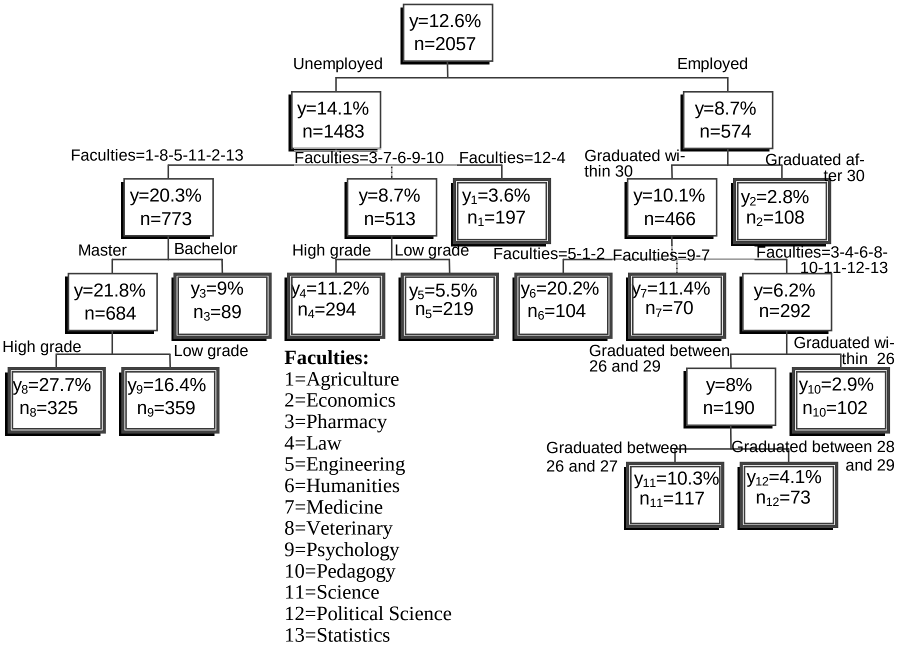


Figure 2. Segmentation tree for the "looks for work using professors' acquaintances" dependent variable

Table 10. Percentage of those in work, average satisfaction level and percentage of interviewees who declare a high degree of congruity between their job and their degree course based on groups derived from segmentation analysis

Group	*n*	*% employed*	*Average satisfaction rate*	*% very coherent with studies*
1	197	43.6	7.5	19.8
2	108	88.3	6.7	15.7
3	89	78.8	7.4	30.7
4	294	60.0	7.2	25.7
5	219	69.6	6.9	21.3
6	104	89.5	7.4	27.5
7	70	89.7	7.5	45.7
8	325	69.1	7.5	41.1
9	359	71.3	7.4	26.7
10	102	82.9	7.1	33.4
11	117	79.8	6.8	24.6
12	73	90.3	7.2	21.1
Total	2817	60.7	7.3	28.4

- a high ratio of students to teachers effectively influences the possibility of establishing relationships of trust between the two groups;
- the experimental nature of most dissertations produced on scientific degree courses compared with the dissertations in the humanities which are more frequently reviews or compilations; an experimental dissertation, more often than not involving issues and methods close to the teacher's research interests, usually involve close collaboration between the student and future supervisor;
- the greater openness of scientific faculty members towards the world outside academe and in particular to the business community.

5. Professional ties

The tendency on the part of the 480 new graduates who were in work to turn to other co-workers and colleagues as part of their job search strategy was evaluated in proportion to the number of contacts obtained to which the graduates could send their CV.

In the segmentation analysis, the following possible predictor variables were added to the variables used in the previous two analyses: "position at work" (above or below the rank of salaried office employee); "the company's business area" (private or state); "the size of the company" (number of workers greater or lesser than twenty).

A summary of the results of this analysis is shown in Figure 3 and Table 11. The category which made greater use of contacts with co-workers (53.8%) was made up of graduates from almost all faculties (except Economics, Pharmaceutical Studies, Law and Classics) who had a "fall-back" job, on a lower level than salaried office worker (teacher, factory worker, shopkeeper, craftsman or woman, "atypical" contract) and who graduated quite late, after the age of 28. Members of this category typically declare that they feel weak in professional terms and express lower than average job satisfaction (6.8%).

Graduates who have a position equivalent to salaried office worker, manager or executive in the state sector, tend to make less use of such contacts, and usually occupy a role which is congruous with what they studied (33.4% compared with an average of 28.4%) and which gives them a high level of job satisfaction (7.6%).

Age at graduation is also directly proportional to the tendency to make use of ties established in the private sector; individuals who graduate later have had the time to establish their own network of contacts at work.

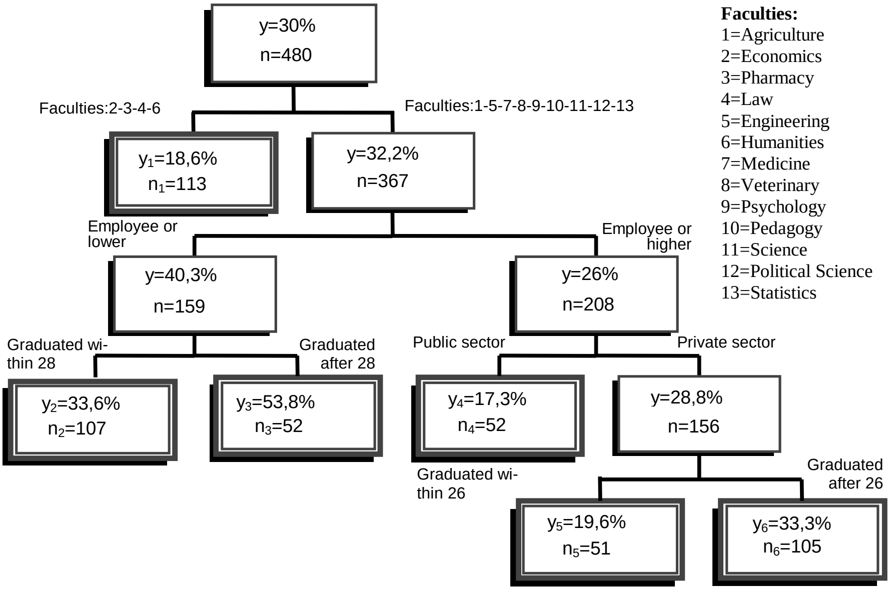


Figure 3. Segmentation tree for the dependent variable "made use of help from co-workers or acquaintances to send CVs"

Table 11. Average satisfaction and percentage of interviewees who state that their job is congruent with what they studied based on groups identified during the segmentation analysis

Group	*n*	*Mean satisfaction*	*% very coherent with studies*
1	113	7.0	17.4
2	107	7.2	35.8
3	52	6.8	33.5
4	52	7.6	33.4
5	51	7.3	41.4
6	105	6.8	8.8
Total	2817	7.3	28.4

6. Final considerations

Individual networks of acquaintance are used by one in two of all Padua University graduates looking for work. It is possible that in areas with a less effervescent job market than the Veneto region, the proportion will be even higher.

It is worth making a distinction between suggestions/recommendations from university teachers and those from family and friends. The former are opportunities offered to the new graduate directly by a teacher, often as compensation for particularly good dissertation performance. Faced with such a proposition, a graduate can only say yes or no. If he or she does not accept, the opportunity will go to someone else, as the tie is strictly a function of the teacher's discretion.

The opportunities that arise in an academic context are thus particularly concrete, and may be "spent" just once for one position; such opportunities are short-lived and are asymmetric in their configuration, from the teacher towards to the student-graduate.

The second category is multiple opportunities, beneficial for those who can make use of them. They are, in this sense, acquisitions which belong entirely to the graduate, shared with the family or social group, but which the individual can "spend" whenever he or she perceives the usefulness in doing so. In other words, out of five recommendations exercised during the hunt for work, four come from the family and social group and one from the academics.

We must underline the extent to which our analyses involve a homogeneous group of people, characterized by the fact that all members posses a superior academic qualification and are of similar age.

In order to make use of one's connections, one needs to have some. Recommendations from teachers usually favour almost exclusively those students who have produced a convincing experimental dissertation. Consequently, this category of ties is narrowly limited to those students who make good academic progress, who do not work and dedicate almost all the time to studies, and who have followed courses strongly geared towards their profession.

The extent of acquaintance through family and friends varies considerably in relation to the new graduates' social background. The higher social classes have at their disposal a greater range of contacts in searching for work.

Nevertheless, those who feel they have greater need use such contacts more frequently and more intensely. Thus, they are used mostly by graduates from the humanities and social sciences, by individuals who gain their qualification at the end of difficult academic or personal experiences, and those who feel they are at a disadvantage.

Among those who that feel they are at a disadvantage, we find numerous graduates who make systematic use of recommendations, to a much greater extent than their counterparts with similar degree qualifications, but to a lesser extent than the student-dissertation supervisor axis. What is interesting is that these contacts are also used by graduates who have perfectly acceptable degrees for job hunting, such as a degree in engineering or other scientific discipline.

It is thus evident that weak ties are called upon to act as shields when graduates or diploma-holders feel exposed, or when they want to use an "external" force to even out their chances of successfully finding a job which, in their view, birth, the life they have lived, and their academic achievement have rendered unequal.

If this statement were true, the use of a network of informal contacts to obtain a job one thought was lost to one rather disqualifies the applicant on moral grounds. Regardless of how we analyse these data, it is clear that those who make use of the network of acquaintance, because they feel weak, run an explicit risk of bearing the brunt of moral judgement; whereas those graduates whose curricula testify to a certain degree of accomplishment and professionalism, play the recommendation card to enhance further their own position.

From this moral judgment, we can exclude those individuals who use contacts in their working environment for no other reason than to learn before possible rivals what opportunities are available in the job market that could improve their own professional position.

Nevertheless, companies and selectors do not perceive the use of informal channels as harmful. First, as the selector gains more information on the candidate from the recommendation itself, it narrows the margin of uncertainty attached to choosing candidate. Second, from the implicit complicity that underpins a recommendation, the selector draws a certain guarantee regarding the behaviour of the new employee, in that the later must ‘answer’ not only to his or her employer, but also to the person who made the recommendation. Moral judgments on the recommendation as an institution are thus relative.

In conclusion, it would be very interesting, but with our data it is not possible to do so, to examine if and to what extent ethical and cultural factors influence the use of informal ties to find work. Given a social status and an academic qualification, some graduates make use of the social network while others do not. An understanding of the extent to which an individual’s values act as a hindrance to his or her using those ties would help us formulate a more precise assessment of this social practice.

It would also be instructive to assess if and for whom in the mid to long term these contacts are useful, including in career progression.

For those already in works who are seeking a better position, we would need to examine either if length of service is connected with the use of ties in order to gain a new job, or whether, with the passing of time in a job, ambitions and contacts lose their importance.

References

BIAN Y., ANG S. (1997) *Guanxi* Networks and Job Mobility in China and Singapore, *Social Forces*, **75**: 981-1005.

CORCORAN M., DATCHER L., DUNCAN G.J. (19080) Information and influence networks in labor markets. In: DUNCAN G.J., MORGAN J.N. (eds) *Five Thousand American Families: Pattern of Economic Progress, vol. VIII*, Institute for Social Research, University of Michigan, Ann Arbor, MI: 1-37.

FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.

FABBRIS L. (2003) *Il monitoraggio dell'inserimento occupazionale e professionale dei laureati e dei diplomati dell'Università di Padova*, Quaderno PHAROS n. 6/2003, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova.

FABBRIS L., MARTINI M.C. (2002) Analisi di segmentazione con una variabile dipendente trasformata in *logit*. In: CARLI SARDI L., DELVECCHIO F. (eds) *Indicatori e metodi per l'analisi dei percorsi universitari e post-universitari*, CLEUP, Padova: 195-214.

FABBRIS L., MARTINI M.C., ROTA G. (2002) Il valore di mercato dei titoli di laurea e diploma di Padova a sei mesi dal conseguimento del titolo. In: CARLI SARDI L., DELVECCHIO F. (eds) *Indicatori e metodi per l'analisi dei percorsi universitari e post-universitari*, CLEUP, Padova: 171-194.

FEVRE R. (1989) Informal Practices, Flexible Firms and Private Labour Markets, *Sociology*, **23**: 91-109.

FLAP H.D., DE GRAAF N.D. (1986) Social Capital and Attainment Occupational Status, *Netherlands Journal of Sociology*, **22**: 124-132.

GRANOVETTER M. (1973) *The Strength of Weak Ties*, The University of Chicago Press, Chicago.

GRANOVETTER M. (1974, 1995) *Getting a Job: a Study of Contacts and Careers*, The University of Chicago Press, Chicago.

GRANOVETTER M., FOLLIS M (eds) (1998) *La forza dei legami deboli e altri saggi*, Liguori Editore, Napoli.

ISTAT (1997) *Rapporto sull'Italia. Edizione 1997*, Il Mulino, Bologna.

KATZ F.E. (1958) Occupational Contact Networks, *Social Forces*, **37**: 252-258.

LIN N., VAUGHN J.C., ENSEL W.M. (1981) Social Resources and Occupational Status Attainment, *Social Forces*, **59**: 1163-1181.

MARSDEN P.V., HURLBERT J.S. (1988) Social resources and mobility outcomes, *Social Forces*, **66**: 1083-1095.

SCHIEVANO C. (2002) LAID-OUT.1: un programma per l'analisi di segmentazione binaria con riferimento ad una variabile dicotomica trasformata in logit. In: PUGGIONI G. (ed) *Modelli e metodi per l'analisi di rischi sociali e sanitari*, CLEUP, Padova, 21-36.

SMELSER N.J., SWEDBERG R (eds) *The Handbook of Economic Sociology*, Princeton University Press, Princeton NJ, Russell Sage Foundation, New York.

WATANABE S. (1994) Strength of Ties and Job-Searching in Japan, *American Sociological Review*, **56**: 60-71.

University Choices and Work Destinies in Italy[1]

Franca Crippa*, Paolo Mariani**

** Psychology Department, University of Milan-Bicocca, Italy*
*** Statistics Department, University of Milan-Bicocca, Italy*

Summary. The need to match academic education with the 'know how' required by the labour market may give birth to both co-operation between universities and enterprises, and to 'flexible' academic curricula. Students can choose from a wide range of University degrees, and within each program can choose the applicative field/major and opt for a short work experience, the internship. The feedback from the university-enterprises interaction may be observed in graduates' job opportunities, since they constitute the third pole in a triad of changing relationship. The case of the Faculty of Statistical Sciences at the University of Milan – Bicocca, in the academic years from 1998/99 to 2002/2003, allows one to identify patterns in undergraduates' educational strategies and poses questions as to the relationship between academic knowledge, individual expectations and actual experiences at work.
Keywords: Competences; Transition times; Non-metric individual differences.

1. Academic and entrepreneurial perspectives

The weakening self-governance of the university system, in favour of the labour market, has raised questions about the usefulness of undergraduate studies with regard to professional requirements (Ottaviani, 2004). The consequent adjustment process gave birth to experiences of co-operation between universities and enterprises, which has put into motion a process that favours some disciplinary areas.

---

[1] Sections 1 and 4 were jointly written by F. Crippa and P. Mariani, Sections 2 e 3 by F. Crippa.

The traditional hiatus between theoretical and applicative disciplines allows the latter a ready and flexible fine-tuning adaptation of curricula, *via* the inclusion in the undergraduate education on the job experiences, hence merging the academic and the entrepreneurial perspectives.

In the creation of a professional identity, this experience widens the students' range of choices and may affect their means entering the labour market: so it is possible to calculate multiple career paths arising from specific choices, in terms of their external efficacy[2] (Fabbris, 2003).

Our study focuses on a particular case: in accordance with the guidelines for university/enterprises interactions, the Faculty of Statistics at the University of Milan-Bicocca includes forms of job training (*internship*) at selected companies' sites in its curricula, under both academic and corporate supervision.

Ever since the so called "3+2" reform, that started in 2001, some undergraduate credits can be acquired for periods of training no less than three months, as well as introducing students into the labour market. This is particularly relevant for statisticians, whose qualifications are still not well defined in the Italian labour force scenario, with a gap between an offer of 1,862 positions for statisticians in 2004[3] and 1,287 graduates in Statistical Sciences in 2003 (Unioncamere-Ministero del lavoro e delle politiche sociali, 2004).

In order to improve its services, a first survey on the Faculty outcomes was carried out in the Academic Year (A.Y.) 2000-2001 (Martini, 1997), thereafter replicated every academic year, named 'Statisticians and Work'. In 2003, it was devoted to the Faculty graduates' work situations.

Data collection was conducted in November 2003 by means of telephone interviews regarding curricular and professional characteristic of the 300 graduates at the Faculty of Statistics, University of Milan-Bicocca from A.Y. 1988/89 to A.Y. 2002/2003, using lists of former students provided by administrative offices[4]. The response rate was 76.7%.

Each interviewee, in addition to the characteristics of his course of studies, provided information on the qualifications gained during her/his educational experience, such as internships, masters or other specialisations believed to be relevant to occupational opportunities. The occupational status at the time of the interview was classified as employed, unemployed and in search of first employment. In addition, it was pinpointed whether any kind of employment was entered into prior to graduation.

---

[2] Average waiting times before occupation, apart possible *post lauream* experiences (such as internships), may be considered indices of external effectiveness of education (Fabbris, 2003).

[3] Estimates of the Excelsior Information System.

[4] Data collection, retrospective dating November 2003, includes more than ten cohorts and spans across a time interval with changing market conditions. Thanks to the favourable conditions of the labour market in Lombardy in these years, such sources of heterogeneity are, at least partially, mitigated.

On the subject of present profession, questions focused on channels and length of job search, activity branches, whether self-employed or on a temporary contract and to what extent the academic knowledge was utilised, and level of job satisfaction.

2. From graduation to work

Descriptive summaries of the data indicate a high rate of employment, namely 89.9% of the subjects under investigation, even before graduation (37.4%) and generally with stable, long lasting contracts (60.3%). An internship was performed by approximately 50% of persons and, for about half of them, leads to a professional position, commensurate for a graduate. Furthermore, the Faculty's Internship Office is referred to often, directly, by enterprises, so that the university adds to its formative function the role of facilitator in the transition to work, representing up to 36.7% of entrants in the labour market.

The present employment position is regarded as befitting their academic degree by 71.5% of subjects; only half of them, however, reckon their functions pertain exclusively to a degree in Statistics, whereas 45.7% think they pertain to other scientific branches, and 5% to any degree. In line with this evidence, specific competences acquired at university are linked to the requirements of the job in the opinion of 66.8% of the graduates, while for the others at least some of their competences are redundant.

As a matter of fact, descriptive statistics play a large role in routine corporate work (66.7%, multiple answers), multivariate analysis absorbs a substantial area (25.7%) while, on the contrary, statistical inference, econometrics and time series analysis are rarely used (respectively, 15.2%, 9.9% 3.9%). The gap between the material studied and effectively used does not seem to discourage newly-graduated statisticians, as they consider themselves satisfied in their professional positions, even if not enthusiastic (Table 1). Neither the graduates express an immediate goal of job improvement, being low the percentage wishing to look for a new job in the next three years (19.1%).

Table 1. Satisfaction for the current profession (graduates in Statistical Sciences at the University of Milan – Bicocca employed in November 2003, percent values)

Satisfaction regarding	*Very*	*Fair*	*A little*	*Not at all*
career perspective	16.8	53.6	19.9	9.7
tasks	26.8	52.0	15.7	5.6
permanence in the position	36.9	29.8	21.2	12.1
wage	18.5	55.9	18.5	7.2
autonomy	45.2	44.7	8.1	2.0
degree of responsibilities	41.1	46.2	10.7	2.0

2.1 Duration in the transition to current professions

Approximately one third of interviewees entered their present job before graduation, 36.4% of whom being employed in the same company which offered them an internship. The internship achieves the value of a substitute for the first trial months typical of many standard contracts in Italy, but free of bureaucratic and administrative red tape.

The remaining subjects were asked how long they waited since they started searching for and when they actually found employed in their current job. Search time for entering the labour market are estimated only in relation to subjects unemployed at graduation, by means of non-parametric methods for study of event history (EHA - *Event History Analysis*), applying actuarial methods (life tables) for discrete failure time variables and assuming the trimester as the time unit.. One of EHA purposes[5] is to estimate the risk of experiencing a certain event and to explain it as a function of a set of covariates (Cox, 1972, Blossfeld & Rower, 2002). In order to avoid inflated estimates, durations for current job searches exclude the military service for males and any kind of *post lauream* learning experience.

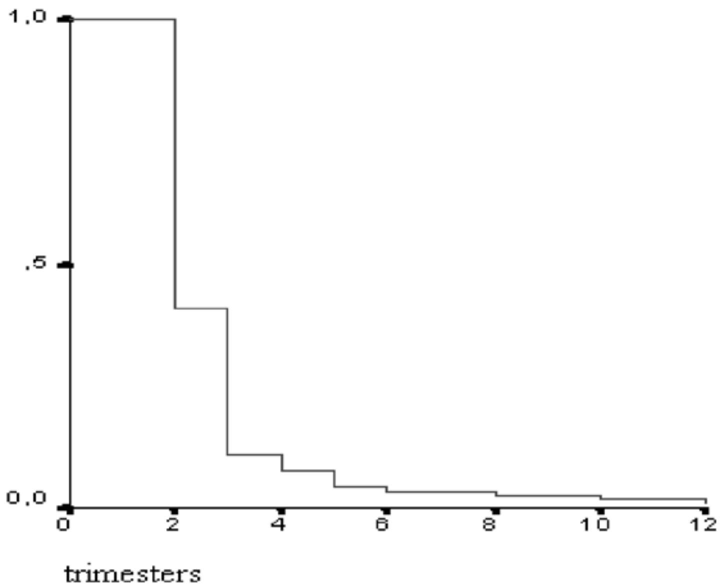


Figure 1. Survival function in the unemployed status estimated according to the actuarial method (graduates in Statistics from the University of Milan–Bicocca unemployed at graduation)

5 In fact, its designation varies in relation to the disciplinary field it is applied to; henceforth it is meant as a synonymous of *hazard models.*

The economy in the metropolitan area is flourishing, with good chances of work, giving the overall survival function its expected 'L' shape (Figure 1). The probability of entering the current position equals 0.59 after six months, 0.80 after nine months, and 0.92 after one year, while the median time is five and a half months.

2.2 Educational choices and transition times

Academic titles were grouped according to both the length of the course of study (two-year and three-year diploma, and four-year 'laurea' programme before the ministerial reform, three-year bachelor programme afterwards[6]) and the disciplinary field (economic, demographic, or purely statistical). In each disciplinary field, students can perform an internship, whilst after graduation she/he can follow *post lauream* courses or work/study abroad.

The differential impact on the labour market, resulting from the range of undergraduate choices, may be transformed into the estimation of differential entry timing for each value of a covariate. In our case, the violation of the assumption of proportional hazard functions implies that the relative risk (the ratio of risks for two distinct values of the same covariate) is not constant in time and thus the semi-parametric Cox model may not be applied.

In order to perform such differential analysis, stratification for academic title, disciplinary field, and both has been carried out, so as to compare stratified survival tables (Blossfeld & Rohwer, 2002).

When hazard differs from one group to another, it needs not to be significantly homogeneous. The statistical test for verifying the equality of risk among strata, applied here in Gehan version, is:

$$\chi^2_{(k-1)} = \sum_{j=1}^{J} \frac{\omega_j \left[D_j - E(D_j)\right]^2}{\omega_j^2 \, var(D_j)} \qquad (1)$$

where:

- j (j=1,2,..,J) is the j^{th} survival table or the j^{th} stratum,
- ω_j are weights, equal to 1 in log-rank test and to the number of subjects at risk of experiencing the event in Gehan test (or generalized Wilcoxon test),
- D_j is the observed number of entries in current professions,
- $E(D_j)$ is the expected number of entries in current professions.

Results provide evidence that the disciplinary choice renders a degree more expendable, particularly when the economic-statistics field and the purely statistical one are compared (Figure 2)[7]. No indicators are found in other direc-

---

6 At the date of the survey, the graduate courses of studies have not been introduced long enough to be concluded by any students.

7 Unfortunately, the demographic-social field is not directly comparable as it was the only one at the Faculty foundation and the others offered only some years later.

tions: the internship, the kind of academic title or the software sophistication generate respectively analogous effects. In replicating the survey, as well as suitable missing data treatment, it would be a great advantage to adopt specific questionnaires, such as the *Event History Calendar interviewing* (Conway, 1996) in order to exploit the information intensively.

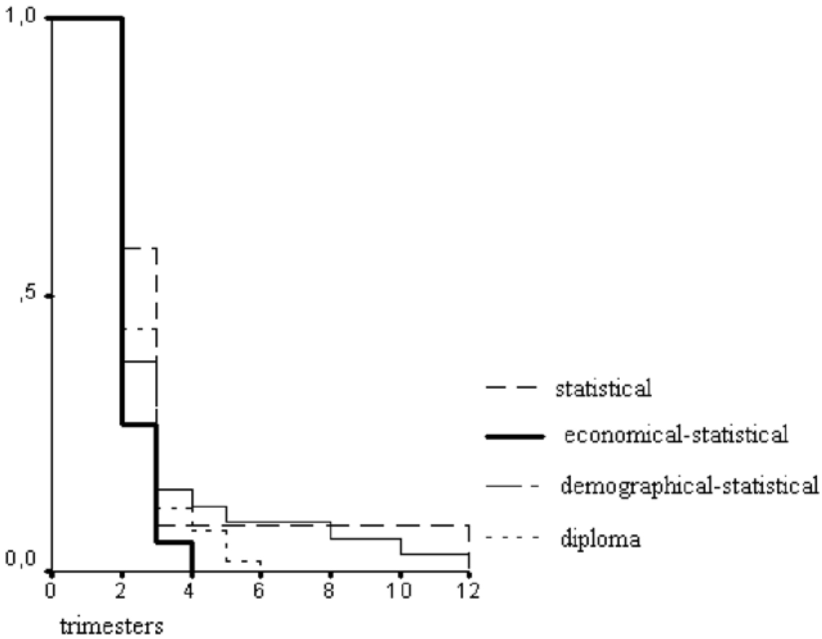


Figure 2. Survival functions in the unemployed status for each disciplinary field, estimated according to the actuarial method (graduates in Statistics from the University of Milan – Bicocca unemployed at graduation)

3. The graduate statistician in the corporate world

A global view of the qualifications and skills that newly graduated statisticians put into practise in their professional life can be achieved by means of comparative analysis, after quantifying categorical variables though ‘optimal scaling’. This procedure projects categories and qualitative scores on a continuum, in order to process quantitative data.

Let $\mathbf{Y}_{\mathbf{nxm}}$ be the observational matrix for n subjects and m variables Y_j (j=1,2,.., m). A variables assumes k_h categories (h=1,2,.., K_j). Let 1 or 0 be the value of x_{jh}, whether the i^{th} subject (i=1,2,…, n) shows or not the h^{th} value on variable Y_j. A vector x_{jh} is generated by assigning these values to a subject. Iterating this process for all K_j values of variables Y_j, the complete indicator matrix $\mathbf{X_j}$ is obtained. Extending the procedure to all *m* variables, matrix $\mathbf{X}=(\mathbf{X}_1\ \mathbf{X_2}, \ldots, \mathbf{X_m})$ is obtained. The scaling parameters for variable Y_j, $\phi_j=(\phi_{j1}\ \phi_{j2}\ \phi_{jKj})$, permit the correspondence of $\mathbf{X_j}$ with $\mathbf{X^*_j}$ as a 'scaled' value of the i^{th} subject for the j^{th} variable:

$$x^*_j = \sum_h \phi_{jh}\, x_{jh}\ , \tag{2}$$

and therefore allow one to compute the individual score matrix for scaled variables $\mathbf{X^*} = (\mathbf{X^*_1}, \mathbf{X^*_2}, \ldots, \mathbf{X^*_m})$.

Certain scaling techniques can be performed while applying a multivariate statistical method, without any distributive assumptions for the variables involved. In our case, categorical principal component analysis (CATPCA) is applied to ordinal data: alternating least squares (De Leeuw *et al.*, 1976) are utilized as an optimal scaling method for the estimation of the covariance of quantified categorical data.

3.1 Just one statistician or many statisticians?

The dimensional reduction of the data provides three factors, linked to the dynamics of entering the labour market (Table 2).

The first dimension represents formal educational qualifications as determined by the academy, such as the type of title, the disciplinary field and final graduation marks. The second represents specialized competencies actually applied in an enterprise during the first years of work, in terms of knowledge, both theoretical and applicative. The two components have equal weights and

Table 2. Factor loadings[a] and reliability of the model (graduates in Statistical Sciences at the University of Milano – Bicocca employed in November 2003)

	Dimensions		
Aspects	*1*	*2*	*3*
Software	0.376	0.815	-0.191
Statistical methodology	0.275	0.774	-0.160
Waiting time	-0.188	0.001	0.709
Entry channel	0.161	0.314	0.839
Graduation mark	0.578	0.259	0.120
Academic title	1.015	- 0.352	0.027
Disciplinary field	1.010	- 0.359	0.017
Model summary: dimensions	*Cronbach's Alfa*	*Explained variance / total variance*	
1	0.729	0.473	
2	0.473	0.299	
3	0.259	0.228	

[a] *Loadings refer to the quantification of each value of a variable, expressing the incidence of such quantification on the detected dimensions, in terms of linear regression. Increases in weight correspond to an increasing association with the dimension.*

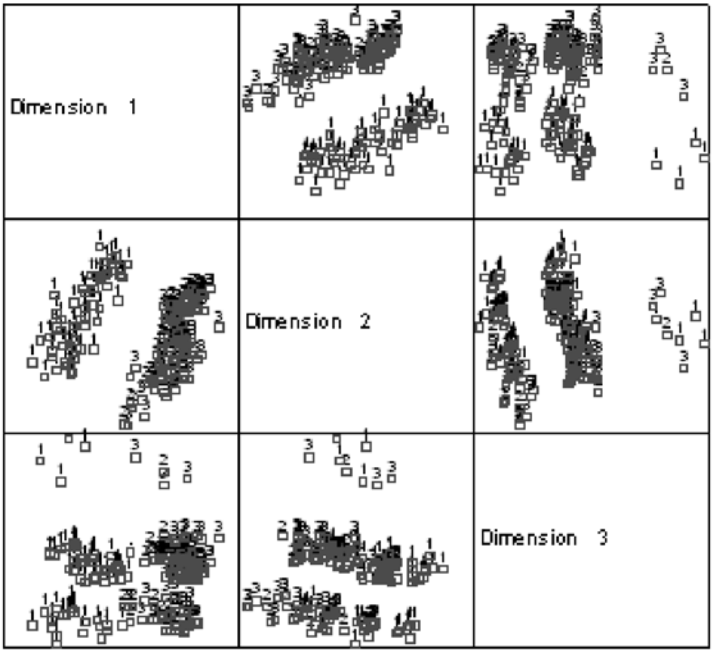


Figure 3. Graphical representation of neo statisticians (graduates in Statistics from the University of Milan – Bicocca employed at interview)
Subjects, represented in each of the first quadrant of the Cartesian plane, can be distinguished according to study programme, 1=diploma, 2= three-year bachelor, 3= four-year 'laurea'.

overlapping graphical representation, highlighting the absence of any distinction between 'theory and praxis' in the working environment. The third factor affects the network graduate statisticians use in looking for a job.

Altogether, the three dimensions adequately explain the characteristics of the employed subjects, as the value of Cronbach's alpha[8], 0.956, demonstrates.

The projection of individual CATPCA co-ordinates on the Cartesian plane indicates the presence of two groups (Figure 3), a first one composed of strictly job-oriented diplomas, and a second one in which three-year and four-year bachelor degree are barely distinguishable. This representation pictures the first set as a sub-par academic experience, somehow of subaltern rank. Three and four-year degrees are very close to each other, hence the so-called "3+2" reform seems to have maintained the same image of the graduate statistician despite the change in duration of the programs.

Actually, the formal description of the 'laurea' corresponds to a slightly higher mastery of complex analytical and computer skills, whilst channels for entering the labour market are the same.

---

8 Cronbach's alfa is an index of internal coherence. It is based on the total eigenvalue and is referred to cardinal variable, either such in origin or scaled.

4. Conclusions

The picture painted so far suggests that students are aware of employers' expectations that they attempt to satisfy by exploiting competences gained at university. In this sense, the internship (particularly if preceding the graduation) is a useful tool for broader co-operation between the study programme and the entrepreneurial world, abetted by the reduced number of students enrolled at the Faculty of Statistics.

The multidisciplinary image of the statistician favours business and economics opportunities, with a more modest requirement in knowledge than the one university provides. Statistical techniques actually used in current job positions seem to confirm a stereotypical academic education veined by speculation that is neglected, at least initially, by enterprises.

On their side, newly graduated statisticians apparently face these situations nonchalantly, with no impact on their overall job satisfaction.

References

BLOSSFELD H.P., ROHWER G. (2002) *Techniques of Event History Modelling, 2nd Edition*, Laurence Erlbaum associate, London.

CONWAY M.A. (1996) Autobiographical knowledge and autobiographical memories. In: RUBIN D.C. (ed) *Remembering our Past: Studies in Autobiographical Memory*, Cambridge University Press, New York: 67-93.

COPPI R., ERBA A. (2002) Prospettive del laureato in statistica sul mercato del lavoro, *Il giornale del SISTAN*, **20**: 14-15.

COX D. R (1972) Regression models and life tables, *Journal of the Royal Statistical Society (B)*, **34**: 187-203.

DE LEEUW J., YOUNG F., TAKANE Y. (1976) Additive structure in qualitative data: an alternating least squares methods with optimal scaling, *Psychometrika*, **41**: 471-504.

FABBRIS L. (2003) *Il monitoraggio occupazionale e professionale dei laureati e dei diplomati dell'Università di Padova*, Quaderno Pharos n.6/2003, Università degli studi di Padova, Padova.

ISTAT (2001-2004) *Università-lavoro. Statistiche per orientarsi*, ISTAT, Roma.

KIERS H.A.L. (1991) Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables, *Psychometrika*, **56**: 197-212.

MARTINI M. (1997) Quale cultura per lo sviluppo economico? *Persone e imprese*, **2-3.**

OTTAVIANI M.G. (2004) Statistici nel mondo del lavoro. Linee guida per la progettazione di un questionario che ne rilevi le competenze nella vita professionale. In: AURELI C.E. (ed) *Strategie metodologiche per lo studio della transizione Università-lavoro*, Cleup, Padova: 311-322.

UNIONCAMERE - MINISTERO DEL LAVORO E DELLE POLITICHE SOCIALI (1999, 2001, 2004) *Sistema informativo Excelsior*, Unioncamere, Roma.

Accept or Refuse a Job? How Graduates Decide

Mario Bonamin, Luigi Fabbris[1]

Statistics Department, University of Padua, Italy

Summary. In this paper, we discuss aspects of a behavioural model of new graduates on the labour market. The model regards the search for, possible refusal, and acceptance of a job by graduates of the University of Padua. Some jobs are considered a makeshift with respect to expectations the graduates matured along with degree achievement. We will analyse data collected by the University of Padua through a long-lasting perspective survey on its own graduates. We compare, in particular, the behaviour of graduates who worked before graduation with that of graduates who did not.
Keywords: Job search; Job refusal; Makeshift job; SHOT Model; Survey on graduates; University of Padua,

1. How graduates choose a job

After degree achievement, a graduate has to define his or her own life strategies. The first decision is that of deciding if, and when, either look for a job, or enter higher university courses, or start an internship, or, otherwise, settle down and create a family before deciding about work and further studies.

Whatever the schedule, the graduate has to face the job market. Fabbris *et al*. (2002) represents this hierarchy of decision with a model, named SHOT – *Search for, Hold back Opportunities, Try again*, where the main choices are made explicit and the graduates who worked before graduation are kept apart from the others who look for a non-episodic job after degree achievement.

The application of SHOT model is appropriate to University of Padua graduates because the local economy is florid and labour market is near full

[1] The authors designed and realised jointly the work described in this paper. However, M. Bonamin wrote the Sections 3 and 4 of the paper whereas L. Fabbris wrote the other sections. The authors wish to thank the anonymous referees for their contribution to a first draft of this paper.

employment and, so, graduates can really choose. On the other side, they have to be competitive and possess, or be willing to achieve, the required competences.

Regarding those who offer work, the graduates are often faced with more than one occupational opportunity, not to just obtain an income, but also to realise professional roles that exploit his/her formative investments.

In such a social environment, graduates feel they could refuse some job offers and accept the most consistent with his/her strategic expectations. In the following, we will discuss the graduates' choice parameters according to the study curriculum and social and personal characteristics of graduates.

For graduates who worked before degree the decision to maintain the current occupation – often chosen to get an income but often different from their expectations – may proceed along with the search of a new position. The novel position may be either a new role within their company, or a new job in another company. Besides, the change of position will not be an issue in this paper.

In the least favourable situations, if the gap between the offered job and the graduate's expectations cannot be filled, he/she may consider the opportunity to look for further education. Even this issue is left aside in this paper.

In the following, we focus on movements of bachelor and master graduates in the labour market (Section 3), the work roles they accept or deny (Section 4) and the possible makeshift position towards low-competence jobs if they are not able to find one that suits their expectations (Section 5).

We conclude in Section 6 with some issues that may weave the plot of the arguments developed in our analyses and may ground future research hypotheses on graduates' work roles and relationships with their companies.

2. The survey on University of Padua graduates

The University of Padua (Fabbris, 2003) has collected the data we analysed with a perspective survey started in 2000 and ended in 2005. Delayed cohorts selected from the graduates list of five adjacent graduation sessions (Figure 1) form the panel.

Cohort	…2000…2001…………2002…………2003…………..2004…………..2005
1	S ------ 1 ------ 2 ------ 3 ------ 4 ------ 5 ------ 6
2	S ------ 1 ------ 2 ------ 3 ------ 4 ------ 5 ------ 6
3	S ------ 1 ------ 2 ------ 3 ------ 4 ------ 5 ---
4	S ------ 1 ------ 2 ------ 3 ------ 4 -----
5	S ------ 1 ------ 2 ------ 3 ------ 4

Figure 1. Panel design of the University of Padua CATI survey (S=sampling)

Interviewing of cohorts was realised in two stages:

1. at time “zero”, i.e. at graduation, a web questionnaire was completed by all students who asked to discuss their graduation thesis. This CAWI – *Computer Assisted Web-based Interviewing* procedure generated a database where general data and references (mail, telephone, and internet) of students were made available. A stratified sample of graduates was selected from this database;
2. every six months after graduation, for three years, the panel was contacted by telephone with a CATI questionnaire. So, all panellists were contacted four times and most six times, apart from the time-zero occasion. Interviews have been realised by the Statistics Department of the same University by means of a system named CAPTOR – *Computer Assisted Personal Telephone and Oneself Reporting* (Capiluppi, 2000).

The data we analysed concern all interviews within three years of data collection. To this data the time-zero interview was linked. The size of the samples are 2954, 2955, 2847, 2763, 2630, 1467, 735 respondents at the six semester-cadenced occasions. The sizes differ due to cumulative non-responses.

The cohort size was designed to give reliable estimates for single faculty. Faculties themselves defined the estimates’ reliability. So, faculties were asked to add (a sum for interviewing) a number of sample units sufficient for the stated reliability target. This way, units entered the sample with varying probabilities and weights had to be used for unbiased estimation (Fabbris, 2003).

With the data collected, it is possible to organise either a cross-sectional database, or a longitudinal one. The cross-sectional database may be organised as in Table 1, where graduates interviewed at the same time from graduation are put together to define a set six months away, and others twelve, eighteen, and so on, months away.

The longitudinal data are specific for the analysis of time-dependent phenomena. The linkage of panel data collected at subsequent time points allows the analysis of data in a dynamic, rather than cross-sectional, perspective and

Table 1. Databases to be organised with the survey on the University of Padua (P=Population; S=Sample)

Cohort	*Time*						
	0	*1*	*2*	*3*	*4*	*5*	*6*
1	P_1	S_1	S_{1*}	S_{1*}	S_{1*}	S_{1*}	S_{1*}
2	P_2	S_2	S_{2*}	S_{2*}	S_{2*}	S_{2*}	S_{2*}
3	P_3	S_3	S_{3*}	S_{3*}	S_{3*}	S_{3*}	
4	P_4	S_4	S_{4*}	S_{4*}	S_{4*}		
5	P_5	S_5	S_{5*}	S_{5*}	S_{5*}		

(*) A star indicates the survivors of the sample selected for time 1 interview.

the estimation of individual *ceteris paribus* variability that is unconditional to variables whose effect could disturb our analysis.

The population was stratified according to the following characters:

- *Faculty*. Padua University faculties are: Agriculture, Economics, Pharmacy, Law, Engineer, Humanities, Medicine, Veterinary, Psychology, Pedagogy, Science, Political Science, and Statistics;
- *Bachelor or master programme*. Master graduates in Medicine are not included in the sample because they need a post-lauream period of specialisation at University so long that they are useless for this survey purposes.

The data were processed with SAS package (SAS Institute, 1999, 2000) at the Department of Statistics, the University of Padua.

3. Contacts and contracts with companies

The largest part of university students is nearly exclusively devoted to their studies until receiving their degree. However, a substantial number of students (31%), dedicate some of their time to work during their university studies (Table 2). If a minor part of their time is dedicated we refer to them as *'student workers'*, and if they dedicate a major part, '*working students*'.

Some student workers do "odd jobs", that is to say jobs with normally unstable contracts, often low profile and paid in cash, however economically useful for completing their studies. All students show very different behaviour upon receiving their degree:

- The 'working students' try to utilize their degree as soon as it is received and focus on career advancement or a change of professional duties within the company for which they work. Otherwise, in order to reposition themselves on the market, they must undergo a new selective process and, in various cases, suffer a period of unemployment. At six

Table 2. Percentage distribution of graduates of the University of Padua, according to conditions concerning work during university studies and the time passed since graduation

Conditions	*Time passed since receiving degree (months)*						
	0	*6*	*12*	*18*	*24*	*30*	*36*
Employed	31.1	59.5	68.8	76.8	82.2	87.2	88.7
Unemployed	68.9	39.8	30.4	22.5	16.8	11.5	8.8
Other condition	=	0.7	0.8	0.7	1.0	1.3	2.6
Total (*n*)	100.0 (2954)	100.0 (2955)	100.0 (2847)	100.0 (2763)	100.0 (2630)	100.0 (1467)	100.0 (735)

months from graduation, 18% of the new graduates that had been working are unemployed (Table 3). The percentage drops a little at 12 months to 16.9% and at 24 months to 9%. The proportion of employed among the unemployed graduates at graduation tend to converge soon with that of the student workers in the same period (Figure 2). The act of receiving a university degree therefore standardises the expectations of employment and career of the graduate, enforcing the professional consideration of the new graduate already inserted in activities for which a high school degree is sufficient to those who have never been employed and about productive companies have a merely hypothetical representation.

- *The proportion of unemployed graduates that find employment in the first few months after receiving degree exceeds 49%.* Please note closely

Table 3. Percentage of graduates of the University of Padua that work at various dates, according to the conditions with respect to work at the moment of receiving their degree and for the time passed since receiving degree.

Upon Receiving Degree	*Time passed since receiving degree (months)*					
	6	*12*	*18*	*24*	*30*	*36*
Employed (n=876)	82.0	83.1	86.9	91.0	92.9	92.8
Unemployed (n=2079)	49.6	62.7	72.5	78.4	84.3	86.2
Total (n=2955)	59.5	68.8	76.8	82.2	87.2	88.7

Figure 2. Percentage of graduates of the University of Padua that work at various dates, according to the conditions with respect to work at graduation and time passed since receiving degree

that this is a gross amount, and includes a denominator of those that have decided to continue the formative activities with other courses for degrees, for masters, or of a university specification, in Italy and elsewhere. The percentage of new graduates employed raises rapidly to 62.7% at 12 months, to 78.4% at 24 months and to 98.4 at 36 months, always a gross number of the formative activities over a long period.

- *At just a year since receiving a degree, about 90% of those that have found work act as if their position were stable.* The tendency to stabilise at work is estimated based on the proportion of employed graduates for at least two successive surveys. The proportion of those employed that last for at least six months is 82% following six months from graduating, to 87.7% at a year and to 95.5% at 24 months (Table 4).
- *If for no other reason except to try, all seek employment after receiving university degree.* Among those who do not find it within a year or so, over half decide to obtain ulterior formal instruction or create the situation to be able to register for a professional board. In the first six month there are, instead, many frequent brief formative programmes, for professional refinement (English courses, computer skills, and others), or stage *post lauream*, eventually associated to an European Social Fund (ESF-FSE) funded course.
- *6.9% of graduates employed within six months from their degree change occupation.* The proportion falls rapidly at just a year from receiving the degree (1.4%) and oscillates about 2% between 12 and 30 months (Table 5). It is not necessary to conduct a specific survey to imagine that for many graduates it was not easy to find their first employment, many of them had to undergo recruitment interviews and compete with other new graduates for the same job. In conclusion, while those that have found a job immediately after their degree could be presented within a brief time period a possibly better position, those that find employment later do not feel like risking a change and accept more easily an offered employment position.

Table 4. Percentage of graduates of the University of Padua that work at various dates, according to their conditions with respect to work at a period of 6 months prior and the time passed since receiving degree

Six months prior	*Time passed since receiving degree*					
	6	*12*	*18*	*24*	*30*	*36*
Employed	82.0	87.7	93.7	95.5	96.7	98.4
Unemployed	49.6	41.6	37.8	38.5	39.9	23.6
Total	59.5	68.8	76.8	82.2	87.2	88.7
(n)	(2955)	(2847)	(2763)	(2630)	(1467)	(735)

Table 5. Percentage of University of Padua graduates that have changed employment at various times since receiving degree, for conditions with respect to the employment 6 months prior, type of work relation, and activity of graduates

Six months upon receiving degree	*Months passed since receiving degree*					
	6	*12*	*18*	*24*	*30*	*36*
Relation of Employment						
Normally Hired	2.8	1.3	1.0	1.4	0.6	3.2
Independent Contractor	7.2	0.8	1.4	2.9	0.3	7.6
Unusual Contract	4.9	1.6	3.3	5.9	2.6	7.7
For the activity of a graduate						
Any university degree sufficient	3.1	1.3	1.2	2.2	1.3	4.3
Bachelor degree sufficient	3.0	1.6	2.4	1.1	=	=
Specific degree required	4.2	1.2	1.6	3.5	0.8	5.4
Worked 6 months prior and changed work	6.9	1.4	1.5	2.5	0.8	4.5
(n, worked 6 months prior)	*(710)*	*(1469)*	*(1794)*	*(1884)*	*(1160)*	*(619)*

- *There are no longer initial gender handicaps.* In various researches (see among others, Fabbris *et al.*, 2002) it was revealed that, in the productive contexts of substantial full employment, like the area in which a graduate of Padua would fit in, the percentage differences between male and female graduates are within one percentage point (Table 8).
- *Some difference exists for the time frame of employment between Master and Bachelor graduates*, in that the bachelor graduates are more immediately hired of the two.
- *The employment research channels followed by graduates are multiple.* A graduate does not want to preclude himself any possibility and tries every channel, public or private, to find employment. Not excluding the Placement Office of the University of Padua, the personal acquaintances, family, friends, business colleagues and the professor with whom the thesis was realised to have useful contacts for a job interview (Boaretto *et al.*, 2006). There are those who, less secure of their means, use even repeatedly the same channels, as if the number of attempts could increment the probability of success, and others that search only in the places in which they have a high chance to find employment. On average, the graduates test research routes that over 46% are three groups of channels, public, private and foreign (Table 6). The proportion of attempts falls rapidly with the passing months, even if the speed of the decline is different for those that work with respect to those that are still looking for employment. For all, the progressive knowledge of the rules of the market, which comes with the experience of searching for employment and

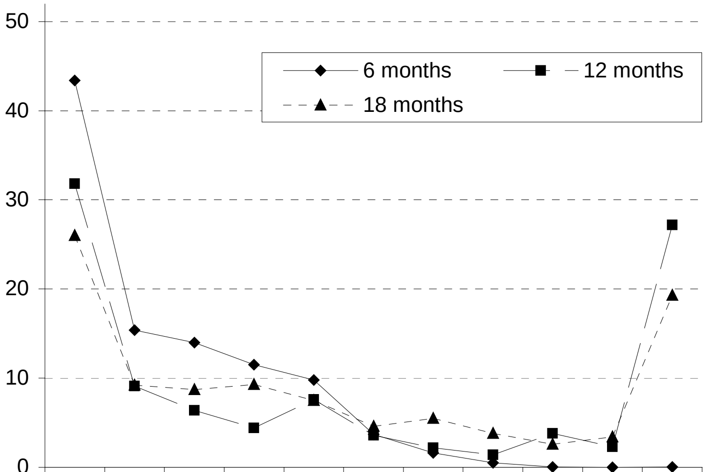
Table 6. Average number of employment channels attempted by University of Padua graduates and percentage not researching actively employment, for conditions with respect to employment at graduation, six months prior and time passed since receiving degree (between parentheses the size of the sample)

	Months passed since receiving degree					
	6	*12*	*18*	*24*	*30*	*36*
Upon receiving degree						
Employed	1.24 (876)	1.00 (840)	0.72 (827)	0.40 (781)	0.36 (493)	0.28 (269)
Unemployed	1.61 (2081)	1.27 (2007)	0.87 (1936)	0.54 (1849)	0.47 (974)	0.37 (466)
Six month prior						
Employed	1.24 (876)	0.87 (1683)	0.62 (1916)	0.32 (1974)	0.28 (1203)	0.21 (631)
Unemployed	1.61 (2081)	1.64 (1146)	1.29 (821)	1.07 (633)	1.20 (255)	1.18 (88)
Sector of activity						
Public	2.95 (432)	3.45 (328)	3.08 (226)	2.64 (138)	2.64 (71)	2.47 (25)
Private	2.54 (655)	2.45 (532)	2.47 (287)	2.45 (143)	2.21 (52)	2.29 (25)
Foreign	2.60 (289)	2.56 (205)	2.86 (116)	2.34 (69)	2.41 (46)	2.36 (16)
Self Employed	0.00 (22)	0.00 (17)	0.04 (16)	0.00 (13)	0.00 (7)	0.00 (7)
Total	1.50 (2957)	1.19 (2847)	0.83 (2763)	0.49 (2630)	0.43 (1467)	0.34 (735)
% not researching actively employment	43.7	56.6	67.5	79.1	81.7	85.3

with the tightening of the perception of the possibilities of success, brings a simplification to the ways to effect the search. At 18 months from obtaining a degree, who already works gives slight attempts in only one direction (0.6 channels tested between 12 and 18 months), who does not work depends even more on public placement and maintains contacts with the private one, but the research is more specific (1.3 channels between 12 and 18 months).

The multiple regression analysis (Fabbris, 1997), applied with the number of research *channels* tested in the first six months from graduation as criterion variable and forcing the faculties as predictors[2], arrives at the following considerations (Table 7):

[2] Faculty estimates refer to the Faculty of Political Science.



	0	1	2	3	4	5	6	7	8	9	10+
6 months	43,4	15,4	14	11,5	9,8	3,7	1,6	0,5	0,05	0	0,02
12 months	31,8	9,1	6,4	4,4	7,6	3,6	2,2	1,4	3,8	2,3	27,2
18 months	26	9,2	8,7	9,3	7,5	4,6	5,5	3,8	2,6	3,4	19,3

Figure 3. Number of active research channels tested by University of Padua graduates, according to time passed since receiving degree

Table 7. Synthetic results from the application of the analysis of regression referred to the number of research channels of employment tested by graduates of the University of Padua in the first 6 months since obtaining degree

	Estimate of β	*s(β)*	*p value*
Intercept	-0.546	0.283	0.054
Agriculture	1.582	0.282	< 0.001
Economics	1.639	0.298	< 0.001
Pharmacy	0.758	0.283	0.008
Law	1.087	0.273	< 0.001
Engineer	1.449	0.270	< 0.001
Humanities	1.837	0.268	< 0.001
Medicine	0.566	0.301	0.060
Psychology	0.607	0.273	0.026
Pedagogy	1.329	0.275	< 0.001
Science	1.571	0.266	< 0.001
Political Science	2.026	0.279	< 0.001
Statistics	2.189	0.288	< 0.001
Gender 1=M 2=F	0.436	0.074	< 0.001
Employed at 6 months	0.259	0.070	0.002
Employed at 0 time	-0.446	0.073	< 0.001

- *The behaviour of the graduate on the employment market depends on the department of origin.*
- *The female graduates search for employment more hardly than men do.* If the number of tested channels is a reflection of the confidence of their own means, the male graduates are shown to be more confident than the female graduates are. It should be noted that female students numerically prevail in the humanities and that the degrees supplied by these faculties are “weaker” in the research for a first employment. Hence, the faculty obscures the analytical relevance of the gender of the graduates, meaning to say that the major mobility in the employment research is observed between feminine *vs* non-feminine faculties.
- *The unemployed do research that is more active at time of the degree, while those that already had a job during their studies search in a more specific way.* At six months from obtaining the degree, those that had a job are searching more than those that did not have one. That is probably because the former ones already had a taste of the employment market, while many of the latter ones, originally uncertain whether to prolong their studies or search intensively for a job, may have decided to enter another period of training at the university. Hence, six months from obtaining a degree are enough for Master or Bachelor graduates to define their life strategies.

4. The refusal of offered jobs

In conclusion of the processes of the employment research, a good part of the graduates of the University of Padua is employed.

At six months, 82% of the Bachelor graduates are employed and 58% of the Master graduates. Altogether, 61% of those who obtained a degree at the Uni-

Table 8. Percentage of University of Padua graduates that work, according to the time span since obtaining the degree, gender and type of degree

Type of degree		*Months passed since receiving degree*						
		0	*6*	*12*	*18*	*24*	*30*	*36*
Males	*Master*	28.2	58.9	69.8	76.3	81.6	87.2	89.2
	Bachelor	38.7	71.3	76.9	81.1	88.8	94.9	97.4
	Total	29.5	60.4	70.7	76.9	82.5	88.2	90.4
Females	*Master*	30.0	55.0	64.4	74.4	80.3	84.9	85.1
	Bachelor	41.7	90.2	91.3	94.2	95.5	95.4	96.5
	Total	31.3	58.8	67.5	76.8	82.1	86.4	87.2
Total	*Master*	29.3	56.7	66.7	75.2	80.9	85.9	87.0
	Bachelor	40.3	81.7	84.9	88.7	92.6	95.2	96.8
	Total	30.5	59.5	68.8	76.8	82.2	87.2	88.7

Table 9. Percentage of graduates of the University of Padua that have refused at least one employment offer in the first six months after obtaining a degree, according to the study programme, conditions at the time of degree and motive for refusal

In that period	*At time of obtaining degree*					
	Employed		*Not Employed*		*Total*	
	Master (n=92)	*Bachel. (n=7)*	*Master (n=450)*	*Bachel. (n=40)*	*Master (n=542)*	*Bachel. (n=47)*
worked	4.2	12.5	2.7	1.8	2.9	3.2
were not interested	33.3	12.9	28.6	25.9	29.4	24.3
were precarious	0.0	0.0	2.2	0.0	1.9	0.0
were inconsistent with study	5.0	0.0	4.0	3.6	4.1	3.2
were inadequate for degree	0.8	0.0	4.2	5.4	3.7	4.7
were far from home	2.1	0.0	2.9	3.9	2.8	3.4
wanted to continue studies	15.6	12.9	24.6	9.7	23.2	10.1
other motives	57.3	74.5	43.2	63.0	45.6	64.4

versity of Padua are employed. At 12 months from the degree, the percentage of graduates rises to 70% and at 24 months to 82% (Table 8).

The insertion in employment is a bilateral process of offer and acceptance/refusal of a job. In general, a company proposes job positions to the applicants and they refuse some offers. In the first six months from obtaining a degree, half of those that leave with a Master or a Bachelor degree from the University of Padua refuse at least one offer of employment (Table 10).

Table 10. Percentage of graduates of the University of Padua that, at different times from obtaining degree, have refused at least one employment offer, according to the employment condition six months prior

	Six months prior		
	Employed	*Not employed*	*Total*
At 6 months from obtaining degree			
Employed	52.8	47.1	49.4
Not employed	50.6	49.0	49.2
Total	51,1	48,7	49.2
At 12 months from obtaining degree			
Employed	67.5	41.4	55.9
Not employed	50.9	35.2	38.6
Total	55.9	38.5	38.8
At 18 months from obtaining degree			
Employed	0.0	40.5	34.9
Not employed	41.3	32.1	33.9
Total	34.9	33.9	33.9

The proportion of refusals declines with the passing of time, being 39% between the 6 and 12 months and of 34% in the successive six months. The motives of the decrement are double:

a. the decrease of interest of the companies for the graduates that have not yet found employment after a certain time since obtaining degree,
b. the decreasing energy of the graduates after repeating the unsuccessful trials of employment research.

The insufficient satisfaction for the accepted job has different causes, as Master or Bachelor graduated categories (Table 11):

- *The bachelor graduates show a low propensity to continue their studies,* both in the cases that before obtaining the university degree they were in employment and that they were without employment experience. The employment offers for the bachelor graduates already in service are rare enough because they must be more advantageous than the position addressed (which is coherent with the possession of a high school degree), and, actually because does not involve elevated positions, are a often regulated by stiff contracts. The refusal of the bachelor graduates is, therefore, much more than those of the master graduates, tied to practical considerations, of existential organisations, and 'affected' toward the owning company.
- The master graduates show behaviours however differentiated according to whether they were employed or not at the time of the degree. Those that were employed, having already their own problem for the existence solved, are able to wait for an interesting occasion to change employ-

Table 11. Percentage of graduates of the University of Padua that have changed an employment activity in the first six months after obtaining degree, according to study programme, period in which they changed the activity and motive

Motives	*Referral period*					
	First 6 months		*6-12 months*		*12-18 months*	
	Master (n=417)	*Bachelor (n=80)*	*Master (n=277)*	*Bachelor (n=28)*	*Master (n=240)*	*Bachelor (n=27)*
Just wanted to change	0.3	0.0	0.0	0.0	0.6	5.6
Distance from residence	0.0	0.0	0.3	0.0	0.0	0.0
Inadequate retribution	0.9	3.0	3.0	2.3	2.3	0.0
Professional duties low	3.4	1.6	3.4	12.6	7.9	6.2
Unstable employment	4.3	5.6	3.7	1.5	5.7	22.7
Family problems	0.0	1.7	1.1	0.0	1.1	2.6
Inadequate internal rapport	23.5	16.9	25.2	15.1	25.4	16.2
Expired contract	3.4	1.6	3.3	0.0	3.7	4.3
Military service	0.6	0.5	2.4	0.0	3.7	0.0
Dismissal	8.8	4.9	23.9	21.6	12.1	2.5
Other motives	54.8	64.2	33.8	46.8	37.6	39.9

ment. The unemployed, instead, refuse an employment often to continue their studies. Unfortunately, we are not able to establish if the desire to continue the studies was born after the offers refused, or if it was a pre-defined choice empowered by the perception of inadequacy of the achieved degree in relation to the market. It is, in fact, plausible that the offers of employment advanced by the companies are few for the prestigious positions and many for second calibre ones. However so, the major part of the graduates refuse employment because they feel a hiatus between their aspirations and the desirability of the offer received.

A logistic regression analysis, applied on dichotomous variable "refusal of an offered job" shortly after obtaining degree (Table 12) as a criterion variable,

- confirms that the most important variable for the interpretation of the frequency of refusals is the faculty. The faculties with the highest "risk" of job refusal no matter the study programme (master vs. bachelor) and the possible position of worker student are Economy and Engineering.

Table 12. Results of the logistic regression analysis referred to the refusal of at least an offered job on University of Padua graduates six months since obtaining degree[3]

	Estimate of β	*$s(\beta)$*	*Significance*	*Estimate of ψ*
Intercept	-0.057	0.401	0.887	=
Agriculture	-0.363	0.328	0.269	0.70
Economy	0.494	0.377	0.190	1.64
Pharmacy	0.251	0.339	0.459	1.28
Law	-0.412	0.259	0.111	0.66
Engineering	0.417	0.300	0.165	1.52
Humanities	-0.119	0.278	0.668	0.89
Medicine	0.418	0.906	0.645	1.52
Veterinary	0.226	0.422	0.592	1.25
Psychology	-0.590	0.274	0.031	0.55
Pedagogy	0.244	0.363	0.501	1.28
Science	-0.219	0.254	0.387	0.80
Statistics	0.229	0.378	0.544	1.26
Gender 0=M 1=F	-0.173	0.135	0.201	0.84
Master=0 Bachelor=1	0.382	0.275	0.165	1.46
Worked (=1) at graduation	0.227	0.167	0.174	1.25

[3] Faculties have been forced into the model as explanatory variables. Faculty estimates refer to Political Science Faculty. A stepwise selection technique (Fabbris, 1997) was applied for the selection of other predictors. The other predictors we considered were: (i) employed at 6 months, (ii) age at graduation, (iii) graduation vote. None of these variables reached a significance level of 20%.

- adds a few other significant variables. The *gender*, index of a greater propensity of males to refuse incongruous jobs, *having obtained a Bachelor (vs. Master) degree,* since it is easier for an employer to make offers with a low professional content to lower training figures, *and having been a worker student* indicate that the comprehension of the functioning rules of the employment market makes one evaluate more critically the opportunity offered.

5. Makeshift working positions

For a graduate, job dissatisfaction may be caused by the acceptance of a makeshift position. It may be correlated to:

- *unqualified professional duties,*
- *precarious employment contract or remuneration,*
- *unfavourable place and conditions of employment.*

The makeshift attitude of the graduate on undesirable work positions shows various gradations of severity. For a graduate, the acceptance of waged job is probably the least desired condition. In the first six months since obtaining degree, it regards 2.3% of those that have obtained a degree from the University of Padua. Fortunately, the proportion is limited and evaporates with the passage of time, it is reduced, in fact, to 1% at one year and to 0.5% at eighteen months after obtaining degree (Table 13).

The work positions that do not require a university degree are a step lower on the scale of severity of dissatisfaction causes. At six months, 14% of the graduates of the University of Padua affirm that their work could be done by anyone even without a high school degree. Since this proportion remains at 10% at 18 months, it may concern the nature of the activities involved and not an undesired job contract.

During the insertion stage, there are frequent references to atypical contracts, or time-limited contracts, or even working without a contract. These conditions regard one in three activities within 18 months from the degree achievement. Since many employment contracts that pre-exist to graduation are stable, it implies that the major part of the “ex-novo” jobs is with precarious or atypical contracts[4].

Some graduates undergo a limitation in remuneration instead of a penalization of contract. Less that 600 Euro per month, that represents about half of the average pay, is the earning of 11% of the graduates at six months from obtaining degree, of 6% at one year, and of 4% at 18 months. The graduates that have to face this limitation are included in the periods of apprenticeship, for

4 In the handicraft firms, it is not rare to have an offer of an apprentice contract even to engineers destined to do qualified activities.

Table 13. Indicators of acceptance of makeshift positions on the part of Master or Bachelor graduates of the University of Padua in the first 18 months from obtaining degree, according to type of degree and working conditions at time of graduation

	months since degrees	*% factory workers*	*% degree not necessary*	*% work contract precarious*[5]	*% remunerated to 600 euro*	*% work outside region*
Padua University graduates	*6*	2,3	14,2	32,7	11,3	7,6
	12	1,0	11,4	34,3	5,9	5,3
	18	0,5	10,3	33,1	4,0	3,1
Employed at time of graduation						
Master degree	*6*	1,7	23,3	20,4	14,8	6,4
	12	0,4	14,8	21,7	6,7	4,7
	18	0,3	14,1	20,2	5,7	2,9
Bachelor degree	*6*	5,5	=	18,3	4,5	3,4
	12	2,4	=	20,6	0,3	1,3
	18	0,0	=	22,5	0,0	0,9
Unemployed at time of graduation						
Master degree	*6*	1,2	14,1	39,7	12,6	10,9
	12	0,9	12,1	38,5	7,9	8,0
	18	0,3	12,5	36,4	4,9	4,0
Bachelor degree	*6*	7,6	=	45,4	8,4	4,6
	12	0,4	=	37,7	3,8	0,9
	18	0,9	=	36,5	2,1	1,2

the admission to the State exam, with badly paid and precarious activities. The low pay couples with every other manifestation of difficulty of work insertion.

Graduates only if necessary accept a place of work far from the area of residence. In addition, the majority of them consider working with rotating hours, or during the weekends, exceptional. With the passing of time also the percentage of graduates that were initially employed outside the region diminishes, a clear symptom that the acceptance of an employment that is distant from home is a makeshift choice caused by the difficulty of finding local employment. The proportion at six months is 8%, falls to 5% at one year and to 3% at eighteen months.

6. The insertion-and-adaptation process

The research of employment opportunities, the choice between the opportunities offered, the evaluation of the employment position in rapport with the pos-

---

5 Percentage ratio of master/bachelor graduates employed with atypical or stiff subordinate contracts and total of master/bachelor employees.

possibility to integrate their own training, the family pressures, all this requires quick and definite choices to graduates. These elements occur in a sequence repeated even more than once, until the decision is made whether to enter the job market or continue with studies in order to obtain a degree that is superior or more specific for the imagined job.

The analysis of the transition to employment of the graduates of the University of Padua demonstrates that getting a job is just a matter of time. The results of the research for employment may depend on the capacity of the new graduates to present themselves adequately to who is offering employment. For the major part of the new graduates, and particularly for those that possess a degree in the humanities or the social sciences, a positive result in the research depends on their ability to adapt to the offers of the market.

In our analysis we have found that graduates who refuse certain jobs are either more prepared on a professional level, or possess a technical or economics degree, or, having worked, know the rules of the employment market. The refusal of an offered job is in direct relation with the image that the new graduates have of their own degrees. However, there is no correspondence between the graduates' ideal image and the market's opportunistic view for the same degree. For instance, the frequent refusal for first employment positions by Economy and Business graduates is not correlated with a proportional offer of high rank positions for the same class of graduates.

The adaptation of the graduates to the market is, instead, progressive. Some modest positions on the professional level, like the positioning in a job as a factory worker or similar, is in fact transitory, and may last even less than six months. A low standard of remuneration may last, for one graduate in twenty, even one or two years, however in the mid run it converges towards higher values.

Massive – above 30% also over 18 months from the insertion – are the percentages of employment with precarious contracts that goes to say atypical, for a limited amount of time, with the possibility of dismissal on the part of the employer, and even "by verbal agreement". These ways to regulate the first employment rapports are so common that their value is shared, with few exceptions, both by who asks for and by who offers employment. Actually, the most desirable employment positions are regulated by unstable contracts, while the lower ones are accompanied with longer contracts. Although it is a mass phenomenon, the instability of contracts is transitional. That is demonstrated by the progressive flow of the newly assumed toward stable positions and toward productive organizations of more ample dimensions.

References

BIGGERI L., BINI M., GRILLI L. (2001) The transition from University to Work: a multilevel approach to the analysis of the time to get the first job, *Journal of the Royal Statistical Society, Series A*, **162(2)**: 293-305.

BOARETTO A., ROTA G., SILVESTRI I., FABBRIS L. (2006) Networks of “weak” ties of Padua University graduates searching for employment. In: FABBRIS L. (ed) *Effectiveness of University Education in Italy: Employability, Competences, Human Capital*, Springer, Berlin Heidelberg New York: 123-139.

CAPILUPPI C. (2000) Il sistema CAPTOR. In: FABBRIS L. (ed) *Il questionario elettronico*, CLEUP, Padova: 227-248.

FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.

FABBRIS L. (2003) *Il monitoraggio dell’inserimento occupazionale e professionale dei laureati e dei diplomati dell’Università di Padova*, Quaderno PHAROS n. 6/2003, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova.

FABBRIS L., ROTA G., TREVISANELLO D. (2002) SHOT: un modello per la rappresentazione dei processi di ricerca del lavoro dei laureati di Padova. In: CARLI SARDI L., DELVECCHIO F. (eds) *Indicatori e metodi per l’analisi dei percorsi universitari e post-universitari*, CLEUP, Padova: 125-146.

SAS INSTITUTE INC. (1999) *SAS/STAT® User’s Guide, Version 8*, SAS Institute Inc., Cary NC.

SAS Institute Inc. (2000) *SAS/STAT® Software: Changes and Enhancements, Release 8.1*. SAS Institute Inc., Cary, NC.

A Multilevel Chain Graph Model for the Analysis of Graduates' Employment

Anna Gottard, Leonardo Grilli, Carla Rampichini

Statistics Department "Giuseppe Parenti", University of Florence, Italy

Summary. The main goal of the present paper is the analysis of the working position of graduates using multilevel and chain graph models, extended to the case of correlated data. After a brief introduction to multilevel modelling and a description of the conditional independence implied by the model, we describe chain graphs for multilevel models. The model put forward can analyse the factors influencing the graduates' job position, using the data collected on students of the University of Florence who graduated in the year 2000.
Keywords: Chain graph models; Logistic regression; Multilevel models; University system evaluation.

1. Multilevel and chain graph modelling

The university system evaluation requires *ad hoc* methods and statistical models able to capture the complexity of the phenomenon. Such a complexity originates from many facets, such as:

(*a*) the hierarchical structure of the data, entailing a correlation among the observations and requiring the consideration of effects at different levels of the hierarchy. This hierarchical structure (students within classes, classes within study programmes, and so on) is substantial for the analysis and the underestimation of cluster effects and the fact that some of the assumption of the usual regression models are not satisfied, may lead to incorrect standard errors of the estimated coefficients;

(*b*) the presence of variables referring to different moments along the students' careers (e.g. parents education, high school grades, exam grades, graduation grades). This aspect implies a logical and temporal order among the involved variables that must be taken into consideration to shed light on the way students achieve their final result (e.g. getting a job), and to distinguish between direct and indirect effects.

Multilevel models (Snijders & Bosker, 1999; Goldstein, 2003) allow us to cope with the intra-class correlation and to analyse in a proper way the cluster effects. For such reasons, multilevel models are widely applied in the education evaluation framework. Chain graph models (Cox & Wermuth, 1996) are a useful tool for the representation of the process described at point (b).

In the following, we propose a method for the integration of multilevel and chain graph models that allows:

(i) to properly model the relationships among the probability of finding a job after the degree, the students' careers and their individual characteristics;

(ii) to stress the contribution of the study programme to the student's success in the labour market;

(iii) to distinguish among direct and indirect effects of the background and career variables.

In Section 2, we describe the two-level linear model and its extension to a binary response. In Section 3 the multilevel graph model derived from the integration among the multilevel model and the chain graph model is illustrated. In the fourth Section, we present the data at hand and the main results of the empirical analysis, and in the fifth we conclude by giving some lines for future research.

2. The linear random intercept model

Let us consider a two-level hierarchical structure, where Y_{ij} is the response variable for the i-th subject (first level unit) of the j-th cluster (second level unit), $i=1,2,\ldots,n_j$, $j=1,2,\ldots,J$. For each subject, a vector $\mathbf{X}_{ij}$ of individual (e.g. gender, high school rank) and cluster (e.g. number of enrolled students for each program course) variables is available.

Let us assume Y_{ij} is a continuous variable. If the relationship between the response Y_{ij} and the covariates $\mathbf{X}_{ij}$ is linear, it is possible to specify the following linear random intercept model:

$$\begin{aligned} Y_{ij} &= \alpha_j + \boldsymbol{\beta}'\mathbf{X}_{ij} + \varepsilon_{ij} \\ \alpha_j &= \alpha + U_j \end{aligned} \tag{1}$$

where ε_{ij} are the first level residuals, while U_j are the second level ones. Residuals are assumed to be independent and normally distributed, with zero mean and variances $\mathrm{Var}(\varepsilon_{ij})=\sigma^2$ at subject level and $\mathrm{Var}(U_j)=\tau^2$ at cluster level. Moreover, as it is common in regression models, the correlations among residuals at both levels and the covariates are assumed null. The independence hypotheses among the observations following from this model are:

$$Y_{ij} \not\perp\!\!\!\perp Y_{i'j} \mid \mathbf{X}, \qquad \forall i \neq i', \forall j \tag{2}$$
$$Y_{ij} \perp\!\!\!\perp Y_{i'j'} \mid \mathbf{X}, \qquad \forall i, i', \forall j \neq j',$$

where $\mathbf{X}=\{\mathbf{X}_{ij} : i = 1,2,\ldots,n_j,\ j=1,2,\ldots, J\}$.

It can be seen from relationships (2), conditionally on the covariates $\mathbf{X}$, that observations from different clusters are independent, while observations belonging to the same cluster are dependent. The intraclass correlation coefficient

$$\rho = Corr(Y_{ij}, Y_{i'j'}) = \begin{cases} 0 & \text{se } j \neq j' \\ \dfrac{\tau^2}{\tau^2+\sigma^2} & \text{se } j = j' \end{cases}$$

measures the within-cluster dependence. Moreover, conditionally on the covariates $\mathbf{X}$ and the second-level errors U_j also the observations belonging to the same cluster are independent:

$$Y_{ij} \perp\!\!\!\perp Y_{i'j} \mid \mathbf{X}, U_j, \qquad \forall i \neq i', \forall j \tag{3}$$

For each cluster j the joint probability distribution can be factorised as follows[1]:

$$\begin{aligned} f(\mathbf{y}_j, u_j, \mathbf{x}) &= f(\mathbf{y}_j \mid u_j, \mathbf{x}) f(u_j \mid \mathbf{x}) f(\mathbf{x}) \\ &= f(\mathbf{y}_j \mid u_j, \mathbf{x}) f(u_j) f(\mathbf{x}) \\ &= \left[\prod_{i=1}^{n_j} f(y_{ij} \mid u_j, \mathbf{x}) \right] f(u_j) f(\mathbf{x}) \end{aligned} \tag{4}$$

where $\mathbf{y}'_j = \{y_{1j}, y_{2j},\ldots, y_{n_j j}\}$. Actually, from the independence among u_j and $\mathbf{X}$ it follows that $f(u_j \mid \mathbf{x}) = f(u_j)$, while for the conditional independence (3) the conditional density $f(\mathbf{y}_j \mid u_j, \mathbf{x})$ corresponds to the product of the n_j individual densities.

In general, the effect of a first-level covariate can be decomposed into two parts: within and between clusters, according to the covariate variance decomposition (Snijders & Bosker, 1999). For instance, in the linear model with only one covariate X, the OLS total coefficient $\hat{\beta}_T$ is a linear combination of the coefficient of the regression among cluster means $\hat{\beta}_B$ and of those within clusters $\hat{\beta}_W$:

$$\hat{\beta}_T = \hat{\eta}_X^2 \cdot \hat{\beta}_W + \left(1-\hat{\eta}_X^2\right) \cdot \hat{\beta}_B \tag{5}$$

---

1 In factorisation (4) it is the same if one considers the matrix of the covariates $\mathbf{X}$ of all the individuals or only the sub-matrix $\mathbf{X}_j$ of the covariates of the subjects of the j-th cluster, so, for the sake of simplicity, only the sub-matrix $\mathbf{X}_j$ is considered.

where $\hat{\eta}_X^2$ is the correlation ratio of X. As a result $\hat{\beta}_T$ assumes an intermediate value among $\hat{\beta}_B$ and $\hat{\beta}_W$. The *between* and *within* coefficients have a different interpretation and they can take opposite values. Thus, the total coefficient can be non-significant, whilst the *between* and *within* coefficients are significant but opposite in sign.

Therefore, it is better to specify the model in order to estimate both the *between* and the *within* coefficients of each subject-level covariate. A way to perform such an estimation is to insert in the model both the covariate X_{ij} and the cluster mean $\bar{X}_{.j}$:

$$Y_{ij} = \ldots + \beta_W X_{ij} + (\beta_B - \beta_W)\bar{X}_{.j} + \ldots \ . \tag{6}$$

In model (6), the coefficient of the cluster mean represents the difference among the *within* and *between* coefficients, so the usual test for the $\bar{X}_{.j}$ coefficient is to be interpreted as a test for the difference among the *within* and *between* coefficients. If the $\bar{X}_{.j}$ coefficient is not significant, the distinction among the *between* and *within* effects can be ignored, leaving among the predictors only the raw covariate X_{ij}.

Note that the insertion of the cluster mean of a covariate allows us to eliminate the possible correlation among the covariate and the random effects U_j (Snijders & Bosker, 1999).

When the response variable is binary, it is possible to assume the linear random intercept model (1) for a continuous latent variable that generates the observed binary variable Y_{ij}^{obs} as follows:

$$Y_{ij}^{obs} = \begin{cases} 0 & \text{if } Y_{ij} \le 0 \\ 1 & \text{if } Y_{ij} > 0 \end{cases}.$$

Assuming a standard logistic distribution for the first-level errors ε_{ij} (equation 1), the logistic random intercept model for the probability of response is:

$$P(Y_{ij}^{obs} = 1 \mid u_j, \mathbf{x}_{ij}) = \frac{1}{1 + \exp(-(\alpha + \boldsymbol{\beta}'\mathbf{x}_{ij} + u_j))}.$$

Many multilevel analysis textbooks (e.g. Snijders & Bosker, 1999) describe the properties of such a model.

2.1 Graphical models for hierarchical data

Graphical models are a class of probabilistic models on a set of random variables whose conditional independence structure can be represented by a graph. A graph is a mathematical object consisting of two sets: a set of nodes and a set of undirected or directed edges (arrows) between nodes. In the graph associated with a particular model, each node corresponds to a random variable in the model. Usually, a discrete random variable is depicted as a circle, ○, a continuous random variable is represented by a dot, ● , an edge between two nodes stands for an association between the variables or, more precisely, the absence of a connection between two nodes indicates conditional independence between the corresponding variables.

A chain graph admits both undirected and directed edges, and (partially) directed cycles are forbidden. This implies that, starting from a node, it is not possible to go back to it through the edges and arrows of the graph.

In a chain graph, nodes can be partitioned into an ordered sequence of blocks. Nodes in a same block can be connected by undirected edges, while only arrows can connect nodes in different blocks. The arrows stay for an asymmetric relationship between the variables, while undirected edges indicate a symmetric relationship.

A chain graph model for a set of random variables is specified by assuming that their joint distribution satisfies the chain graph Markov properties. Therefore, chain graph models are a class of probabilistic models allowing for both symmetric and asymmetric relationships between variables, assuming a sort of logical and temporal order among the variables. Each variable in a block has to be considered explanatory of the variables in the following blocks.

Because of the partial ordering among the variables, it is possible to distinguish the set of pure explanatory variables (usually in the last block on the right), from the set of pure response variables (last block on the left) and from the set of intermediate variables, that are both explanatory and responses, positioned in the intermediate blocks. In this work, we refer to the chain graph Markov properties proposed by Lauritzen & Wermuth (1989) and Fridenberg (1990). These properties are usually termed LWF Markov properties.

An important Markov property for chain graphs is the global Markov property, which is based on the definition of the 'moral graph'. Starting from a given chain graph, a moral graph can be obtained by connecting parents of common children (or children belonging to the same chain component), and then converting all the arrows into undirected edges. See, for example, Figure 1, where the graph in (b) is the moral graph of the chain graph in (a). More details can be found in Lauritzen (1996).

The global Markov property combines the concept of conditional independence to that of separation between nodes in the moral graph. If, in the moral graph, a set of nodes S separates the nodes in A from the nodes in B, that is, each path from A to B passes by some node in S, then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_S$, where $\mathbf{X}_k$ is the vector of random variables represented by nodes in k, $k = A, B, C$. Markov properties induce a factorization of the joint distribution of the variables in a model, which is useful for the inferential procedures (Lauritzen, 1996).

Graphical models are apt to represent the conditional independence relationships among a set of variables, if statistical units are independent. This assumption is no longer valid whenever the data have a hierarchical structure. Gottard & Rampichini (2004) propose to overcome this issue by representing in a graph all the variables of a generic group j, given that the J groups are assumed independent and identically distributed. For instance, in a two-level model all the variables of the vector $(\mathbf{Y}_j, U_j, X_{1j}, \ldots, X_{Kj})$ of the j-th group are represented in the graph. Whatever the cluster sizes n_j are, it is enough to depict the minimal sub-graph suitable to read the conditional independence structure from the graph. For instance, in the case of a two-level structure, only two elementary observations have to be included in the minimal sub-graph.

This solution implies additional definitions. An *individual node* is a node that represents a random variable for a specific statistical unit. A *grouping latent node,* is a node representing a latent random variable U_j being a separator between the individual nodes, such that, $Y_{ij} \perp\!\!\!\perp Y_{i'j} | U_j$. Such node is represented by the symbol ⦸. A *deterministic node* is a node representing a random variable whose conditional distribution is degenerate. This node is represented with a double line block.

The conditional independence structure of a two-level random intercept model can be represented by a chain graph where: the last block on the left, made up of pure response variables, contains two *individual nodes* and the second-last block contains a *grouping latent node*. LWF Markov properties can be used to encode the conditional independence structure of such a graph.

Figure 1 shows an example of a chain graph for a two-level random intercept model with only one explanatory variable. Therefore, the main advantage of this proposal is that the usual LWF Markov properties and the factorization criterion are valid in such chain graphs. For instance, in Figure 1, the pairwise chain graph Markov property suggests that the latent variable U_j is marginally independent from the explanatory variable Z_1. Moreover, due to the global Markov property, looking at the moral graph in (*b*), one can see that U_j is not independent of Z_1 conditionally on the response variables Y_j.

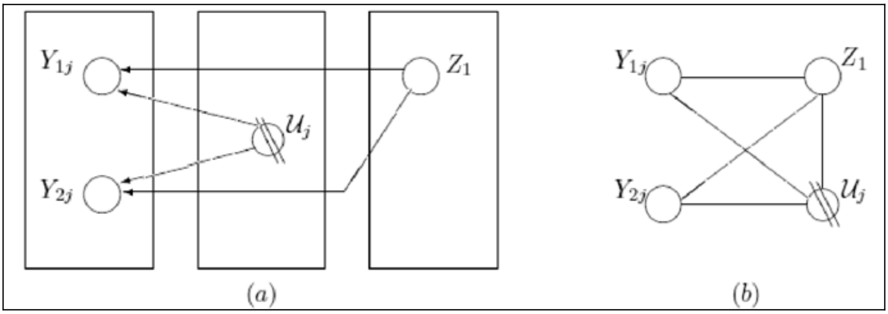


Figure 1. Example of (a) two-level random intercept graphical model, and (b) corresponding moral graph.

3. A multilevel graphical chain applied to graduates' employment

We adopted the multilevel graphical chain model described in Section 2.1 to analyse the data collected on the students who graduated in year 2000 at the University of Florence. They had been interviewed at approximately two years from the attainment of the degree. We analysed the occupational condition at the interview for the graduates who, at that time, were working or were seeking for an occupation. Our aim is to determine the factors that influence job finding, with reference to both individual characteristics and type of degree.

The data include 2,917 graduates employed or seeking a job: 46% had an occupation with tenure, the other 54% was unemployed or with a temporary occupation. The graduates under consideration had 56 types of degrees, with a number of graduates per course ranging from 4 (Chemistry) to 504 (Architecture), and a median of 22. Graduates (first level units) and degree programmes (second level units) thus characterize the hierarchical structure.

The response variable is binary with a value of 1, if the graduate has a stable occupation to the date of the interview, and 0 otherwise.

The use of a graphical model allows the study of the joint distribution of all the involved variables, bringing to light direct and indirect relationships. It is just this point that differs our contribution from others where the response variable is studied through a multilevel model, but the relationships among the explanatory variables are not modelled (e.g. Chiandotto & Bacci, 2004).

The specification of the graphical chain model requires, first, the covariates to be ordered according to the prior knowledge of the phenomenon, following a logical and/or time order. The variables used in our analysis[2] and their block ordering are shown in Table 1.

The variables in block 5 are the cluster means of the corresponding subject level variables. As shown in equation (6), the insertion of the cluster mean allows to decompose the total effect of a variable into a *between* and a *within* component. Since the conditional distribution of a cluster mean degenerates, that is $f(\overline{x}_j \mid x_{1j}, \ldots, x_{n_j j}) = 1$, a cluster mean is represented in the graph as a deterministic node, located in a block following the block containing the corresponding individual-level variable and preceding the block containing the cluster latent variable.

Block 6 contains only the random effect U_j, represented as a grouping latent node, whose role is to model the variance of the response.

---

[2] The variables were selected on a logical basis and thanks to past analyses. Our aim was to design a relatively simple model able to catch the key features of the process under scrutiny.

Table 1. Block ordering of the variables. Graduates of the year 2000, University of Florence.

Block	*Variable*	*Description*
1 exogenous	MALE	Gender: 1=male, 0=female
	MOTHER EDUC.	Mother's education: secondary school or low, high school (ref. cat.), degree
2 intermediate	LICEO	High school: 1=lyceum, 0=other
	SCH. MARK	High school final mark: 36-60 (mean=48.0)
3 intermediate	DIPLOMA	Type of degree programme: 1=diploma (3 years), 0=laurea (usually 4 years)
4 intermediate	AGE	Age at graduation: 21-50 (mean=27.6)
	UNIV. MARK	Average exam mark: 18-30 (mean=26.8)
5 cluster means	c.m. SCH. MARK	Degree programme mean of SCH. MARK
	c.m. AGE	Degree programme mean of AGE
	c.m. UNIV. MARK	Degree programme mean of UNIV. MARK
6 cluster latent node	$U_j \sim N(0,\tau^2)$	Degree programme latent variable
7 response	EMPLOYED	Employment: 1=stable occupation, 0=other

The block ordering shown in Table 1 and the independence assumptions of the multilevel model imply the following factorization of the joint distribution:

$$\begin{aligned} f(\mathbf{y}_j, u_j, \mathbf{x}) &= f(\mathbf{y}_j \mid u_j, \mathbf{x})\, f(u_j)\, f(\mathbf{x}) \\ f(\mathbf{x}) &= f(\mathbf{x}_{[4]} \mid \mathbf{x}_{[3]}, \mathbf{x}_{[2]}, \mathbf{x}_{[1]})\, f(\mathbf{x}_{[3]} \mid \mathbf{x}_{[2]}, \mathbf{x}_{[1]})\, f(\mathbf{x}_{[2]} \mid \mathbf{x}_{[1]}) \end{aligned} \tag{7}$$

where $\mathbf{X}_{[k]}$ denotes the variables of the k-th block, k=1,2,3,4, for example $\mathbf{X}_{[2]}$={LICEO, SCH. MARK}.

The fitting of the multilevel graphical chain model that corresponds to factorization (7) requires the fitting of four regression models, some of which have a multivariate response.

Given the alternation between categorical and continuous variables in consecutive blocks, we adopted the estimation procedure by Cox & Wermuth (1996). The procedure consists in fitting, for every endogenous variable, a univariate regression model whose explanatory variables are those in the same block and those in the previous ones. When the endogenous variable is continuous, a linear regression is fitted, while in the binary case a logistic model is used. The multilevel (random intercept) model is the one for the response variable EMPLOYED.

All the regression models are fitted by maximum likelihood. For the random intercept logistic model, the likelihood is approximated through adaptive Gaussian quadrature using the `gllamm` command of *Stata* (Rabe-Hesketh *et al.*, 2004).

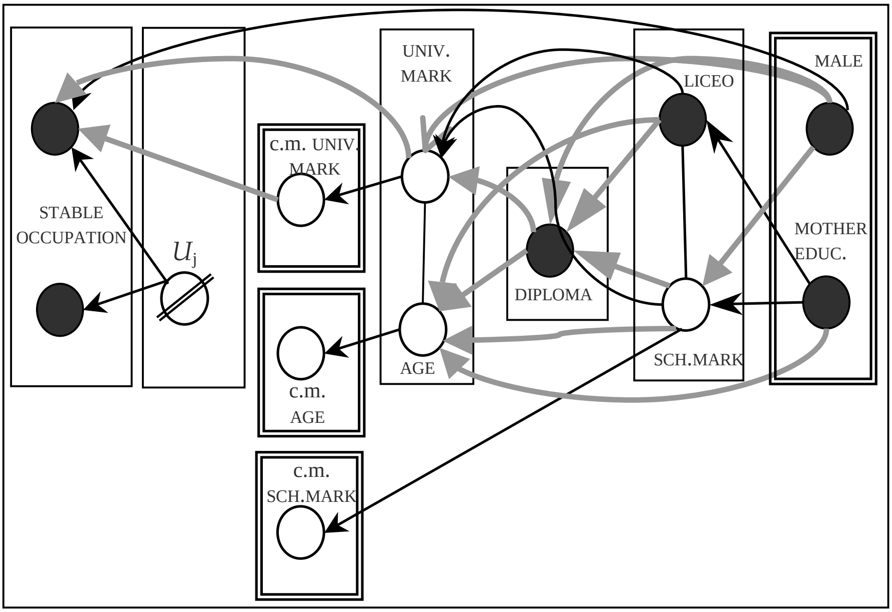


Figure 2. Multilevel graphical chain model: black (grey) arrows are positive (negative) effects at 10% significance level. Graduates of the year 2000, University of Florence.

The resulting graphical model is shown in Figure 2. The arrows are drawn when the *p*-value of the corresponding regression coefficient is less than 0.10.

In order to ease the reading of the graph the positive effects are represented by black arrows and the negative ones by grey arrows. The arrows pointing to the cluster means represent a deterministic relationship (as suggested by the double line block). The sign of the relationship between the response variable and the latent node U_j is not identifiable. Moreover, the two individual nodes in the final block are identically distributed, that is the dependence structure of the two nodes is the same even if, in order to simplify the reading of the graph, the arrows have been only traced for one of the two individual nodes.

The estimates concerning the models for the intermediate variables are not shown, as the essential information for the aims of the analysis is encapsulated in the graph of Figure 2. The estimates concerning the random intercept logistic model for the probability of a stable occupation are reported in Table 2.

Figure 2 shows that the response variable EMPLOYED directly depends only on MALE and UNIV. MARK. Therefore, the variables MALE and UNIV. MARK constitute a separator set between the response variable and the other covariates, in the sense that EMPLOYED is independent from AGE, DIPLOMA, LICEO, SCH. MARK and MOTHER EDUC. conditional on MALE and UNIV. MARK.

If only the model for the response variable was fitted, one would conclude that the factors relevant for employment were MALE and UNIV. MARK. Actu-

Table 2. Random intercept logistic model for the probability of stable occupation. Graduates of the year 2000, University of Florence.

Parameter	*Estimate*	*Standard error*	*p-value*
Intercept	4.925	6.024	0.414
MALE	0.372	0.087	0.000
MOTHER EDUC. (secondary sch.)	-0.020	0.090	0.820
MOTHER EDUC. (degree)	-0.115	0.135	0.397
LICEO	-0.089	0.087	0.308
SCH. MARK	-0.003	0.007	0.595
c.m. SCH. MARK	0.070	0.050	0.164
DIPLOMA	0.612	0.396	0.122
AGE	-0.005	0.016	0.737
c.m. AGE	-0.034	0.106	0.747
UNIV. MARK	-0.055	0.030	0.069
c.m. UNIV. MARK	-0.221	0.119	0.063
Cluster residual variance τ^2	0.510	-0.148	

ally, also other covariates influence the result, even if indirectly. This highlights the potentiality of the graphical chain model: given a block ordering of the variables, such a model allows to study the phenomenon taking into account the whole dependence structure.

The presence of an arrow between the cluster mean of UNIV. MARK and the response variable points out that UNIV. MARK has distinct *within* and *between* effects. From expression (6) it follows that the *within* effect is the coefficient of the variable at the subject level (-0.055), while the *between* effect is the sum of the coefficient of the variable at the subject level and the coefficient of the cluster mean (-0.221), and it turns out to be -0.276. Both the effects are negative, but the *between* effect is stronger, so the negative effect of the mark is largely due to the degree programme: a higher mark is associated with a lower probability of a stable occupation, because high marks are more frequent in degree programmes that usually yield modest occupational opportunities, e.g. the Humanities.

The effect at the subject level, even if significant and negative, is low. The effect at the subject level may be negative because the graduates with higher marks have greater ambitions and therefore are more demanding in job search.

If *between* and *within* effects are not disentangled, namely if the model contains UNIV. MARK without its cluster mean, the coefficient is -0.068 (s.e. 0.029). Such a coefficient is the total effect of the variable UNIV. MARK and therefore is difficult to interpret. However, a reader not accustomed to multilevel analysis is likely to misinterpret this effect as an effect at the subject level.

The unobserved factors at degree programme level are relevant: the likelihood ratio test comparing the models with and without random effects is significant (test statistic 108.8 with 1 degree of freedom) and the intraclass correlation coefficient ρ=0.134, that is 13.4% of the unexplained variance is at the degree programme level. Such value is rather high given the kind of model and the field of application.

The probability of stable occupation $P\left(Y_{ij}=1|u_j,\mathbf{x}_{ij}\right)$ is a function of the subject level covariates $\mathbf{X}_{ij}$ and the degree programme random effect U_j. To assess the role of the random effect, let us consider a particular graduate who is a male, has a mother with high education, attended a high school other than LICEO, had a high school final mark 48, obtained a 'laurea' degree, graduated at 27, and had an average examination mark of 26. For such a graduate the probability of a stable occupation is 0.56 if graduated in a mean course (u_j = 0), 0.72 if graduated in a course yielding high occupational chances ($u_j=+2\hat{\tau}$) and 0.38 if graduated in a course yielding low occupational chances ($u_j=-2\hat{\tau}$).

After parameter estimation, the residuals $\hat{u}_j$ can be calculated with the Empirical Bayes method (Snijders & Bosker, 1999). The degree programmes with a positive (negative) residual yield graduates with a probability of a stable occupation higher (lower) than predicted. The ranking based on the residuals has Nursing in the first position and Physics in the last position.

Figure 3 can be used to compare the residuals in pairs: two residuals are significantly different (at 95% average confidence level) if and only if the corresponding intervals do not overlap. The interval length is a decreasing function of the sample size of the degree programme: two extreme examples are

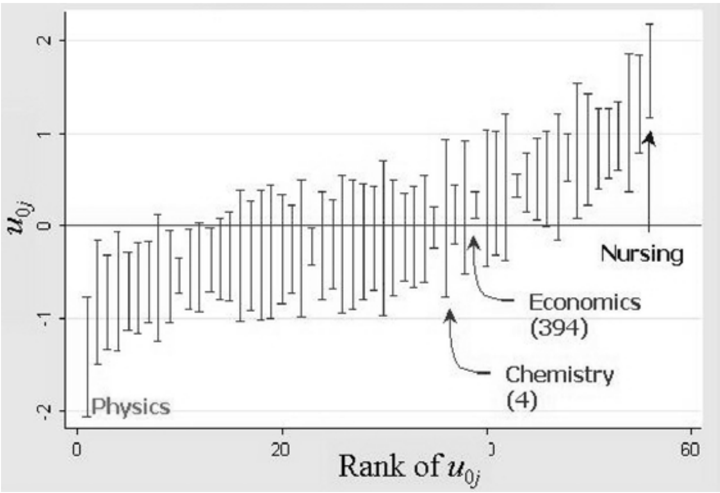


Figure 3. Intervals for pairwise comparisons between residuals at the degree programme level (95% average confidence level). Graduates of the year 2000, University of Florence.

Chemistry (4 graduates) and Economics (394 graduates). For a study programme with few graduates the considerable length of the interval hinders the comparison with the other programmes, as the differences are nearly not significant.

The residuals $\hat{u}_j$ incorporate all the unobserved factors at the study programme level, so they can be interpreted as a measure of external effectiveness of the study programme, though not adjusted for the conditions of the labour market. For example, the job opportunities yielded by Nursing depended not only on the quality of the programme, but also on the labour market needs (see, about this, Chiandotto & Grilli, 2003.

4. Final remarks

In this paper, we presented and applied a method of analysis based on the integration between chain graph and multilevel models. This method has the advantage to make explicit the assumptions on the ordering of the variables and on the conditional independences underlying the multilevel model. Moreover, the use of the graph helps to visualize the direct and indirect effects on the response variable and to read in a simple and direct way the conditional independences among the variables. It is not necessary that the variables follow a joint Gaussian distribution, nor a Conditional Gaussian one, so the estimates depend on the assumed block ordering of the variables. With a different block ordering the estimates could change and so the conditional independences. However, in our application, the adopted block ordering was plausible because it follows a logical and/or time ordering.

The potentiality of this class of models is still to be explored. It would be useful to extend the methodology in the following directions: several multilevel regressions in the same graph, modelling of the process of formation of the clusters, regression of cluster-level variables on individual-level variables.

The application based on data from the graduates employed or seeking a job, so the considered joint distribution is conditioned on this subset of graduates. For example, the relationship between the type of degree programme chosen by the student (DIPLOMA) and their characteristics (MALE, MOTHER EDUC., LICEO, SCH. MARK) is not referred to all the students enrolling in university, but only to students who eventually graduated and searched for jobs.

This choice follows the need to have a random sample from the joint distribution, as is customary in graphical models. The consideration of a wider sample, such as a cohort of freshmen, requires an adequate representation of the selection process in the graphical model: for example, for the freshmen who do not graduate, all the variables associated with graduation and successive job search are not observable, and they cannot be treated as missing at random. The development of multilevel graphical models able to represent also the process of selection of the statistical units is an interesting topic for future search.

References

CHIANDOTTO B., BACCI S. (2004) Un modello multilivello per l'analisi della condizione occupazionale dei laureati dell'Ateneo fiorentino. In: CROCETTA C. (ed) *Modelli di statistici per l'analisi della transizione Università–lavoro*, Cleup, Padova: 211-234.

CHIANDOTTO B., GRILLI L. (2003) *La domanda di lavoro nella provincia di Firenze: Analisi integrative sui dati dell'indagine Excelsior 2003*, University of Florence (http://www.unifi.it/aut_dida/indexval.html).

COX D., WERMUTH N. (1996) *Multivariate Dependencies. Models, Analysis and Interpretation*, Chapman & Hall, London.

FRYDENBERG M. (1990) The chain graph Markov property, *Scandinavian Journal of Statistics*, **17**: 333-353.

GOLDSTEIN H. (2003) *Multilevel Statistical Models (3rd edition)*, Arnold, London.

GOTTARD A., RAMPICHINI C. (2004) Chain Graphs for Multilevel Models, *Working Paper*, Department of Statistics 'G. Parenti', n. 8/2004, Florence.

LAURITZEN S. (1996) *Graphical Models*, Clarendon Press, Oxford.

LAURITZEN S., WERMUTH N. (1989) Graphical models for the association between variables, some of which are qualitative and some quantitative, *Annals of Statistics*, **17**: 31–57.

RABE-HESKETH S., SKRONDAL A., PICKLES A. (2004) *GLLAMM Manual*, U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.

SNIJDERS T., BOSKER R. (1999) *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling*, Sage, London.

Transition from University to the Job Market. A Time Analysis of the University of Cagliari Graduates

Mariano Porcu, Nicola Tedesco[1]

Department of Social and Economics Research, University of Cagliari, Italy

Summary. To investigate the transition from University to work, and indirectly assess the effectiveness of university education, time analysis can play a key role. The period that elapses between graduation and finding work is usually influenced by various factors. In this work, an analysis of such determinants is performed by means of non-parametric and parametric survival models.
Keywords: Transition to work; Time intervals; Graduates; Kaplan-Meier model; Weibull model.

1. Transition times between graduation and employment

The transition from university to the job market is a basic aspect to be considered when assessing the efficacy of the higher education system in Italy (Giambalvo, 1996; Ortu *et al.*, 2000). Any analysis of the outcomes of the university system has to evaluate the success of graduates in finding a job in a short time.

The occupation of graduates must be evaluated both in relation to the employment rate and job qualification, and to time needed to reach a given target. The length of a transition period may be influenced by several factors. By analysing these factors, we can understand how the missions of state-run universities to educate its citizens and to satisfy the demands of the labour society are accomplished.

---

[1] This paper is the result of the joint research of the two authors. M. Porcu was responsible for the final editing of Sections 1, 2 and 4, whereas N. Tedesco was responsible for Section 3. The authors wish to thank the anonymous referees for their precious suggestions.

Moreover, considering the “3+2” reform of the Italian university system, this aspect is an essential instrument to correct and improve the process of university education. Therefore, in this work, we will study transition times with semi-parametric and parametric techniques on a sample of graduates of University of Cagliari, in order to evaluate the importance of some covariates.

After the presentation of some general traits of the graduates of the University of Cagliari (Section 2), we will analyse their transition times (Section 3) and draw some general conclusions from the obtained results (Section 4).

2. The University of Cagliari graduates

Our data comes from a survey on graduates of the University of Cagliari, undertaken in autumn 2003. Our research, realized with a CATI system, involved 4,363 graduates (aged 37 years or less at graduation) who had their degree in years 1999 and 2000.

We selected a sample size of 1,112 (one out of four population units). For each subject, the professional status at the interview was noted: i.e. whether employed, unemployed (before employed), seeking for first employment, enrolled in a masters or PhD course, not interested in finding an occupation. Moreover, those in employment were asked whether they had started in their current job before or after completing their degree.

According to these variables, we selected a sub-sample of 906 graduates who were employed when interviewed and had begun their present job after the graduation. We concentrated our analysis on the employment status during a given period. Each of the 906 graduates was observed for a period ranging from 35 to 56 months (TIME) and their unemployment (transition) times were evaluated retrospectively. If a subject was not working at the end of survey period, his/her exposure time was truncated.

The criterion variable TIME is computed as difference between the student’s first employment time and that of graduation. If, during the observational period, the subject had obtained a master/PhD/specialization title, the time span (both truncated and complete) was recalculated subtracting the frequency period because we hypothesised that during post-graduate education subjects did not look for jobs.

For an explanatory analysis of transition times, we considered the covariates: sex (SEX), type of degree (TY_DEG), degree final mark (MARK), delay in finishing university studies (obtained by subtracting the statutory time for university studies from the time taken) (LAG), completion of master, or Ph.D., or other specialisation courses (POST).

Concerning the variable MARK (the interval is 66-110), we decided to create two levels (≤ 105 and ≥ 106) in line with its distribution (Fig. 1). The variable LAG was categorised in three levels: “≤ 1 year”, “2-3 years”, “≥ 4 years”.

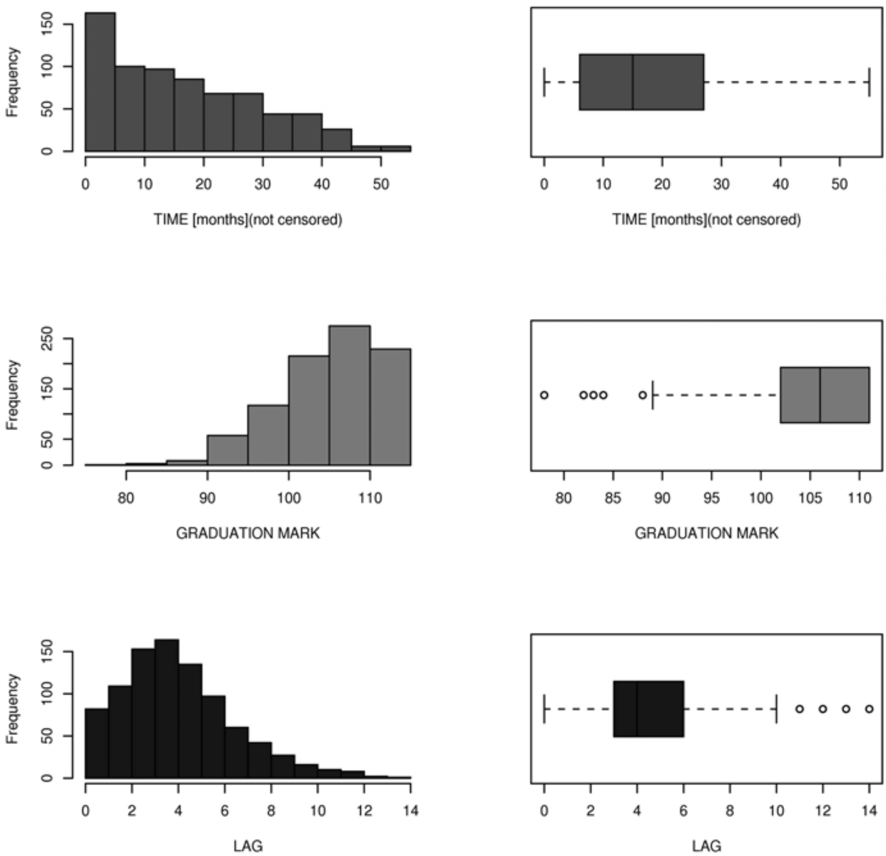


Figure 1. Graphics for variables TIME (only not censured), MARK and LAG

As for variable TY_DEG, we consider four groups: Economics-Law-Social Sciences (EGS), Scientific-Technical subjects (SCT), Life and Health Sciences (SVS) and Humanities-Education-Behavioural Sciences (UEC)[2] (the classification rationale is explained by Porcu & Puggioni, 2003).

Table 1, and Figures 1 and 2 show the distribution of the examined variables. It is important to point out the negative asymmetry for variable MARK and high values for the average (105.0) and for the median (106.0). Contrariwise, the variable TIME shows a positive asymmetry (the average being equal to 17.4 and the median being 15.0 months) according to previous studies (Ortu *et al.*, 2000).

---

2 EGS group: degrees in Economics, Law and Political Science; SCT group: Engineering, Physics, Mathematics, Chemistry and Geology; SVS group: Medicine, Biology, Natural Sciences and Pharmacology; UEC group: Literature, Languages and Educational Sciences (Pedagogy and Psychology).

Table 1. Synthetic statistics for variables TIME, MARK and LAG on sample of survey.

Covariates	*Min*	Q_1	M_e	μ	Q_3	*Max*	σ
*TIME**	0.00	6.00	15.00	17.36	27.00	55.00	12.94
MARK	78.00	102.00	106.00	105.02	110	110	5.73
LAG	0.00	3.00	4.00	4.47	6.00	14.00	2.47

* Relative to non truncated observations

One noteworthy finding, considering its negative implication for an evaluation of the performance of our university system, is the average value for the variable LAG (4.47 years) and the median (4.00 years). We can conclude that in Sardinia the average time a student spends at university amounts to twice the statutory duration with a consequent and evident increase in costs for a university system that is clearly inefficient.

In Figure 2, we can observe distributions of some variables on a sub-sample of 906 subjects. The employment rate is 78% (707 events of interest and 199 censured), in which there is a prevalence of female (62%) graduates not involved in post-graduate education (53%), and the principal degree categories are EGS (36%) and UEC (30%).

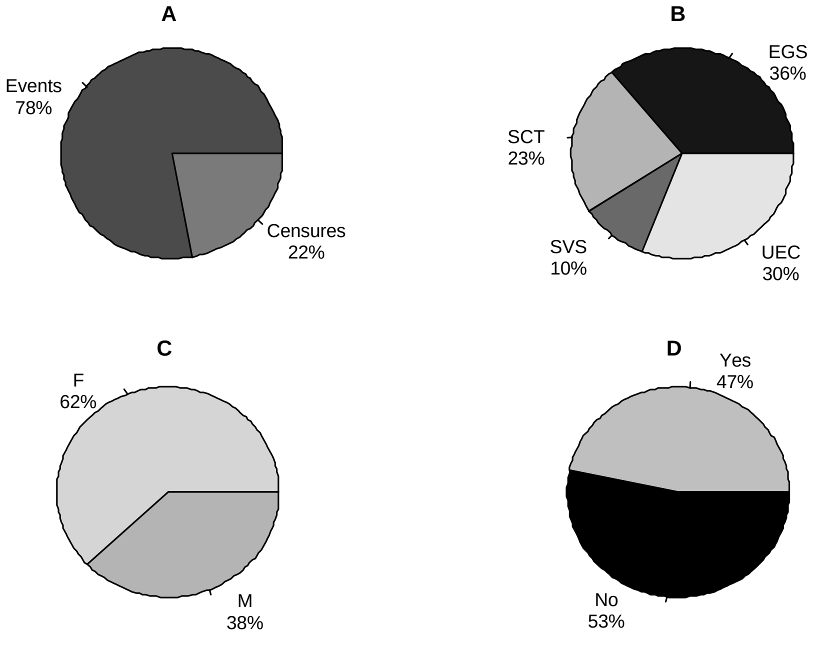


Figure 2. Variables indicating the event (**A**) and TY_DEG (**B**), SEX (**C**) e POST (**D**)

3. Analysis of transition times

In order to study the transition times and the importance of some covariates on times, we initially applied the Kaplan-Meier (KM) method, a non parametric technique (Section 3.1) and then a parametric one (Section 3.2). The total duration of the follow-up was 56 months (March 1999 – October 2003), and the longest time-event was 55 months.

The survival function on all observations (Fig. 3) displayed a uniform trend with a median of 21 months and with 25% of the graduates who found a job within 8 months from graduation. We found that truncated observations (the unemployed) appear after 35 months (the minimum period of observation referring to graduates that are more recent).

The regularity of the survival function is an indirect indicator of the labour market's inability to absorb graduates. When there is a good level of demand for graduate employees, we would expect the survival function to have an "L" form. In other words, the majority of graduates would find a job in a short time and only a minority would require longer finding employment. In our opinion, this aspect is all the more relevant if we consider that the variable TIME does not incorporate the time for post-graduate education, as if someone who is studying for a post-graduate qualification would not in the job market[3].

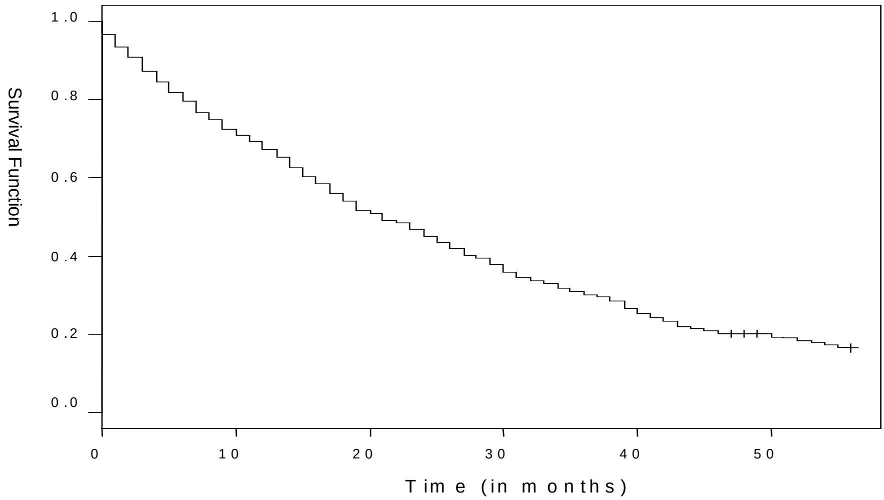


Figure 3. Kaplan-Meier survival function on 906 observations

---

[3] The hypothesis is conservative. In a context of relative backwardness of an economic system such as Sardinia's, it is plausible that a large percentage of graduates view post graduate education as a sort of status in which they can study and work simultaneously.

3.1 Non parametric analysis: the effect of covariates

We carried out a non-parametric analysis of transition times in relation to some covariates SEX, MARK and POST (Fig. 4) considering the stratification of observations according to the levels of covariates. No covariates proved significant because the survival curves for each level of covariates were similar, in particular for the POST variable. This is an interesting result for three reasons: a) the local labour market's inability to employ graduates with an post graduate qualification; b) the low quality of post-graduate education in Sardinia[4]; c) the possibility that post-graduate education is a way of overcoming the gaping divide between demand for and supply of graduates in Sardinia.

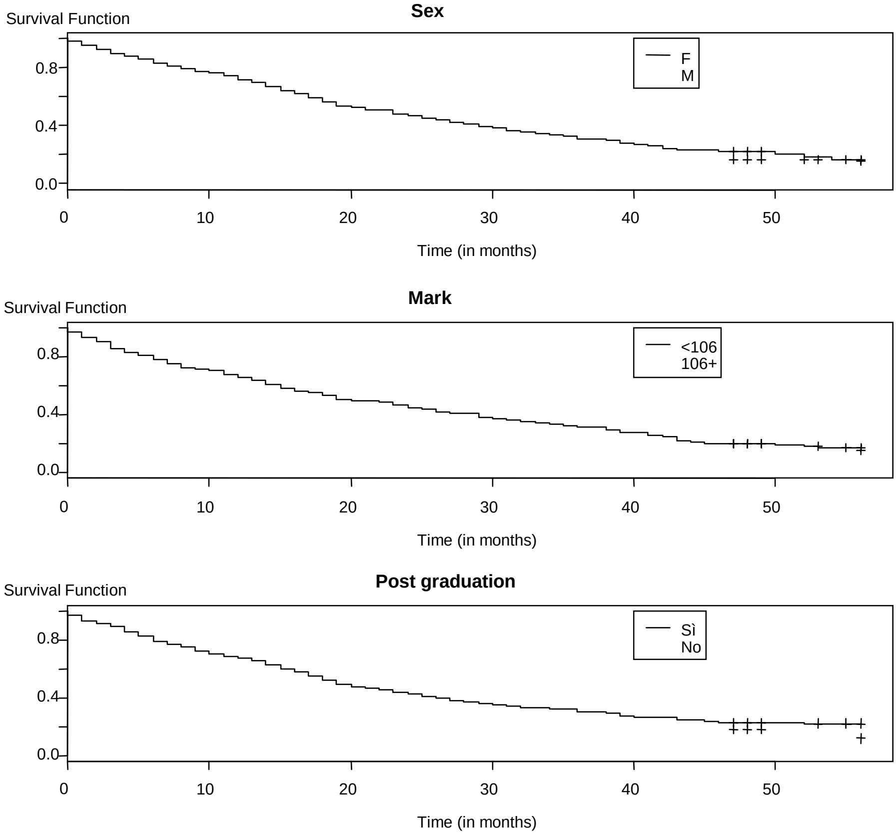


Figure 4. KM survival function for variables SEX, MARK and POST

[4] In our sample, in most cases post-graduate education was not a PhD or a Master programme but a shorter duration course.

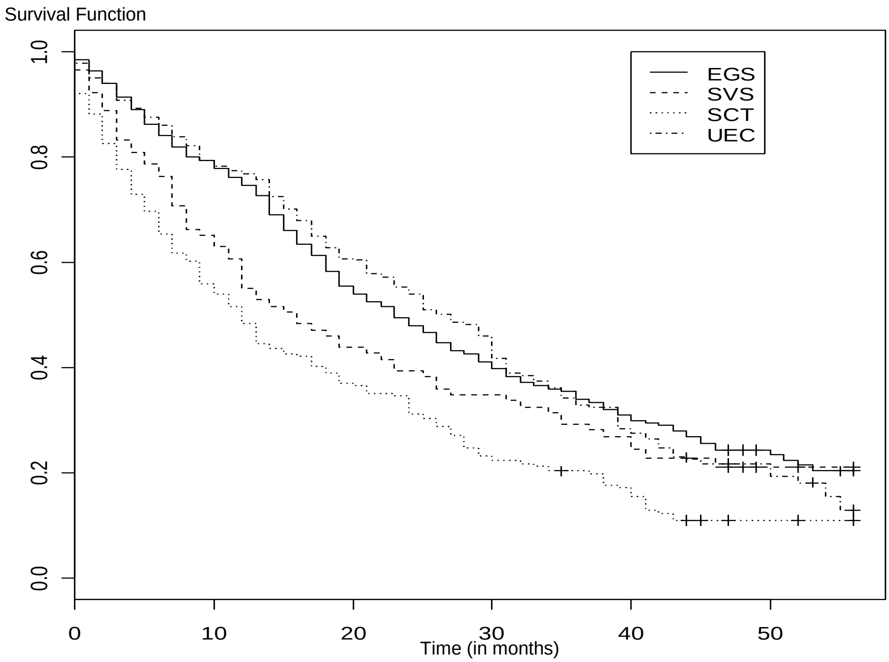


Figure 5. KM Survival Functions for variable TY_DEG

The stratified analysis of transition times by type of degree (TY_DEG) is rather interesting. As we can observe (Figure 5), survival functions for graduates SCT and SVS show how their transition times are lower than they are for other graduates (in particular for SCT) and that the shape of their curves is more similar to “L” than in the other groups, which is the typical situation for short time events.

The comparison between medians and quartiles (Table 2) provides us with further points for reflection. On average, an SCT graduate finds a job in half the time it takes UEC or EGS graduates to find one. It is interesting to notice how the longest observation time of the event for SCT graduates (43 months) is lower by almost one year than for graduates of groups EGS (53 months) and UEC (55 months).

Such evidence is a well-known factor among those who analyse university education performance; our evidence stems from a context (Sardinia) of relatively low economic development. Therefore, this would indicate that even in a region where the services, and in particular the public ones, are particularly demanding for the new employment, graduates of groups EGS and UEC do not have a greater probability of finding a job[5].

[5] Nevertheless, it is reasonable to believe that this evidence is influenced by the persistence of a blocking of employment opportunities in the public sector.

Log-rank, Wilcoxon and Likelihood Ratio tests were applied to test the hypothesis of homogeneity of survival curves for each category of covariate TY_DEG. They are significant in all cases (p<0.0001).

The analysis of transition times for the LAG variable (Figure 6) does not show any differences in distributions of times, even though those who complete a degree course with a lag of one year have shorter transition times (Table 2). Moreover, for the same category of graduates, Q_1 and M_e are quite similar to the corresponding values for other categories, while values of Q_3 and *Max* appear lower.

Table 2. Some statistics on transition times for covariates TY_DEG and LAG

Variables	*Min*	Q_1	M_e	Q_3	*Max*
TY_DEG:					
EGS	0	12	23	46	53
SCT	0	4	12	28	43
SVS	0	7	16	40	46
UEC	0	14	26,5	42	55
LAG:					
≤ 1	0	8	18	34	45
2-3	0	8	19	38	55
>3	0	9	23	43	54

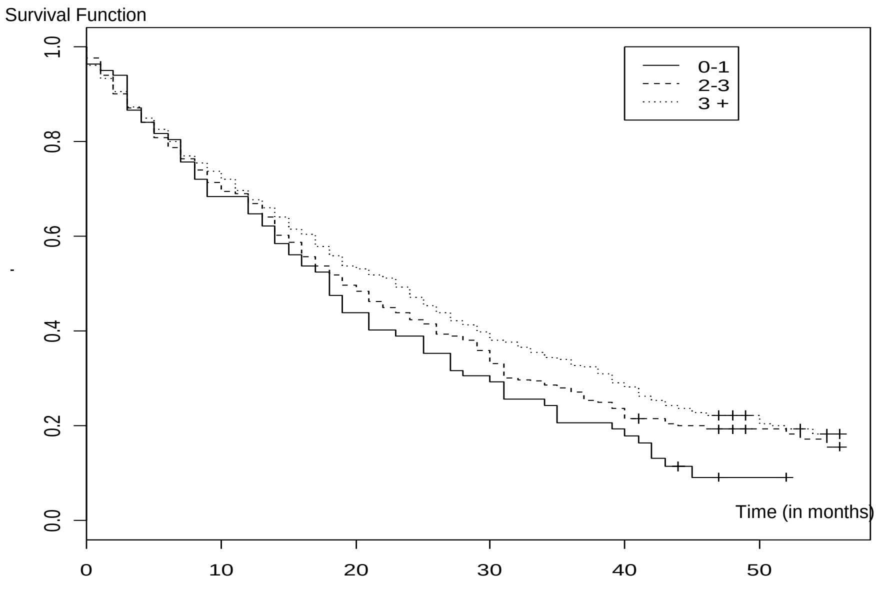


Figure 6. KM Survival Functions for variable LAG

This meant that only Log-Rank and LR tests were significant (at 5% level), but this was the case for the Wilcoxon Test (not significant because of the intersection of curves in the initial part of the follow-up).

3.2 Parametric analysis of covariate effects

Non-parametric analysis enabled us to distinguish between different durations of times in a status of unemployment. Nevertheless, it could not actually measure the strength of the effects of covariates. That is why we applied a parametric Weibull survival model[6].

The hazard function of the model is defined as

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}) h_0(t),$$

where $h_0(t) = \lambda\gamma t^{\gamma-1}$ distributes as the Weibull function, with λ being a scale parameter and γ a shape parameter.

The results of the application of the model (Table 3), show clearly how the principal covariates that influence the likelihood of finding employment are LAG (p=0.0136) and TY_DEG (p<0.0001).

Table 3. Significance of factors and parameters applying a Weibull model

Covariates	*Estimates*	*Standard error*	*Chi square*	*p-value*
INTERCEPT	*3.6241*	*0.1327*	*746.23*	*<0.0001*
SEX				*0,6940*
F *vs* M	0.0389	0.0989	0.15	0.6940
MARK				*0.9686*
≥ 106 *vs* < 106	0.0037	0.0929	0.01	0.9686
CORSO				*0.7660*
Yes *vs* No	0.0262	0.0880	0.09	0.7660
LAG				*0.0136*
≤ 1 *vs* > 3 years	-0.4269	0.1541	7.68	0.0056
2-3 *vs* > 3 years	-0.1682	0.0999	2.84	0.0922
TY_DEG				*<0.0001*
EGS *vs* UEC	-0.0652	0.1163	0.31	0.5747
SCT *vs* UEC	-0.6104	0.1305	21.87	<0.0001
SVS *vs* UEC.	-0.2329	0.1587	2.15	0.1422
λ scale parameter	1.1359	0.0375		
γ Weibull shape	0.8803	0.0290		

[6] We applied the parametric Weibull model to test the existence of a linear relation between *log*(*times*) and *log*[-*log* $S(t)$], where $S(t)$ represents the survival function calculated with the Kaplan-Meier method . We did not apply the semi-parametric Cox model because we did not verify for all covariates the hypothesis of the proportional odds of events for each different category of covariates.

These results were by no means unexpected because they were disclosed in the non parametric KM analysis and in other studies (Tedesco, 2002; Porcu & Puggioni, 2003a), albeit with different research objectives.

The analysis of estimated parameters clarifies the role of the above-mentioned factors.

First of all, with regard to the time taken to obtain a degree, it is evident that a lag of 2 to 3 years, with respect to a baseline lag of 3 or more, does not cause any significant reduction of transition times (p>0.09).

On the contrary, what appears to be evident and statistically significant is the reduction of transition times for those who finish their university career either within the statutory time or with a minimum lag (one additional year) compared to the baseline ($\hat{\beta} = -0.4269$; p=0.0056).

As regards type of degree, the only significant parameter is for the category SCT ($\hat{\beta} = -0.6104$; p<0.0001); the estimate for this parameter is particularly high because *exp*(-0.6104) = 0.5431: in other words, the probability of a SCT graduate getting a job is almost twice as high as it is for a UEC graduate (baseline) .

Results for the SVS group were quite interesting: the parameter is not significantly different from that obtained with the non-parametric KM analysis. The high value for dispersion of the estimate (0.1587) is probably due to the substantial number of those not employed (more than 25%) in this group of graduates and this causes the parameter to be non significant.

4. Conclusions

Our research represents the first analysis of transition times undertaken on a sample of graduates at the University of Cagliari with the aim of defining a general context in which to carry out an in-depth study of the problem of transition from university to job market.

We observed a linear survival function which shows that the labour market can absorb the supply of graduates, which shows the existence of endogenous and exogenous obstacles to employment. One such exogenous obstacle is the lack of new openings for jobs in the public sector, which, in a region such as Sardinia, represents the principal employer for graduates in several areas, and in particular in the education sector.

In contrast with the results of other research on the efficiency of employment markets, no significant differences were recorded between genders (but lack of new job positions in the public sector can mask the effect of gender). Similar results have been obtained with the variables degree classification (MARK) and postgraduate education. This may indicate that these are not discriminating factors in obtaining a job. This is in contrast with general opinion, in particular in the opinion of students.

The flattening toward high level frustrates any possibility of selection based on the criteria of final degree mark, while for post-graduate education the result is surprising because it seems that the choice to attend a postgraduate programme is dictated more by the lack of available employment than the idea of getting a higher qualification for the market.

Among the significant factors, we quoted the effect of the type of degree and, to a lesser extent, that of the duration of university studies.

As far as type of degree is concerned, only the SCT (Scientific-Technical) group shows significantly shorter transition time, as well as a lower percentage of censures (unemployed) at the end of the observation period.

A similar trend was observed for graduates of group SVS (Life and Health Sciences), but the large number on censures (explained by the high percentage of graduates that attend specialisation programmes for medical doctors) and the low values for transition, or those finding work in a short time made the parameter in the Weibull model non significant.

As for the duration of studies, we have the confirmation that those who get their degree in a short time have a higher probability of obtaining a job in a short time in comparison with those who complete their degree with a lag of four years or more (this was the case for more than 50% of our sample).

In conclusion, the labour market in Sardinia, displays a striking inability to select graduates based on their *curricula*. Probably, the labour market is saturated in traditional economic sectors (public administration) and, because there is no innovative post-industrial development policy, we might see a rise in the phenomenon of graduate emigration, which recent research (Porcu & Puggioni, 2003b) puts at about 10%.

References

COLLETT D. (1994) *Modelling Survival Data in Medical Research*, Chapman & Hall, London.

GIAMBALVO O. (1996) Tempi di attesa per l'inserimento nel mercato del lavoro di una leva di laureati. In: *Rivista Italiana di Economia Demografia e Statistica*, **L(1)**: 243-267.

KLEIN J.P., MOESCHBERGER M.L. (2003) *Survival Analysis. Techniques for Censored and Truncated Data, 2nd ed.*, Springer, New York.

PORCU M., PUGGIONI G. (2003a) Laurea e abbandono universitario. Uno studio comparativo dei due eventi su un campione di immatricolati dell'Università di Cagliari. In: FABBRIS L. (ed) *LAID-OUT: scoprire i rischi con l'analisi di segmentazione*, vol.3, CLEUP, Padova: 25-40.

PORCU M., PUGGIONI G. (2003b) Transizione Università–Lavoro. Analisi preliminare sui laureati della Sardegna nel periodo 1992 – 2001. In: CIVARDI M. (ed) *Transizione Università–Lavoro: la definizione delle competenze*, CLEUP, Padova: 81-96.

TEDESCO N. (2002) Analisi di segmentazione di una coorte di immatricolati dell'Università di Cagliari. In: G PUGGIONI (ed) *Modelli e metodi per l'analisi dei rischi sociali e sanitari*, CLEUP, Padova: 141-160.

TEDESCO N., PUGGIONI G., ORTU. A. (2000) Inserimento nel mercato del lavoro dei laureati in Scienze Politiche dell'Università di Cagliari dal 1987 al 1996. In: CIVARDI M., FABBRIS L. (ed) *Valutazione della didattica con sistemi computer-assisted*, CLEUP, Padova: 135-154.

The Determinants of Graduates' Placement. Analysis of Interactions Using Boolean Logit Models

Mariano Porcu, Giuseppe Puggioni, Nicola Tedesco[1]

Department of Economic and Social Sciences, University of Cagliari, Italy

Summary. In this analysis of the occupational placement of graduates, we define the role played by some covariates assembled to predict the dichotomous event occupied/unoccupied. These covariates influence the response variable singularly and jointly. This work aims to evaluate this joint effect by means of a recently developed technique known as Boolean logit. We applied an exploratory binary segmentation analysis to support the analysis.
Keywords: Graduates' placement; Segmentation analysis; Boolean regression analysis; Logit regression analysis.

1. Introduction

In the evaluation of the performances of the university educational system, the search for the determinants of the occupational placement of graduates is an important issue at stake. The issue has been approached with different methodologies (Chiandotto, 2004; Civardi & Zavarrone, 2004). An approach that appears to have an important role is the logit model, based on causal dependence between a response variable and a set of predictors. The dichotomous variable *employed*/*unemployed* is considered as dependent on a set of p predictors

$$y = f(x_1,\ldots,x_p).$$

[1] This paper is the result of the joint research of the three authors. M. Porcu was responsible for the final editing of Sections 1, 2, 5 and 6, whereas N. Tedesco was responsible for Section 4, and G. Puggioni for Section 3. The authors wish to thank the anonymous referees for their precious suggestions.

Predictors influence the response variable singularly, and in combination with each other. Such responses outline a framework of analysis based on the conceptual category of *causal complexity*. According to Braumoeller (2003), causal complexity is a concept in which "*multiple causes interact with one other, and the way in which they interact is described by the logical operators* and *and* or".

A number of concepts can be considered as special cases of causal complexity, that is:

- multiple conjoint causation: X_1 and X_2 and X_3 produce Y;
- substitutability: X_1 or X_2 or X_3 produce Y;
- contexts: X_2 produces Y but only in the presence of X_1;
- necessary and sufficient conditions: X_1 and X_2 produce Y, either X_1 or X_2 produce Y;
- INUS conditions[2]: (X_1 and X_2) or (X_3 and X_4) produce Y.

Complex causation is a problem for the majority of standard statistical techniques. The problem is that causal complexity implies non-additivity, which arises from the cumulative process of the influence of the independent variables on the response variable. This means that the presence or absence of one independent variable mitigates (or even nullifies) the impact of another. So, from a practical point of view, the problem arises of how to "capture" causal complexity with standard statistical techniques.

A number of methodological proposals have been put forward and much attention has been paid to studying them (Frosini, 2004). With reference to the dichotomous event *employed*/*unemployed*, we observe that in several research studies it has been stated that the event could be considered as the outcome of a process of causal complexity.

2. Modelling the interactions

The logistic regression model is often used to model the probability of a certain event as a function of a set of explanatory variables. The influence of the explanatory variables on the response is considered linear on a logit scale

$$\log(\pi/(1-\pi)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ .$$

The possible joint effects among covariates are usually taken into account by fitting the product among the variables into the model itself (Hosmer & Lemeshow, 1989):

$$\log(\pi/(1-\pi)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \{X_1 \times X_2\} \ .$$

---

2 The acronym INUS, created by Mackie, defines a particular kind of causal relationship. It makes reference to "an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result" (Braumoeller, 2003).

This modelling forces the researcher to keep the interactions among the variables on a very simple level, generally one-way or two-ways, for both technical (e.g. sparsity of data, power of tests) and theoretical reasons (e.g. parsimony principle). Consequently, only the main effects are usually included in the model, although the joint effects are more useful in predicting or in selecting groups, especially for social analysis purposes.

2.1 The Boolean logit

A method that can take into account the relationships among variables that are rooted in the concepts of causal complexity is the so-called *Boolean logit* (Braumoeller, 2003). It allows the researcher to estimate the influence of the interactions among predictors on the binary response Y. It is postulated that Y is thought to be produced by a *Boolean* or *logic* combination of some conditions $A_1, \ldots, A_k, \ldots$, e.g.:

$$A_1 \text{ and } (A_2 \text{ or } A_3) \rightarrow Pr(Y=1) = \pi = Pr(A_1) \times Pr(A_2 \cup A_3).$$

where $Pr(.)$ denotes the probability of the argument. The probability of occurrence for each condition

$$Pr(A_K) = p_k$$

is expressed by means of a logit (or probit) model (Braumoeller, 2003):

$$\boldsymbol{p_k} = \frac{\exp(\boldsymbol{\beta_k X})}{1+\exp(\boldsymbol{\beta_k X})}.$$

The k index means that each "condition" depends on its own explanatory variables $X = \{X_j\}$ through corresponding parameters β_k. The same X_j can be included in different p_k with no multicollinearity arising. Obviously, if there is only one "condition", Boolean logit reverts to the standard logit.

Boolean logit can be helpful in solving the problem of statistical estimation when causal complexity is present. To fit it, we need to postulate a model for π and to express π as a function of explanatory variables and relevant parameters through the p_k. As an example, if we assume that

$$\pi = Pr(A_1) \times Pr(A_2),$$

$$\text{logit}(p_1) = \mathbf{x}_1'\boldsymbol{\beta}_1; \quad \text{logit}(p_2) = \mathbf{x}_2'\boldsymbol{\beta}_2,$$

the model takes the form

$$\pi_i = p_{1i} \times p_{2i}$$

and the likelihood is

$$\boldsymbol{Lik}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \prod_{\boldsymbol{i}=1}^{\boldsymbol{n}} (\boldsymbol{p}_{1\boldsymbol{i}} \times \boldsymbol{p}_{2\boldsymbol{i}})^{y_i} (1 - \boldsymbol{p}_{1\boldsymbol{i}} \times \boldsymbol{p}_{2\boldsymbol{i}})^{1-y_i}.$$

So, once the occurrence of the event is explained in the language of *complex causation*, the consequent hypothesis can be expressed in probabilistic terms[3]. In conclusion, it might be worthwhile noting that the brackets are quite important since the Boolean statement "(*A and B*) *or C*" is different from "*A and* (*B or C*)".

3. The data

Data were collected with a CATI survey carried on in November 2003. A sample of 1,112 graduate students of the University of Cagliari was selected among those who graduated in the years 1999 and 2000. At the end of the survey, interviewees were classified according to their occupational status: 823 were employed, 150 unemployed, 137 were enrolled in postgraduate educational programmes (2 were missing). The whole group of employed was then split into two subsets, one of which included graduates (756) that found employment after graduation, the other graduates that were in work before they finished university studies (67).

Before performing the analysis, we decided to fix some eligibility criteria for the observation, i.e.

occupational status:

- employed;
- unemployed;

for the employed:

- job found after graduation;
- no more than 36 months to get the job.

Because of such criteria, and in order to model the dichotomous event "*Y*" *employed/unemployed*, we considered a total amount of 837 observations. Among them, 687 were employed (*Y*=1), 150 unemployed (*Y*=0).

During the survey, a lot of information on demographic as well social information on the graduates was collected. Evaluations on their educational programmes and previous working experiences were also collected together with data on time spent looking for a job (Porcu & Tedesco, 2004; Porcu & Puggioni, 2004).

Some exploratory analysis was carried out. The results (not shown here) led to the selection of a set of variables that could be particularly relevant in undertaking the present analysis, namely *sex, high school attended, mark gained in the high school leaving examinations, degree*[4], *age at graduation, final*

---

[3] A possible alternative to Boolean logit to model causal complexity is the so called *Logic Regression* recently proposed by Ruczinski *et al*. (2003).

[4] The study programmes were classified into four groups. a) Economic-Legal-Social (ELS); b) Scientific-Technical (SCT); c) Health and Life Sciences (SHL); d) Humanities and Behavioural Sciences (HBS).

mark at graduation, post-graduation studies. We will dichotomise all these variables.

4. Groups of variables for the analysis of interactions

One of the major drawbacks of the Boolean logit is the arbitrariness in the choice of the combination of predictors. A good set of predictors is the researcher's beliefs and opinions; such ideas are obviously rooted in her/his own experience on the topic that is being investigated. In any case, the uncertainty and the subjectivity of such a method for choosing the combinations of predictors may weaken the final model.

Moreover, the Boolean logit models require numerous computational resources and a model choice based on comparisons among the results gained with different combinations of predictors that could be excessively time consuming.

In the following, we propose a method to choose the combination of predictors. We will apply an exploratory segmentation analysis to give the researcher useful hints on the influence exercised by a set of predictors on a response variable and on the relationship between them. With such a method, results could vary with respect to the choice of the kind of segmentation (especially considering the chosen "*criterion function*").

Nevertheless, the choice for a binary segmentation analysis based on the likelihood criterion function could be considered suitable in that it allows the researcher not to choose a distance function (Tedesco, 2002). Segmentation could be considered as a representation of the causal complexity data as well. Therefore, our aim is not to lose this kind of information while building up Boolean groups. Essentially, we try not to make the abstract ideas of the researcher prevail but let the data 'speak for itself'.

The software used for segmentation was SAS-RECPAM (Carinci & Pellegrini, 2001; Ciampi, 1991). The chosen criterion function is to maximise the likelihood ratio of the logit *employed/unemployed* with respect to all the possible two by two combinations of the predictors; a minimum number of 40 observations per node/leaf was also incorporated into the model (minimum 10 of them employed). The split was set to α=0,05. Such a restrictive criterion was chosen to form a concise tree with not too many branches. The main objective was, indeed, to explore data for the next Boolean logit analysis.

The possible predictors we set in the segmentation were: the *attending of a post graduate programme* (Yes/No); *sex* (Male/Female), *type of high school attended* (Lyceum[5]/Other); *kind of degree* (ELS, SCT, SHL, HBS); *mark*

---

[5] The Italian "Lyceum" provides a classical education such as the one offered by the old British "Grammar Schools".

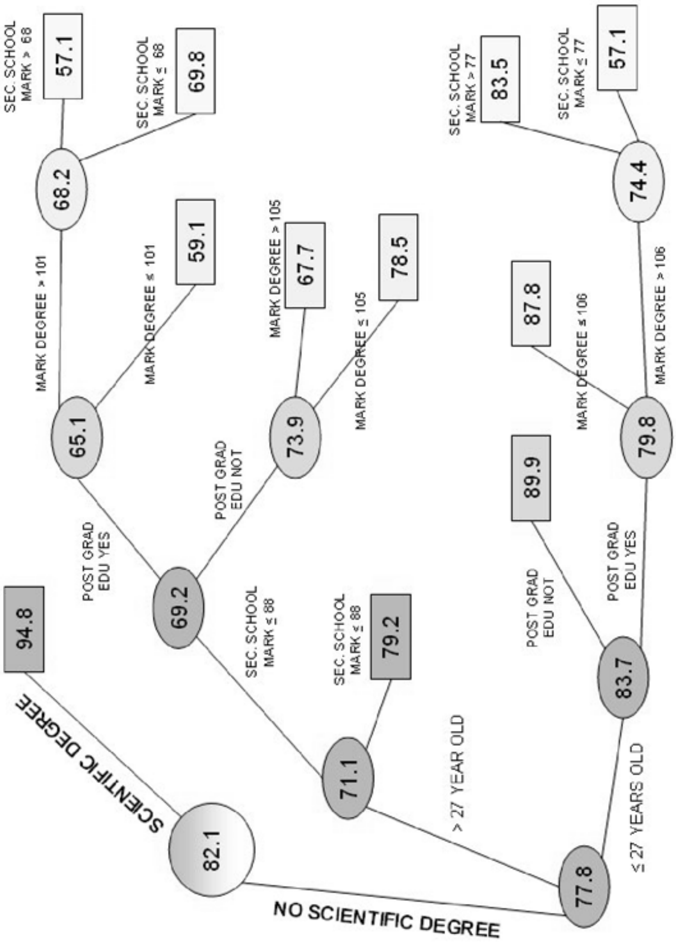


Figure 1. Segmentation tree

gained in the school leaving examinations ; *age at graduation*, *final mark at graduation*. We considered these variables on a continuous scale so to obtain guidance for the setting of thresholds useful to collapse them into categories.

The tree diagram (Figure 1) shows some interesting results. First, we perceive a marked asymmetry because graduates in the SCT group have a high employment rate (94.8%) and do not split further. This means that for these graduates the kind of degree they gained is the only important factor of success in getting employment.

On the other hand, for all the other graduates, the total employment rate is lower (77.8%) but it reaches notably higher values, in particular, for those younger (<27 years old) and who did not attend postgraduate programmes of study (89.9%). Among older graduates (≥27 years old), what seemed to be the cause of notable disadvantage was a low mark at high school final examination, the attendance of post graduates studies and a low mark at graduation.

It is clear that among graduates that are not in the SCT group, the low age at graduation is more useful than postgraduate studies for getting a job. Maybe, this is because post-graduate education is sometime a sort of standby status for those seeking occupation, and training is provided directly by companies.

The mark gained at high school leaving examinations also seems interesting: it occurs several times in segmentation, especially if this evidence is compared with the non-significant role played by variables such as sex, or type of high school. It may suggest that the "quality" of a graduate lies not only in the university education but also in the high school curriculum, as it plays an important role in determining the occupational placement of a graduate.

Focusing again our attention on the non-SCT group and on younger and with a postgraduate education, we highlight that the graduation mark has little influence (87.8% is the percentage of employed with a graduation mark ≤106/110). Among those with a high graduation mark (>106), it is the mark gained at high school leaving examinations that seems to play a role.

In conclusion, for the non-SCT group (considering the values of the GPI[6], see Table 1) marks for high school leaving examinations and the age at graduation play a joint effect on the response variable.

Table 1. GPI values

Covariates	GPI
High school leaving examination mark	100
Degree	96
Age at degree	90
Degree mark	60
Postgraduate studies	55
Type of high school attended	28
Sex	27

5. Modelling the event employment

To model the event *employed* (Y=1) *vs unemployed* (Y=0) we considered the results of the segmentation analysis and we treated them as dichotomous variables (1=Yes; 0=Not):

[6] The GPI (Global Predictive Index) is a measure of the predictive power of a covariate based on increments of the LRS for each covariate in each node with respect to the value of LRS obtained without that predictor. It is a measure of the informative power of the *i*-th predictor. Once the *i* sums are calculated, the bigger value is set at 100 and the remaining are rescaled in relation to this (Ciampi, 1991).

- male sex (SEXM);
- "lyceum" attended as high school (LICCS);
- mark at high school final examination ≥90/100 (DIP90);
- degree Technical-Scientific (SCIEN);
- age at degree ≤ 26 years (LAU26);
- final mark at graduation ≥108 (VOTOHIGH);
- postgraduate studies = Yes (CORPOST).

5.1 Fitting of a standard logit

We will first present the results of the fit of a standard logit model and then the fit of the Boolean logit. The results are summarised in Table 2. We can see that the only significant ($\alpha = 0.05$) variables[7] are DIP90, SCIEN, LAU26 and CORPOST.

Table 2. Point estimates ($\hat{\boldsymbol{\beta}}$) and z-scores (z= $\hat{\boldsymbol{\beta}} / \boldsymbol{SE}(\hat{\boldsymbol{\beta}})$) for the standard logit model (*Log*Lik=-362,937)

Covariates	$\hat{\boldsymbol{\beta}}$	|z|-*score*
SEXM	0.1967	0.916
LICCS	-0.2801	1.417
DIP90	0.5453	**2.068**
SCIEN	1.4855	**4.315**
LAU26	0.6134	**2.875**
VOTOHIGH	-0.2575	1.301
CORPOST	-0.4493	**2.309**

Table 3. Point estimates ($\hat{\boldsymbol{\beta}}$) and z-scores (z= $\hat{\boldsymbol{\beta}} / \boldsymbol{SE}(\hat{\boldsymbol{\beta}})$) of some standard logit models

Covariates	*Basic model*		*Without* SCIEN		*With* SEXM×SCIEN	
	$\hat{\boldsymbol{\beta}}$	|z|-*score*	$\hat{\boldsymbol{\beta}}$	|z|-*score*	$\hat{\boldsymbol{\beta}}$	|z|-*score*
SEXM	0.1967	0.916	0.5464	0.916	0.1780	0.787
LICCS	-0.2801	1.417	-0.2030	1.417	-0.2770	1.398
DIP90	0.5453	2.068	0.8068	2.068	0.5454	2.069
SCIEN	1.4855	4.315	–	–	1.3758	2.544
LAU26	0.6134	2.875	0.5251	2.470	0.6128	2.873
VOTOHIGH	-0.2575	1.301	-0.3212	1.633	-0.2579	1.303
CORPOST	-0.4493	2.309	-0.5115	2.668	-0.4445	2.274
SEXM × SCIEN	–	–	–	–	0.1778	0.256
log*Lik*	-362.937		-374.985		-362.904	

[7] Negative values for $\hat{\beta}$ indicate a lower probability for the event.

After the fitting a second model that took into account the first order interactions among predictors, we observed that none of them seemed to be significant for the response (Table 3).

The causal relationship as the one just described poses the researcher with the problem of how to interpret the effect exercised by the set of predictors on the response variable *Y*. For example, if from the base model in Table 2 we do not include the SCIEN predictor, we can note that the variable SEXM has a significant influence (even though the model is poorer in terms of log*Lik*). Nevertheless, if we fit a new model that considers the interaction term between SCIEN and SEXM, it is not significant.

5.2 Fitting a Boolean logit

To fit a Boolean logit[8] model we have considered the same predictors used for the standard logit (namely, SEXM, LICCS, DIP90, SCIEN, LAU26, VOTOHIGH, CORPOST). To fit the Boolean logit model it is necessary to posit some preliminary conditions. Taking into account the results we obtained with the binary segmentation analysis we decided to set the following conditions:

- A_1 = "ownership of *winning* qualifications for the job market"
- A_2 = "ownership of characteristics pertaining to the educational profile".

A_1 is defined by a set of covariates that refers to some of the most notable characteristics of someone who is seeking to enter the job market, namely *age* and *skills*: LAU26 and SCIEN. A_2 is defined by a set of covariates that refers to the educational background of the graduate, plus the variable sex, namely, DIP90, LICCS, VOTOHIGH, CORPOST and SEXM.

The probability of being employed, $Pr(Y=1)=\pi$ is modelled as the interaction among A_1 and A_2:

$$\pi = Pr(A_1) \times Pr(A_2)$$

The conditions A_1 and A_2 are expressed as an additive function of the explanatory variables considered:

- A_1 = LAU26 + SCIEN
- A_2 = SEXM + DIP90 + LICCS + VOTOHIGH + CORPOST.

As we can note from Table 4, the results are rather similar to those we had with the standard logit both in terms of log likelihood and parameter estimates. Nevertheless, the underlying models are quite different: in the standard logit, no interaction term shows significant effects on the response and therefore the model suggests that no variable influences the probability of dropping-out independently of the remaining ones.

On the contrary, in the Boolean logit, the model shows that the response is

[8] To process the data we applied the "Boolean" library available for R package (R Development Core Team, 2003).

Table 4. Point estimates ($\hat{\boldsymbol{\beta}}$) and z-scores (z= $\hat{\boldsymbol{\beta}} / SE(\hat{\boldsymbol{\beta}})$) for the standard logit model and two Boolean logit models

Covariates	*Standard*		*Boolean 1*		*Boolean 2*	
	$\hat{\beta}$	|z|-*score*	$\hat{\beta}$	|z|-*score*	$\hat{\beta}$	|z|-*score*
LAU26	0.6134	2.875	1.0330	2.996	1.0363	1.904
SCIEN	1.4855	4.315	2.2442	2.813	2.2573	1.192
LICCS	-0.2801	1.417	-1.4221	1.094	0.0070	0.008
SEXM	0.1967	0.916	0.5035	0.796	0.4973	0.501
DIP90	0.5453	2.068	1.3693	1.715	1.3640	1.326
VOTOHIGH	-0.2575	1.301	-1.2291	1.251	-1.2127	0.527
CORPOST	-0.4493	2.309	-1.6930	1.066	-1.6632	0.408
LICCS	–	–	–	–	-1.4294	0.904
log*Lik*	-362.937		-360.640		-360.639	

produced by multi-vector interactions, namely interactions of vectors of variables. Thus, it has been assumed that a student's dropout from University depends jointly on A_1 and A_2, with the parameters tacitly maintaining an "interaction meaning". The parameters DIP90 and CORPOST show no significant influence on the response Y and this means that in interacting with the other variables these covariates lose their predictive power.

The Boolean logit allows the researcher to fit models where the same covariate is set in more than one "condition". For example, the variable *lyceum as high school* (LICCS) could be inserted both in condition A_1 and in condition A_2. The results of the fitting are in Table 4 in the "Boolean 2" columns. In spite of the variable LICCS being statistically not significant, it acts on the response in opposite directions if considered in A_1 or in A_2.

6. Conclusions

The use of standard logit to model the probability of a dichotomous event as the effect of a causal relationship from a set of predictors gives researchers some useful tools to understand social phenomena. These tools (namely the log-odds ratios) allow the researcher to interpret the role played by each predictor on the response controlling for the remaining parameters.

Modelling the probability of graduates of finding an occupation, the standard logit allows us to identify factors that negatively influence the probability of 'getting an occupation'. Among these factors, the graduates who obtained high marks at graduation and in their postgraduate studies are likely to be older and therefore less attractive for the companies when they try to enter the job market.

We have seen that other factors play an opposite role (they contribute to increasing the event probability). Among these, we can highlight the case of science graduates and the importance of completing a study programme when still relatively young. We should remember that the fitted standard logit does not take into account possible interaction among covariates. Such an assumption entails a form of additive causal dependence that makes it difficult for the model to catch the full complexity of the phenomenon studied.

Boolean logit is neither an alternative nor a method better than the standard logistic one, but it does offer an advantage: it allows the researcher to consider models that consider causal complexity. Causal complexity mechanisms make it possible to improve the predictive power of the variable response models. Their major drawback lays in the subjective choice of the probability statements that lead to the combination of predictors.

However, the likelihood-based criteria for choosing the best model tend to mitigate this subjectivity. Another drawback is that (unlike standard logit) the parameters are not (log) odds-ratios for the response. Finally, the method is computationally time consuming.

Nevertheless, considering the encouraging results gained in this and in other applications (Muggeo & Porcu, 2004) we can conclude that Boolean logit is a useful tool for implementing sensitivity analyses of other models and for re-enforcing the evidence that emerges regarding the meaning of the predictors studied.

References

BRAUMOELLER B.F. (2003) Causal complexity and the study of politics, *Political Analysis*, **11**: 209-233.

CARINCI F., PELLEGRINI F. (2001) *RECPAM/SAS (Recursive Partitioning and Amalgamation): a statistical tool for criterion-driven data-mining*, Technical Report (http://med.monash.edu.au\publichealt).

CHIANDOTTO B. (2004) La situazione occupazionale dei laureati: dall'indagine alla pianificazione degli interventi sui percorsi formativi. In: CIVARDI M. (ed) *Transizione Università-Lavoro: la definizione delle competenze*, CLEUP, Padova: 1-18.

CIAMPI A. (1991) Generalized Regression Tree, *Computational Statistics and Data Analysis*, **12(1)**: 57-78.

CIVARDI M., ZAVARRONE E. (2004) Proposta di un modello generatore delle competenze acquisite attraverso la formazione universitaria. In: AURELI CUTILLO E. (ed) *Strategie metodologiche per lo studio della transizione Università-Lavoro*, CLEUP, Padova: 141-152.

FROSINI B.V. (2004) Causality and Causal Models. In: *Atti della XLII Riunione della Società Italiana di Statistica, Volume 1*, CLEUP, Padova: 3-32.

HOSMER D.W., LEMESHOW S. (1989) *Applied Logistic Regression*, John Wiley & Sons, New York.

MUGGEO V., PORCU M. (2004) Factors that Cause University Students to Drop Out. An Alternative Modelling of Interaction Terms in Logistic Regression Models. In: *Atti della XLII Riunione della Società Italiana di Statistica, Volume 2*, CLEUP, Padova: 511-514.

PORCU M., PUGGIONI G. (2004) Condizione occupazionale e prima valutazione del fenomeno dell'emigrazione dei laureati dell'Ateneo di Cagliari. In: D'OVIDIO F. (ed) *Professioni e competenze nel lavoro dei laureati*, CLEUP, Padova: 317-326.

PORCU M., TEDESCO N. (2004) Dall'Università al Lavoro: analisi dei tempi di passaggio dei laureati dell'Ateneo di Cagliari. In: AURELI CUTILLO E. (ed) *Strategie metodologiche per lo studio della transizione Università-Lavoro*, CLEUP, Padova: 281-295.

RUCZINSKI I., KOOPERBERG C., LEBLANC M. (2003) Logic Regression, *Journal of Computational and Graphical Statistics*, **12**: 475-511.

R DEVELOPMENT CORE TEAM (2003) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, (http://R-project.org).

TEDESCO N. (2002) Analisi di segmentazione di una coorte di immatricolati dell'Università di Cagliari. In: PUGGIONI G. (ed) *Modelli e metodi per l'analisi dei rischi sociali e sanitari*, CLEUP, Padova: 141-160.

The Transition from University to the Job Market. An Emigration Analysis of the University of Cagliari Graduates

Mariano Porcu, Giuseppe Puggioni, Nicola Tedesco[1]

Department of Economic and Social Sciences, University of Cagliari, Italy

Summary. In this work, we analyze data provided by official records on graduates of Sardinian universities in the period 1991-2001 with the aim of obtaining contextually relevant information for the realisation of a survey on the transition from university to the job market. Examining data from the survey, we looked in particular at certain aspects of the conditions of graduates in search of employment, their postgraduate education and the opportunities for entering the labour market. We also present some initial results regarding the emigration of graduates and the evaluation of human capital.
Keywords: Graduates, Emigration, Survey, CATI, Human Capital.

1. Introduction

In the years immediately before the “3+2” Italian university reform, the percentage of high school graduates enrolling at Sardinian universities was about 65% for males, and as high as 75% for females. At the beginning of the 80s this rate actually decreased, coming more or less into line with national averages.

In this work, we analyse administrative data in order to carry out a survey on University of Cagliari graduates and their transition from university to the job market. The goal is to build up a general framework for analysing a local context characterized by the low input-output dynamics of university students.

Taking a sample of graduates of University of Cagliari in the years 1999 and 2000, we analyse their positions “vis-à-vis” the employment market,

[1] This paper is the result of the joint ideas and efforts of the three authors. M. Porcu was responsible for the final editing of Sections 2.1, 3 and 5, G. Puggioni for Sections 1 and 6, whereas N. Tedesco was responsible for the other Sections.

their postgraduate education and their opportunities for entering the labour market.

In addition, we present some initial results on the emigration of graduates and the evaluation of human capital. With regard to this latter aspect, we should underline that this analysis is preliminary. In fact, according to some recent definitions of human capital, recent theories have recognised that studies on human capital must take into account the concepts of social capital and training. These theories (Morgan, 2000) have begun to define the concept of social capital in terms of “resource capital” development and how this can be related to the resources that an organised society is prepared to make available for the cultural and social development of it members.

We should also note what all sectors of the economy have for some time been stressing, i.e. the need to have a workforce educated and trained in strict view of the new skills required by technological innovation. A region such as Sardinia, which is characterised by a slow economic development and a new phenomenon of emigration, is an interesting context in which to study which aspects determine the emigration of graduates.

2. Time series dynamics for graduates and higher education programmes at Sardinian universities

In the period 1991-2001 in Sardinia we had 27,599 graduates: 0.64% not resident in the region, 67.8% graduates at the University of Cagliari, 62.1% female (61.1% in Cagliari and 64.1% in Sassari).

Time series of indexes (base year 1992=100) of graduates in the period 1992-2000 in Sardinia, as pointed out by the high value of the correlation coefficient (r=+0.97), show a similar increasing trend for the two universities. This trend, as shown in Figure 1, has become more evident in the last three years (1998-2000); for the University of Cagliari the index goes from 151 to 190. In the same period for the University of Sassari it goes from 136 to 167. If we examine the time series for female graduates, we see that the trend is the same as it is for graduates as a whole (MF *vs* F r=+0.98).

The number and typology of university courses in Sardinia is less varied than it is on the Italian mainland, though not greatly so (Figure 2). The two Sardinian universities differ in the courses they offer because some courses are run in one university, but not in the other. What we can derive from this situation is a non-negligible dissimilarity (z=0.34) between distributions of graduates according to type of degree in the two universities.

The comparison between Cagliari and Sassari according to type of degree show a slightly higher concentration[2] for Cagliari (Cagliari g=0.42; Sassari

[2] Concentration was measured with Gini’s index of concentration g (Leti, 1983).

g=0.39). For the purpose of this work, the different degrees have been classified into four groups: a) Economic-Legal-Social (ELS); b) Scientific - Technical (SCT); c) Sciences of Health and Life (SHL); d) Humanities and Behavioural Sciences (HBS).

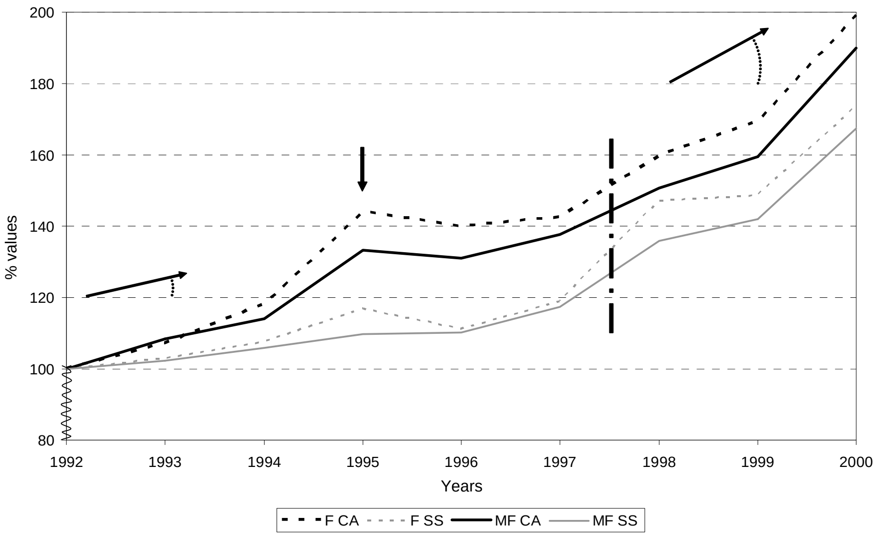


Figure 1. Graduates in Sardinia from 1992 to 2000 (Index: base year 1992=100)

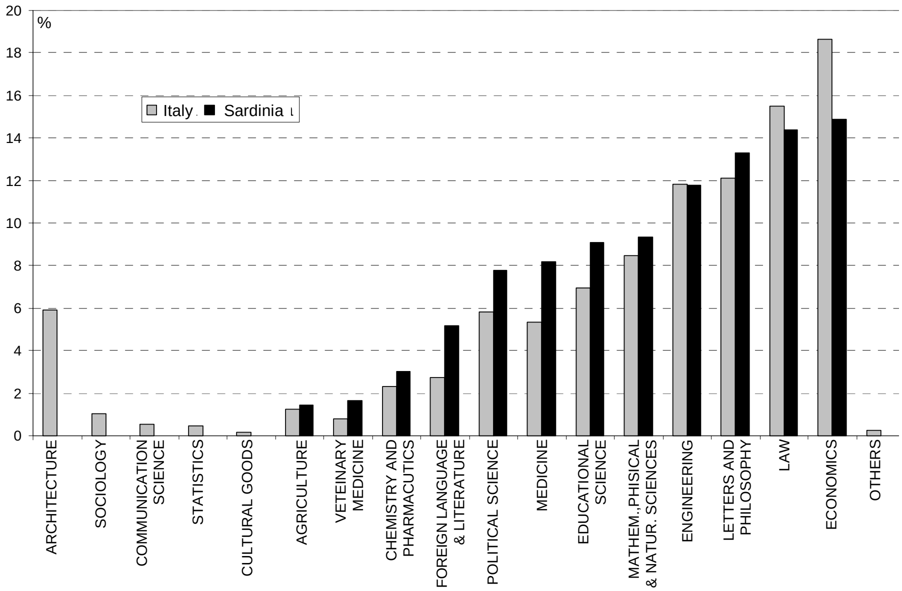


Figure 2. Percentage of graduates in 1999-2000 in Italy and Sardinia, by faculty

2.1 *Input-output* dynamic of graduates in Sardinia

The factors that influence social and economic life in Sardinia relate to the labour market (Bottazzi, 1999). Data on this aspect is common knowledge: during the 1950s and 1960s, with a trend evident in regions at a similar stage of development, there was the apparently strange phenomenon of a simultaneous decrease in the number of employed and of workforces. This decrease resulted, for that period, in a low unemployment rate with values of approximately 5%.

The reduction in workforce rate in the transition from an agricultural to a modern industrial society is a well-known phenomenon; emigration from Sardinia (in those years there was an average annual negative balance of about 10,000 units) reduced the number of the working population. From the mid-seventies, the situation drastically altered when the unemployment rate began to grow constantly until it reached peaks of up to 20% in the 1980s, values that remained, with some oscillations, for most of the 1990s.

The main causes of this situation (Bottazzi, 1998) were the extraordinary increase in the numbers of those in search of work[3] due principally to the effects of the sixties' *baby boom*[4], by the slowdown of migratory dynamics and, moreover, by the increasing number of women entering the labour market.

In this context of high unemployment, what reduces the risk of exclusion from the labour market is an educational qualification (Bottazzi, 1999). This is particularly true for those who have obtained a diploma at a vocational type high school. Graduates also have a greater probability of entering the labour market and show a greater inclination to seek employment outside Sardinia.

With regard to this latter phenomenon, from an analysis of changes of residence from and to Sardinia between the two last censuses, we have made an initial evaluation of the graduates' emigration. Considering that almost all graduates in Sardinia are resident in the region, according to public registry records for graduates up to 35 years of age on completion of a degree (94.8% of graduates of decade), the emigration rate is about 10% and the proportional migration index is -8.3. This result is just approximate evaluation because the registry office counts emigrations only from changes of residence and those who emigrate from Sardinia usually maintain their official residence for a not necessarily short period.

This information will be of particular interest because if it is true that the social capital of graduates represents a fundamental condition to guarantee constant scientific and technological progress, it is ever more a key requirement to activate development policies in Sardinia that are coherent to present

---

[3] In the period under consideration, the number of jobs available increased more than employment levels (+1.4% against +0.4% annual average) but the workforce numbers increased by 2.1% per year.

[4] By *baby-boom* we refer to the rapid increase in births in Italy during the 60s.

economic wisdom. Furthermore, as is well known, we can add that the emigration of graduates is usually a selective process: younger and more qualified graduates choose to emigrate to develop themselves professionally.

If we look at figures for the year 2003 in an international context in which Western countries are trying to reduce immigration, the USA issued almost 200.000 green cards for highly qualified foreign professionals. The United States have effectively enriched their human capital, while the ‘donor’ countries of origin of new immigrants have lost “highly valuable resources”. These countries have of course incurred the heavy costs of their education.

Gini (1953) examined the effects of this process as early as 1940. He quantified (based on conservative estimates and not ignoring relative advantages for countries of origin owing to a decrease in the labour supply) what was a great accumulation of wealth (notwithstanding money remittances and re-emigrations to homelands) for the USA economy[5].

3. A survey on University of Cagliari graduates

In an initial step aimed at studying graduates’ transition from university to the job market, the survey was carried out on a sample of graduates from the University of Cagliari aged 35 years or less at the moment of graduation, over a two-year period (1999 and 2000). The decision to consider graduates from these two years was justified for two reasons. Firstly, it was possible to select a sample from a large population and secondly, because of the need to observe a fairly long retrospective period (≥ 34 months) for each of the graduates in order to study their transition for three years. We chose the cut-off age of 35 because older graduates probably come from the category of students that were already working before their graduation[6].

Thus, total population consisted of 4,363 units, classified by sex, age, groups of faculties (Table 1). Subsequently, using a stratified sample design by faculty, with selection proportional to the number of each stratum, a sample of 26% was selected from the total population.

Estimating emigration trends of Sardinian graduates with administrative data will only provide an approximate evaluation of the phenomenon because of the limitations of these sources. In fact, the registration or cancellation of resident status will not necessarily coincide with physical transfer to another town.

In November 2003 a CATI survey was carried out on a sample of graduates from the University of Cagliari (Porcu & Puggioni, 2003). This sample of

5 For an evaluation of this kind of migration, see also Todisco (2000) and Fondazione ISMU (2002).

6 We should note that graduates over 35 numbered 216 (4.7% of population). From this group we eliminated 16 graduates non-resident in Sardinia at graduation.

Table 1. Graduates of the University of Cagliari in the years 1999 and 2000 by sex, class of age and faculty group

Groups	*Female*				*Male*				*Total*			*Total for Groups*	
	< 26	*27-29*	*30-35*	*Total*	*< 26*	*27-29*	*30-35*	*Total*	*< 26*	*27-29*	*30-35*	*N*	*%*
ELS	326	396	215	937	204	225	155	584	530	621	370	1521	*38.8*
SCT	105	150	46	301	157	327	213	697	262	477	259	998	*22.9*
SHL	195	170	87	452	63	90	71	224	258	260	158	676	*15.5*
HBS	429	352	215	996	64	57	51	172	493	409	266	1168	*26.8*
Tot	1055	1068	563	2686	488	699	490	1677	1543	1765	1053	4363	*100,0*
%				61.6				38.4	35.4	40.5	24.1		*100.0*

Table 2. Age at graduation by faculty: some statistics for sample and population

Faculty groups	*Mean*	*Sd*	M_e	$Q_3 - Q_1$	μ *(pop.)*
ELS	27.78	2.64	28.00	3.00	27.78
SCT	28.37	2.74	28.00	4.00	28.11
SHL	27.52	2.47	27.00	3.00	27.75
HBS	27.72	2.94	27.00	4.00	27.52

1,112 units represented almost a quarter of all graduates aged below 36, who graduated in either 1999 or 2000 according to official administrative records[7].

With respect to age at interview, sex and high school attended, the sample showed good overlap with the overall population. Considering those interviewed with regard to their faculty grouping, even the mean age at graduation seemed to overlap with the population (Table 2). A total number of 697 female and 415 male graduates were interviewed. 62% of those interviewed had attended “Lyceum”[8] and 17.5% had a Technical school background.

4. Occupational placement of the graduates

Of the 1,112 graduates interviewed, a proportion of 74.0% were working at the time of survey; 13.7% graduates were still looking for a job. All the graduates had had at least 34 months in which to get a job; 12.3% were still involved in postgraduate studies.

As regards gender, male graduates that were in work comprised 78.3% of their total; the corresponding figure for female graduates was 71.4%. The un-

[7] Only the 5% of graduates were aged over 35 at graduation in 1999-2000.

[8] The “Lyceum” provides a classical education such as the one offered by the old British “Grammar Schools”.

employment rate among males was 7.5% and 11.0% among females. Both males and females have the same percentage of graduates choosing to continue their studies (11.6% males, 12.8% females). As a benchmark for comparisons, we can note that in Sardinia in 2003 the total unemployment rate was 18.5%; the total unemployment rate in Italy for graduates aged 25 to 34 was 13.6% (ISTAT, 2004).

If we look at the categories ordered by degree type, we note that the highest rates of employment are among graduates from the faculties of Engineering (88.7%), Educational Sciences (84.9%), Pharmacy (81.3%), Economics (77.3%) and Sciences (77.1%). The lowest rate of employment was registered for graduates from the faculty of Medicine, but there is a high percentage (76.8%) of graduates in medicine who are still enrolled in postgraduate courses.

The majority (72.1%) of graduates find work as company/state employees. Only for Law faculty graduates do we note the opposite: employee workers comprise 32.3% *vs.* freelance workers (67.7%). More than half (56%) of those in employment are not working in the same city in which they hold official resident status.

After Medicine, the faculty of Law registers the highest percentage of graduates that have continued their studies in a postgraduate programme. The percentage of Philosophy graduates in work totals 65.8% (19.2% are unemployed, against an overall unemployment rate of 9.7%). Next we find the faculty of Political Science: 70.0% of graduates are in work and 18.0% are unemployed.

A high proportion (59.1%) of the employees is working for private companies; 34.7% are working in the public sector (the 2001 census revealed that 25% of the total number of workers is engaged in the public sector).

If we classify employees according to the type of employment contract held, we find that 49.1% of employees have permanent contracts and 26.6% have fixed term contracts, with 20% of the latter being termed atypical. The majority of those working for private firms (59.3%) have permanent contracts. 44.7% of those working in the public sector have fixed term contracts. The percentage of atypical contracts is the same in both the private and public sector (20%).

The highest number of graduates in employment with a permanent contract is from the faculties of Economics (74.5%), Pharmacy (68.0%), Engineering (65.7%) and Law (51.6%). On the contrary, graduates from the faculties of Philosophy (50.8%) and Sciences (40.5%) tend to be working in precarious conditions or on fixed term contracts. Graduates from the faculties of Humanities and Languages are among those working with the highest number of atypical contracts, with percentage values of 45.6% and 36.0% respectively.

We can note that for the public sector the majority of workers (72.8%) are winners of job competitions (the normal procedure in Italy for assigning candidates for public sector jobs); the most common means of obtaining a job in the private sector is by job interview (65.2%) or through direct hiring (15.2%).

91.9% of the 823 graduates who were working at the time of the survey started their job after graduation; 59.8% of them had obtained a postgraduate qualification and of these 74.7% stated that they had completed such studies.

71.8% of the workers that started their job after graduation found employment as regular employees. It is only for the Law faculty that we see the opposite trend, with 70.7% of graduates starting their careers as freelance workers. The overall percentage figure for private sector employees is 66.1%. The faculty with the highest amount of graduates in the private sector is Engineering (79.8%). More than a half (55.8%) the employees are working in the service sector (28.6%) and education (27.3%). For all the groups of faculties considered, the majority of employees are working in these sectors. The exception is the prevalence of engineering graduates finding work in the building sector (29.8%); only 12.5% of engineers are employed in the education sector.

5. Graduates that migrate

Of the 756 interviewed who started their working career after graduation (Table 3), 10.6% said that they were working outside Sardinia. One in five of these were working abroad. Almost all (94%) are employees.

More than 2/3 of the 80 emigrant graduates came from the faculties of Engineering, Sciences and Economics (graduates from these faculties make up 44.3% of the total). Therefore, the faculty of Engineering produces graduates that, in comparison with other faculties, have greater success in job placement not only in the local labour market but also outside Sardinia.

It is particularly interesting to highlight some characteristics of graduates that emigrated compared with those of graduates that are working in Sardinia.

Table 3. Comparison between some distributions for graduates that are working in Sardinia and those who have emigrated

Faculty	*Graduates that work:*	
	in Sardinia	*out of Sardinia*
	F / M x 100	
Engineering	24.5	16.0
Economics	127.3 }156	140.0
Sciences	213.1	140.0
	Average mark at graduation	
Engineering	104.0	105.8
Economics	102.9	106.2
Sciences	106.0	106.7
	Average age at graduation	
Engineering	28.4	27.8
Economics	27.7	27.3
Sciences	27.5	27.2

We only consider graduates from the three faculties that produce the majority of emigrant graduates (i.e. Engineering, Economics and Sciences).

As expected, the female/male percent rate for graduates who have emigrated is lower than that of those who work in Sardinia: Engineering 16% (25% in Sardinia), Economics and Sciences 140% (156% in Sardinia).

On examining the graduate population as a whole, we see that the average mark at graduation for emigrant graduates is significantly higher (α=0.05) compared to those that are working in Sardinia. On the other hand, the average age at graduation does not seem to be significantly different (p=0.408).

It is important to note that when we ask graduates that have emigrated and graduates that work in Sardinia their opinion on the relevance of their degree to their job, only 11% of the engineers that are working in Sardinia think that it is not relevant; the same opinion is expressed by 29% of emigrant engineers. Graduates in Economics or Sciences say the exact opposite; 23% of those who work in Sardinia said that the degree is not relevant. For graduates who have emigrated the percentage saying their degree is not relevant, is 17%.

With regard to using competencies acquired during their university studies, almost all graduates (93% for Engineering, 91% for Economics and Sciences) working in Sardinia use them at work. Among those who have emigrated, the percentage is considerably lower, especially among engineers (74%).

Such a result seems to be in contrast with the opinions expressed on the relevance of the competence acquired. Indeed, almost 40% of emigrant graduates say that the skills acquired are relevant. Regarding graduates that work in Sardinia, such an opinion was expressed only by 23% of graduates in Engineering and by 26% of graduates in Economics or Sciences. Such issues raise two points:

Table 4. Comparison between some distributions of opinions expressed by graduates working in Sardinia and by those who have emigrated

Faculty	*Graduates that work*	
	in Sardinia	*out of Sardinia*
	% degree and work not relevant	
Engineering	11.1	28.6
Economics	18.2	27.3
Sciences	29.8	8.3
	% use competences	
Engineering	93.4	74.1
Economics	95.9	83.3
Sciences	84.6	80.0
	% relevance of competence acquired	
Engineering	23.0	39.3
Economics	21.2	36.4
Sciences	32.8	41.7

- Is the local job market looking for competences different from those provided by the degree courses at the University of Cagliari?
- Given the competences required by the job market, do the graduates that have emigrated consider they have adequate skills, as compared with those of other workers with the same skills?

Finally, it is also important to note that a greater percentage of emigrant engineers are looking for a new job compared to those who are working in Sardinia (37% *vs* 31%). For those that have a degree in Economics or Sciences we can note that 37% of emigrant graduates are looking for a new job but this percentage increases notably among those working in Sardinia. For all the three kinds of degrees, the reason for wanting to change job is dictated by financial motives. If only emigrant workers are considered, the percentage of Science or Economics graduates that want to change their occupation for financial reasons is close to 90%; among engineers, it is only 50%.

6. Final remarks

The evaluation of graduate emigration obtained analysing administrative data seems to give results that underestimate the real importance of the phenomenon. The exploratory analysis of the data collected with the sample survey carried out on graduates of the University of Cagliari enabled us to obtain an initial evaluation of the phenomenon of graduate emigration. First, the total unemployment rate for Sardinian graduates aged 25-34 is 13.5%, which is very close to the national average. This is an important statistic, considering that Sardinian has a much higher unemployment rate than the rest of Italy; so it seems quite clear that a university degree does positively influence the probability of getting a job. Yet, this does not mean that all working graduates in Sardinia are doing a job that is relevant to their degree course.

Another important factor concerns those graduates that emigrated to find a job. The data are particularly worrying not in strictly quantitative terms, but qualitatively. Emigrant graduates tend to obtain good grades in a technical-scientific degree. In a context such as the Sardinian labour market, that needs well-designed policies to support local development, it will be crucial to be able to exploit qualitative human capital, in other words the very skills possessed in particular by those that emigrate to find work.

References

BOTTAZZI G. (1998) *Disoccupazione e mercato del lavoro*, Isic-Tecnofor, Nuoro.

BOTTAZZI G. (1999) *Eppur si muove! Saggio sulle peculiarità del processo di modernizzazione in Sardegna*, CUEC, Cagliari.

FONDAZIONE ISMU (2002) *Ottavo rapporto sulle migrazioni 2002*, Franco Angeli, Milano.

GINI C. (1953) Cause apparenti e cause effettive della prosperità americana. In: *Scritti di sociologia e politica in onore di Luigi Sturzo, Vol. II*, Zanichelli, Bologna: 225-252.

ISTAT (2004) *Annuario delle Forze di Lavoro 2003*, ISTAT, Roma

LETI G. (1983) *Statistica descrittiva*, Il Mulino, Bologna.

MORGAN S.L. (2000) Social capital, capital goods and the production of learning, *The Journal of Socio-Economics*, **29**: 591-595.

PORCU G., PUGGIONI G. (2003) Laurea ed abbandono universitario. Uno studio comparativo dei due eventi su un campione di immatricolati dell'Università di Cagliari. In: FABBRIS L. (ed) *LAID-OUT: scoprire i rischi con l'analisi di segmentazione,* CLEUP, Padova: 25-40.

TODISCO E. (2000) Mobilità dei cervelli e mobilità delle conoscenze. In: *Proceedings of the International Meeting "Migrazioni: scenari per il XXI secolo", session "Cultural migration: arts and folks moving" Florence 2000, 27-30 September* (http://www.cestim.org/index_dossier.htm).

University Studies and Employment. An Application of the Principal Strata Approach to Causal Analysis

Leonardo Grilli, Fabrizia Mealli

Statistics Department "G. Parenti", University of Florence, Italy

Summary. In this paper, we propose a methodology, based on the principal strata approach to causal inference, for assessing the relative effectiveness of two university study programmes with respect to the employment status of their graduates. The analysis relies on a parametric model fitted by maximum likelihood. In that context, we discuss some relevant modelling issues and the implications of the results for policy.
Keywords: Causal effects; Effectiveness; Employment; Principal strata.

1. Students' propensity for employment

Traditional analyses of the effect of study programmes on employment (also called *external effectiveness analyses*) are performed on the sole basis of graduated students, neglecting the fact that the students who are able to graduate in a programme are in principle different from those who are able to graduate in another one. In other words, two programmes could select different kinds of students with specific propensity for employment.

The analysis of the employment of graduates mixes the "direct" effect of the study programme on the employment status with the "indirect" effect through the graduation status. The possibility to disentangle the two effects could be important from a political point of view. For example, if there is a direct effect on employment, then the programme with smaller effectiveness should try to adapt its contents in order to match labour market requirements.

If, instead, the occupational success of a programme is merely due to different selection criteria of the university career (e.g., one programme is more difficult than the other one and thus selects better students), the problem becomes an issue of educational policy. We should evaluate whether it is desirable for the society to graduate students with low ability or to allow the existence of study programmes with different difficulty levels.

To study the direct effect of programmes on employment, avoiding the possible bias caused by different graduation processes, it is necessary to envisage a joint study of graduation and employment. In this respect, a convenient framework is that of principal stratification (Frankgakis & Rubin, 2002), a development of the potential outcomes approach to causal inference (Rubin, 1974). Barnard *et al*. (2003) recently used the framework of principal stratification for the analysis of a complex randomized experiment in the educational context.

In the following, the *treatment variable* is the degree programme, while the *intermediate (post-treatment) variable* defining the principal strata is the graduation status (graduated or not). The key point is that if a student does not graduate, the *outcome variable*, that is the employment status, cannot be defined for assessing the external effectiveness of a given programme. This is an example of the so-called *censoring by death* (Zhang & Rubin, 2003).

The present analysis is limited to the comparison of two study programmes. The extension to three or more degree programmes entails some technical difficulties, but the conceptual framework would remain unaltered.

We compare the programmes of Economics and Political Science of the University of Florence, which are supposed to be similar with respect to the contents of the courses and to the job market opportunities. In the light of this similarity, the choice of a student to enrol in a programme should be weakly related to unobserved characteristics that potentially affect also graduation and employment status, so the ignorability assumption discussed later seems reasonable.

The paper is organised as follows. Section 2 describes the data, while Section 3 outlines the principal strata framework and the probabilistic structure used to model the data at hand. Section 4 describes the model fitting and shows the main results. Section 5 concludes the paper with some remarks.

2. The data

A joint analysis of the academic careers and job opportunities of university students requires the merging of two data sources: an administrative database of a cohort of freshmen and survey data on employment of the graduates belonging to that cohort. The two sources are: (*i*) the administrative database of the 1992 cohort of freshmen enrolled in the two programmes to be compared; and (*ii*) three surveys on the occupational status of the graduates of the years 1998, 1999 and 2000, respectively. The matriculation number is used to merge the datasets.

Overall, 1941 freshmen belong to the examined 1992 cohort: 1068 enrolled in Economics and 873 in Political Science.

Table 1. Percent distribution of curriculum status of the 1992 cohort by the end of the year 2000

Status	*Economics*	*Political Science*
Dropped	51.0	60.9
Graduated	25.3	20.2
Still enrolled	23.7	18.9
(n)	(1068)	(873)

Table 2. Employment status at the interview of the 1992 cohort

Status	*Economics*	*Political Science*
Graduates (n)	(270)	(176)
% graduates interviewed	69.3	56.2
% graduates with permanent job	51.9	36.4

The choice of the 1992 cohort is motivated by the availability of survey data for the graduates of the years from 1998 to 2000. Only 21 students of the 1992 cohort graduated before 1998, while among the students who did not drop out, the majority of them graduated in the triennium 1998-2000.

The status of the students by the end of the year 2000 is summarised in Table 1. For the students still enrolled at the end of 2000 there are no data on their employment status. Therefore, for the purpose of the present analysis, *graduation* means “graduation within nine years from enrolment”. This definition is not particularly harmful for the analysis since graduation after nine years typically concerns students who already have a regular job during their studies.

All the interviewees responded to the question on the employment status. Apart from 21 students who graduated before 1998 and are out of target, almost all missing interviews are due to failed contact. The outcome variable of the analysis is a binary indicator of permanent job at the time of the interview, i.e. from one to two years after graduation. The employment status at the interview for the subset of graduated students is reported in Table 2.

The administrative database includes some additional information on every student of the 1992 cohort, which is used to define five dichotomous covariates: gender, residence (Florence *vs.* others), high-school degree (Gymnasium *vs.* others), high school grade (high grade, i.e. 50-60, *vs.* low grade, i.e. 36-49), late enrolment (i.e. the student did not enrol soon after high school). Table 3 reports the sample means of the covariates.

The covariates have different distributions in the two programmes, i.e. in the two treatment groups, since the assignment mechanism is not random. In particular, the high school grade is higher for the Economics students. The most striking difference concerns late enrolment, which is relatively rare in Economics, but reaches 22% in Political Science.

Table 3. Percent estimates of the covariates for the two study programmes

Covariate	*Economics (n=1068)*	*Political Science (n=873)*
Female	0.41	0.54
Residence in Florence	0.23	0.31
High school: gymnasium	0.34	0.45
High grade at high school	0.37	0.25
Late enrolment	0.06	0.22

3. The principal strata representation

Let n denote the total number of individuals under study, i.e. the dimension of the 1992 cohort of freshmen enrolled in Economics or Political Science. The *treatment variable* Z_i is thus defined as $Z_i = 1$ if student i was enrolled in Economics, and $Z_i = 0$ if student i was enrolled in Political Science.

Now let z_i denote the realized value of Z_i and let $\mathbf{z}$ denote the vector of z_i for all n individuals. In the potential outcomes framework every post-treatment variable, i.e. any relevant variable that takes its value after treatment assignment, depends on the vector of treatment assignments $\mathbf{z}$. However, in the present application it is reasonable to make the standard SUTVA (*Stable Unit Treatment Value Assumption*): for any individual i every post-treatment variable depends on $\mathbf{z}$ only through z_i. This excludes possible interactions between individuals.

Given SUTVA, every post-treatment variable has many potential versions as the number of possible treatments (two in the present application). Therefore, the post-treatment variables can be defined as follows.

The first potential post-treatment variables are the intermediate variables $S_i(z)$, z=0,1: $S_i(z) = 1$ if student i graduated by the end of 2000 (i.e. within 9 years) when enrolled in study programme z, and $S_i(z) = 0$ otherwise.

The availability of survey data for the graduates up to the year 2000 suggested to identify S as to represent the event “the student graduated by the end of 2000”; in this way the groups “Dropped” and “Still in course” collapsed in the same category.

Other potential post-treatment variables are the response indicators $R_i(z)$, z=0,1: $R_i(z) = 1$ if student i responded to the question on the employment status when enrolled in programme z and graduated, and $R_i(z) = 0$ otherwise.

The last potential post-treatment variables are the outcome variables $Y_i(z)$, z=0, 1: $Y_i(z) = 1$ if student i had a permanent job at the time of the interview, i.e. from one to two years after graduation, when enrolled in programme z and graduated, and $Y_i(z) = 0$ otherwise.

For each individual, the treatment variable assumes a single value; for every post-treatment variable, only one of the two versions can be observed. Therefore, we introduce the notation $S_i^{\mathrm{obs}} = S_i(Z_i)$, $R_i^{\mathrm{obs}} = R_i(Z_i)$, $Y_i^{\mathrm{obs}} = Y_i\,(Z_i)$.

Since both the treatment variable and the intermediate variable are dichotomous, four principal strata can be defined through the latent variable L_i:

- L_i = '*GG*' (Graduated, Graduated) if $S_i(1)=1$ and $S_i(0)=1$: students who are able to graduate in both study programmes;
- L_i = '*GN*' (Graduated, Not graduated) if $S_i(1)=1$ and $S_i(0)=0$: students who are able to graduate in the first programme (Economics), but are *not* in the second (Political science);
- L_i = '*NG*' (Not graduated, Graduated) if $S_i(1)=0$ and $S_i(0)=1$: students who are *not* able to graduate in the first programme (Economics), but are able to graduate in the second (Political science);
- L_i = '*NN*' (Not graduated, Not graduated) if $S_i(1)=0$ and $S_i(0)=0$: students who are *not* able to graduate in either programme.

Note that each student belongs to a single stratum, although the data cannot in general reveal which stratum a person belongs to. In other words, the principal strata are latent classes and the data only allow estimating the probability that a given student belongs to a certain latent class. Also, note that the couple of potential values of the intermediate variable define the principal strata, so the strata are not affected by the treatment and thus can be viewed as an unobserved pre-treatment covariate.

The relationship between the observed groups, defined by Z_i and S_i^{obs}, and the principal strata is described in Table 4, along with the corresponding supports of R_i^{obs} and Y_i^{obs}.

For the post-treatment variables the sample proportions in the two treatment groups are: $P_{S,1} = 0.253$ that is the sample proportion of graduates among students enrolled in Economics (Z_i=1), and $P_{S,0} = 0.202$ is the analogue for Political science (Z_i =0); $P_{Y,1} = 0.516$ is the sample proportion of Economics students with a permanent job (Z_i=1), who graduated ($S_i^{obs} = 1$) and responded to the interview ($R_i^{obs} = 1$), while $P_{Y,0} = 0.364$ is the analogue for Political Science (Z_i =0).

Therefore, Economics has a higher graduation rate and a higher employment rate among the graduates. The analysis should assess if the better performance of Economics can be attributed to a positive causal effect.

Table 4. Observed groups, principal strata and variables of the study

Observed group $O(Z, S^{obs})$	Z_i	S_i^{obs}	R_i^{obs}	Y_i^{obs}	*Latent group* L_i *(principal stratum)*
$O(1,1)$	1	1	$\in \{0,1\}$	$\in \{0,1\}$	*GG* or *GN*
$O(1,0)$	1	0	not defined	not defined	*NG* or *NN*
$O(0,1)$	0	1	$\in \{0,1\}$	$\in \{0,1\}$	*GG* or *NG*
$O(0,0)$	0	0	not defined	not defined	*GN* or *NN*

Since the purpose of the study is to evaluate the effectiveness of graduation in a given programme with respect to the job market, the outcome variable Y is defined only for the graduates. Therefore the causal effect $Y_i(1)$-$Y_i(0)$ on the employment status is properly defined only for the *GG* stratum, i.e. the students who are able to graduate in both programmes.

In principle, if data were available, the outcome variable Y could be defined for all the enrolled students, allowing comparisons within the other strata. Anyway, such comparisons would not address the issue of relative effectiveness of graduation in different programmes.

The parameter of main interest is thus the *average* causal effect (ACE) for the *GG* stratum. When interest rests only on the population at hand, this estimand is simply the difference of the means of the two potential outcomes of Y for the individuals belonging to the *GG* stratum: $\overline{Y}_{GG}(1) - \overline{Y}_{GG}(0)$.

However, we are interested on the more general data generation mechanism, so the results will be implicitly referred to a superpopulation and expressed in probability terms. The estimand is thus the difference between the probabilities of having a permanent job under the two treatments, again for the *GG* stratum: $E(Y_{GG}(1)) - E(Y_{GG}(0)) = P(Y_{GG}(1) = 1) - P(Y_{GG}(0) = 1)$.

The probabilities of the principal strata are also of interest since they allow deepening the analysis of the effectiveness of the programmes with respect to graduation, as explained in Section 4.

Since Z is not randomised, there may be some confounders that influence both Z and S or both Z and Y: in such a case, the simple association between Z and Y cannot be interpreted as a causal effect. The available covariates $\mathbf{x}_i$, described in Table 3, may alleviate this problem and this is what underlies the following assumption (*Unconfoundedness of treatment assignment*): $Z_i \perp \{S_i(0), S_i(1), Y_i(0), Y_i(1)\} \mid \mathbf{x}_i$.

In the present application, this assumption would be violated if students with the same observed covariates would base their enrolment decision on reliable predictions on graduation and employment that depend on unobserved covariates. However, this behaviour seems unlikely, since the two competing degree programmes have many common features.

The data on the graduates' outcomes also suffer from a problem of missing data: in fact, even if the outcome variable Y is defined only for the graduates, it is available only for the interviewed ones. We assume that the information about Y is *Missing at Random*: $R_i(z) \perp Y_i(z) | \{\mathbf{x}_i, S_i(z)=1\}$ for each z=0,1. Under this assumption, the response mechanism is ignorable, so the analysis can be safely based on the available responses (conditional on observed covariates). Overall, the assumption of missing at random seems reasonable in the data at hand[1] because almost all missing interviews are due to failed contact.

[1] Mealli *et al*. (2004) discuss alternative assumptions on the response mechanism.

Under the assumptions of SUTVA, treatment unconfoundedness and missing at random, the data generating process can be defined in terms of the following two sets of probabilities:

A. *Probabilities of the principal strata*: $\pi_{GG:i}$, $\pi_{NG:i}$, $\pi_{GN:i}$, $\pi_{NN:i}$, where, for example, $\pi_{GN:i} = \Pr(L_i = 'GN' \mid \mathbf{x}_i)$ is the probability that student i belongs to principal stratum *GN*, i.e. he or she is able to graduate in Economics but not in Political Science;

B. *Probabilities of outcome conditional on the principal stratum*: $\gamma_{1,GG:i}$, $\gamma_{0,GG:i}$, $\gamma_{1,GN:i}$, $\gamma_{0,NG:i}$, where, for instance, $\gamma_{0,GG:i} = \Pr(Y_i^{obs} = 1 \mid Z_i = 0, L_i = 'GG', \mathbf{x}_i)$ is the probability that student i has a permanent job when he/she belongs to the principal stratum *GG* and graduated in Political science $(Z_i = 0)$. From the unconfoundedness assumption, $\gamma_{0,GG:i} = \Pr(Y_i(0) = 1 \mid L_i = 'GG', \mathbf{x}_i)$.

The probabilities corresponding to combinations of the programme and principal stratum, other than the four listed, are not defined in the present application. The probabilistic structure is analogous to that of latent class models, except that in the present case belonging to a certain latent class determines not only the values of the probabilities of Y, but also whether they are defined or not. The estimand of main interest is the *average causal effect (ACE) on employment in the GG stratum,* i.e. the difference between the probabilities of the outcome Y under the two treatments for individuals belonging to the *GG* stratum:

$$\Pr(Y_i(1) = 1 \mid L_i = 'GG', \mathbf{x}_i) - \Pr(Y_i(0) = 1 \mid L_i = 'GG', \mathbf{x}_i) = \gamma_{1,GG:i} - \gamma_{0,GG:i} \, . \quad (1)$$

Also the probabilities of the principal strata $(\pi_{GG:i}, \pi_{NG:i}, \pi_{GN:i}, \pi_{NN:i})$ are interesting in itself, as they throw light on the dynamics of the graduation process in the two programmes. In fact, the *average causal effect (ACE) on graduation* is

$$\Pr(S_i(1)=1 \mid \mathbf{x}_i) - \Pr(S_i(0)=1 \mid \mathbf{x}_i) = (\pi_{GG:i} + \pi_{GN:i}) - (\pi_{GG:i} + \pi_{NG:i}) = \pi_{GN:i} - \pi_{NG:i} \, . \quad (2)$$

Therefore, the probability of the *GG* stratum, $\pi_{GG:i}$, is irrelevant for the ACE on graduation, but it can still describe different scenarios. In particular, as $\pi_{GG:i}$ diminishes, the graduates of the two degree programmes tend to be more heterogeneous. Even in the case of a homogeneous population, the probabilities π's and γ's are not directly estimable from the data without further assumptions. In fact, there are three non redundant π's and four γ's, compared with only four sample proportions ($P_{S,1}$, $P_{S,0}$, $P_{Y,1}$, $P_{Y,0}$). In particular, $P_{S,1}$ and $P_{S,0}$ allow us to get a point estimate of the π's only after fixing one of them, as long as the π's are the same in both treatment arms. Moreover, the γ's cannot be directly estimated, since they are defined conditional on the principal stratum. Rather, the data allow to estimate (through $P_{Y,1}$ and $P_{Y,0}$) the following mixtures of probabilities, so that estimation requires some mixture deconvolution:

$$\gamma_{1,GG:i}\frac{\pi_{GG:i}}{\pi_{GG:i}+\pi_{GN:i}}+\gamma_{1,GN:i}\frac{\pi_{GN:i}}{\pi_{GG:i}+\pi_{GN:i}};$$
$$\gamma_{0,GG:i}\frac{\pi_{GG:i}}{\pi_{GG:i}+\pi_{NG:i}}+\gamma_{0,NG:i}\frac{\pi_{NG:i}}{\pi_{GG:i}+\pi_{NG:i}}. \tag{3}$$

4. Model specification and fitting

Model specification and estimation is a difficult task, since in the principal strata framework the latent groups lead to mixtures of distributions that are difficult to disentangle. The covariates are extremely useful to identify the model: identification can be achieved by several alternative restrictions whose plausibility should be judged case by case, as illustrated by Jo (2002) in the special instance of non-compliance with two latent groups. However, the likelihood function is usually flat, so its maximization is not trivial. A Bayesian analysis (Imbens & Rubin, 1997) may circumvent these difficulties, but, apart from the computational complexity, the choice of suitable prior distributions is tricky.

Here we perform a maximum likelihood analysis, which turns out to be effective for the problem at hand. As noted in the previous section, the data generating process can be defined in terms of two sets of probabilities, the π's, leading to the principal strata sub-model, and the γ's, leading to the outcome sub-model. The variables available for each individual are Z_i, S_i^{obs}, R_i^{obs}, Y_i^{obs} (if $R_i^{obs}=1$) and the vector of covariates $\mathbf{x}_i$.

In the present application, the 19 individuals with missing values on the covariates are simply deleted, so the covariates can be treated as fully observed. Extensions to handle missing values in the covariates are shown in Barnard *et al.* (2003).

Now let us collect the parameters in the vector $\boldsymbol{\theta}$ and the variables for the n individuals in the vectors $\mathbf{Z}$, $\mathbf{S}^{obs}$, $\mathbf{R}^{obs}$ and $\mathbf{Y}^{obs}$ and in the matrix $\mathbf{X}$. Then the likelihood can be written as a product over the four observable groups defined by Z_i and S_i^{obs}, where $i\in O(k,h)$ stands for $Z_i=k$ and $S_i^{obs}=h$:

$$\begin{aligned}
&L(\boldsymbol{\theta}\mid \mathbf{Z},\mathbf{S}^{obs},\mathbf{R}^{obs},\mathbf{Y}^{obs},\mathbf{X})=\\
&\prod_{i\in O(1,1)}\left\{\pi_{GG:i}\left[(\gamma_{1,GG:i})^{Y_i^{obs}}(1-\gamma_{1,GG:i})^{1-Y_i^{obs}}\right]^{R_i^{obs}}+\pi_{GN:i}\left[(\gamma_{1,GN:i})^{Y_i^{obs}}(1-\gamma_{1,GN:i})^{1-Y_i^{obs}}\right]^{R_i^{obs}}\right\}\times\\
&\times\prod_{i\in O(1,0)}\{\pi_{NG:i}+\pi_{NN:i}\}\times\\
&\times\prod_{i\in O(0,1)}\left\{\pi_{GG:i}\left[(\gamma_{0,GG:i})^{Y_i^{obs}}(1-\gamma_{0,GG:i})^{1-Y_i^{obs}}\right]^{R_i^{obs}}+\pi_{GN:i}\left[(\gamma_{0,NG:i})^{Y_i^{obs}}(1-\gamma_{0,NG:i})^{1-Y_i^{obs}}\right]^{R_i^{obs}}\right\}\times\\
&\times\prod_{i\in O(0,0)}\{\pi_{GN:i}+\pi_{NN:i}\}
\end{aligned}\tag{4}$$

The model is based on the assumptions of SUTVA, treatment unconfoundedness and missing at random. In the likelihood (4) the individuals who did not respond to the interview (R_i^{obs} =0) do not contribute to the estimation of the γs, but do contribute to the estimation of the π's. Therefore, the π's are estimated from all the individuals in the sample, while information about the γs is given only by the individuals who graduated and were interviewed (overall, 15% of the sample), so estimation of the γs relies on scarce data.

As in the majority of current applications of the principal strata approach, the treatment and the intermediate variable are binary, leading to four principal strata. While in many settings it is sensible to assume that certain strata are empty (e.g. the assumption of no defiers in an experiment without compliance), in the present context such assumptions are not plausible in the light of the symmetry of the two treatments, so all the strata exist. This level of generality implies a considerable increase in model complexity since, as it is clear from the likelihood (4) that every observed group $O(k,h)$ is generated by a mixture of two distributions to be disentangled.

The probabilities of the principal strata (π's) are subject to some restrictions since they must lie in the [0, 1] interval and must sum to one. Therefore, in order to model the dependence of these probabilities on the covariates it is useful to operate a transformation to a set of unbounded parameters, using a multinomial logit specification (where *NN* is the reference category). For example, the model probability for GG is

$$\pi_{GG:i} = \frac{\exp(\eta^{\pi}_{GG:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}.$$

For the probabilities of the outcome (γs) the transformation to unbounded parameters can be obtained through separate logit specifications. For example,

$$\gamma_{1,GG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{1,GG:i})},$$

and four linear predictors are defined: $\eta^{\gamma}_{1,GG:i}$, $\eta^{\gamma}_{0,GG:i}$, $\eta^{\gamma}_{1,GN:i}$ and $\eta^{\gamma}_{0,NG:i}$.

Then, the η^{π}'s and the η^{γ}'s are assumed to depend linearly on the covariates. In the most general version of the model, each of these parameters has its own set of regression coefficients. In the current application the most general model we consider entails an unconstrained linear specification of the η^{π}'s, e.g. $\eta^{\pi}_{GG:i} = \alpha^{\pi}_{GG} + \boldsymbol{\beta}^{\pi}_{GG}{}'\mathbf{x}_i$, and a linear specification of the η^{γ}'s with a specific intercept, but a common vector of slopes, e.g. $\eta^{\gamma}_{1,GG:i} = \alpha^{\gamma}_{1,GG} + \boldsymbol{\beta}^{\gamma}{}'\mathbf{x}_i$. In this way it is assumed that each covariate has the same effect in each principal stratum and that the ACE on employment in the *GG* stratum is additive on the logit scale, $\alpha^{\gamma}_{1,GG} - \alpha^{\gamma}_{0,GG}$, i.e. is the same for all levels of the covariates. In our application, this specification seems a reasonable one. Other specifications could be devised (Jo, 2002), but in the present case their use is hindered by the limited sample information.

Model identification is possible only with a suitable number of covariates. The model specification just outlined has 27 parameters and 5 covariates, so that conditions for identification are met (Grilli & Mealli, 2006). However, empirical underidentification problems are likely with models of this kind, and in fact in the model selection process we had to put some constraints on some parameters entering the η^{γ}'s.

Maximum likelihood estimation was performed by means of the NLMIXED procedure of the SAS system (SAS Institute, 1999). The procedure has several maximizing algorithms, the default being quasi-Newton with a BFGS update of the Cholesky factor of the approximate Hessian.

For certain values of the covariates, some principal strata have very low predicted probabilities, meaning that they are nearly empty. In particular, for the baseline individual, which was chosen to be the most frequent pattern in the sample and characterised by all the covariates being zero, the *NG* stratum seems to be empty, since the corresponding value on the multinomial logit scale is -7.826 (s.e. 14.763).

Therefore, in order to follow a simple and clear model selection strategy, we redefined the coding of the covariates in order to obtain a new baseline individual with non-negligible probabilities for all the four strata. The goal was achieved simply by switching the coding of the covariate *Late enrolment,* from now on labelled *Regular enrolment.*

In the unrestricted model, six of the estimated β^{π}'s are below –5 with huge or not available standard errors, meaning that when the covariate switches from zero to one the corresponding principal stratum disappears. In particular, with the exception of some students who enrolled late, the *NG* stratum appears to be empty. This is not surprising, since the overall proportion of graduates is modest and is lower for Z_i =0, so that the *NG* stratum (Not graduated if Z_i =1 and Graduated if Z_i =0) is necessarily very limited. In addition, the opposite *GN* stratum seems to be empty in some cases.

Therefore, model selection goes on by fixing to $-\infty$ the aforementioned β^{π}'s, leading to the results shown in Table 5. The reduction from 27 to 21 parameters entails a negligible reduction in the deviance, while the other parameters and standard errors are essentially unchanged. Some of the β^{π}'s are not significant at conventional levels, so the principal strata sub-model could be further refined. However, model selection was stopped since a parsimonious principal stratum sub-model is not an objective of substantive interest and has little effect on the precision of the estimates of the outcome sub-model.

In the outcome sub-model the β^{γ}'s are not significant at conventional levels, though two of them (*Gymnasium* and *Regular enrolment*) are high in magnitude: more data would be needed to assess the influence of the covariates on the outcome. Nonetheless, the primary target of the analysis, i.e. the causal effect on the logit scale $\alpha^{\gamma}_{1,GG} - \alpha^{\gamma}_{0,GG}$, is estimated as 0.666 with a s.e. of 0.301, so it is significantly different from zero at the 5% level.

Table 5. Parameter estimates (and standard errors) of the model

Principal strata submodel (π's)			*Outcome submodel (γ's)*		
α^{π}_{GG}	-4.402	(0.448)	$\alpha^{\gamma}_{1,GG}$	1.262	(1.241)
α^{π}_{GN}	-2.647	(0.752)	$\alpha^{\gamma}_{0,NG}$	-1.365	(1.568)
α^{π}_{NG}	-3.207	(0.835)	$\alpha^{\gamma}_{0,GG}$	0.596	(1.185)
$\beta^{\pi}_{GG,gymnasium}$	1.275	(0.157)	$\alpha^{\gamma}_{1,GN}$	0.484	(1.058)
$\beta^{\pi}_{GN,gymnasium}$	$-\infty$		$\beta^{\gamma}_{gymnasium}$	-0.410	(0.374)
$\beta^{\pi}_{NG,gymnasium}$	$-\infty$		$\beta^{\gamma}_{high\_grade}$	-0.036	(0.263)
$\beta^{\pi}_{GG,high\_grade}$	1.205	(0.146)	$\beta^{\gamma}_{regular\_enrolment}$	-0.932	(0.979)
$\beta^{\pi}_{GN,high\_grade}$	1.113	(0.652)	β^{γ}_{female}	0.070	(0.272)
$\beta^{\pi}_{NG,high\_grade}$	$-\infty$		$\beta^{\gamma}_{Florence}$	0.104	(0.333)
$\beta^{\pi}_{GG,regular\_enrolment}$	2.023	(0.425)	ACE on empl. *GG* $\alpha^{\gamma}_{1,GG}-\alpha^{\gamma}_{0,GG}$	0.666	(0.301)
$\beta^{\pi}_{GN,regular\_enrolment}$	-0.009	(0.792)			
$\beta^{\pi}_{NG,regular\_enrolment}$	$-\infty$				
$\beta^{\pi}_{GG,female}$	0.117	(0.137)			
$\beta^{\pi}_{GN,female}$	-0.622	(0.755)			
$\beta^{\pi}_{NG,female}$	0.991	(1.111)			
$\beta^{\pi}_{GG,Florence}$	0.280	(0.144)			
$\beta^{\pi}_{GN,Florence}$	$-\infty$				
$\beta^{\pi}_{NG,Florence}$	$-\infty$				

Table 6. Estimated percent probabilities for some covariates' patterns

Probability	**00000**	**00100**	**00110**	**00101**	**01100**	**10100**	**11100**	**11111**
$\pi_{GG:i}$	1.1	8.0	9.1	10.9	20.3	24.9	52.5	62.2
$\pi_{GN:i}$	6.3	6.0	3.3	0.0	14.0	0.0	0.0	0.0
$\pi_{NG:i}$	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\pi_{NN:i}$	89.0	86.0	87.6	89.1	65.7	75.1	47.5	37.8
ACE on graduation $\pi_{GN:i}-\pi_{NG:i}$	2.7	6.0	3.3	0.0	14.0	0.0	0.0	0.0
$\gamma_{1,GG:i}$	77.9	58.2	59.9	60.7	57.3	48.0	47.1	51.5
$\gamma_{0,GG:i}$	64.5	41.7	43.4	44.2	40.8	32.2	31.4	35.3
$\gamma_{1,GN:i}$	61.9	39.0	40.7	41.5	38.1	29.8	29.0	32.8
$\gamma_{0,NG:i}$	20.3	9.1	9.7	10.0	8.9	6.3	6.1	7.1
ACE on empl. for *GG* $\gamma_{1,GG:i}-\gamma_{0,GG:i}$	13.5	16.5	16.5	16.4	16.5	15.8	15.7	16.2
$\gamma_{0:i}$	30.6	41.7	43.4	44.2	40.8	32.2	31.4	35.3
$\gamma_{1:i}$	64.2	49.9	54.8	60.7	49.5	48.0	47.1	51.5
$\gamma_{1:i}-\gamma_{0:i}$	33.6	8.2	11.4	16.4	8.6	15.8	15.7	16.2

The pattern (x_1,x_2,x_3,x_4,x_5) stands for *Gymnasium* = x_1, *High grade* = x_2, *Regular enrolment* = x_3, *Female* = x_4, *Florence* = x_5

To aid the interpretation of the results, Table 6 reports the estimated probabilities for some covariates' patterns. The patterns are in increasing order of $\pi_{GG:i}$. Note that $\pi_{GN:i}-\pi_{NG:i}$ is the ACE on graduation (2) and $\gamma_{1,GG:i}-\gamma_{0,GG:i}$ is the ACE on employment in the *GG* stratum (1), while $\gamma_{0:i}$ and $\gamma_{1:i}$ are the probabilities of employment for the two programmes obtained through the mixtures defined in (3).

The estimated proportion of students belonging to the *GG* group varies a lot with the covariates, from a minimum of 1.1% to a maximum of 62.2%. Moreover, the proportion of students belonging to the *GN* and *NG* groups (i.e. the students able to graduate in only one programme) tends to diminish as the *GG* stratum grows even if the *NN* stratum goes down.

On one extreme, the individual with all the covariates equal to one (a female with residence in Florence coming from a gymnasium with high grade and regular enrolment) has the highest probability of graduation (62.2%) which is entirely attributed to the *GG* group.

On the other extreme, the baseline individual (a male with residence outside Florence coming from a school other than a gymnasium with a low grade and late enrolment) has a low probability of graduation in at least one of the programmes (11.0%), mainly attributed to the *GN* and *NG* groups.

Since the ACE on graduation, $\pi_{GN:i}-\pi_{NG:i}$, stems from the *GN* and *NG* groups, it follows that the two degree programmes have a differential effect on the probability of graduation only for students having a weak background. Special guidance policies should be designed for this kind of students.

In general, knowing the sizes of the principal strata is important for guidance: for example, the students who benefit from enrolling in Economics are the ones belonging to the *GN* stratum, as they graduate only in that degree programme.

Looking at the effect on employment, some results are worth stressing. The level of the probability of employment varies a lot with the covariates, ranging from 47.1% to 77.9% for the graduates in Economics, and from 31.4% to 64.5% for the graduates in Political science. However, the ACE on employment in the *GG* stratum, which we assumed constant on the logit scale to avoid identification problems, generates a quite stable differential of about 15% in the employment probability.

Of course, the reliability and also the substantive importance of such an effect depends on the size of the *GG* stratum: for example, the causal effect in the *GG* stratum has little relevance for our baseline individual, which has a probability of only 1.1% to be a *GG*.

5. Concluding remarks

In the paper, two programmes of the University of Florence have been compared in order to evaluate their effectiveness with respect to employment rate.

The principal strata approach to causal inference was used to set up a general framework for the analysis of the problem, with a precise definition of the quantities of interest. Inference was drawn through a model fitted with maximum likelihood. Some care was needed in the model selection strategy to account for the possibility that some principal strata were empty.

The causal effect in the *GG* stratum (i.e. the students who are able to graduate in both programmes) is positive (namely in favour of Economics graduates). Moreover, the model gives some additional insight into the phenomenon, as it shows how the principal strata structure changes with the covariates: this information is crucial to understand the enrolment process and to interpret consciously the estimated causal effect.

Unfortunately, the limited sample information on the employment status led to a lack of statistical significance in most parameters of the outcome submodel and prevented us from exploring structures that are more complex.

The model could be developed also in a Bayesian framework, which entails several difficulties (specification of the priors, computational complexity), but offers some advantages which become crucial as the complexity of the model increases (Barnard *et al.*, 2003). Alternatively, if the assumptions of the parametric model are judged too restrictive, large-sample non-parametric bounds can be calculated (Grilli & Mealli, 2006).

References

BARNARD J., FRANGAKIS C.E., HILL J.L., RUBIN D.B. (2003) Principal stratification approach to broken randomized experiments: a case study of school choice vouchers in New York City, *Journal of the American Statistical Association,* **98**: 299-323.

FRANKGAKIS C.E., RUBIN D.B. (2002) Principal stratification in causal inference, *Biometrics*, **58**: 21-29.

GRILLI L., MEALLI F. (2006) Nonparametric bounds on the causal effects of University studies on job opportunities using principal stratification, in: *Journal of Educational Behavioral Statistics* (In press, under revision).

IMBENS G.W., RUBIN D.B. (1997) Bayesian inference for causal effects in randomized experiments with noncompliance, *Annals of Statistics*, **25**: 305-327.

JO B. (2002) Estimation of intervention effects with noncompliance: alternative model specifications, *Journal of Educational and Behavioral Statistics*, **27**: 385-409.

MEALLI F., IMBENS G.W., FERRO S., BIGGERI A. (2004). Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes, *Biostatistics*, **5**, 2, 207-222.

RUBIN D.B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, **66**: 668-701.

SAS INSTITUTE (1999) *SAS/STAT User's Guide Version 8*. Cary: SAS Institute Inc.

ZHANG J.L., RUBIN D.B. (2003) Estimation of causal effects via principal stratification when some outcomes are truncated by 'death', *Journal of Educational and Behavioral Statistics*, **28**: 353-368.

Is the Length of the First Job Search Spell a Valid Measure of External Effectiveness of University Programmes?

Dalit Contini

Statistics and Applied Mathematics Dept. "Diego de Castro", University of Turin, Italy

Summary. The aim of this paper is to show that labour-market outcomes related to holding a job are not valid measures of the relative impact of different programmes, while better performing measures are those related to holding jobs reaching given quality standards. For this purpose, we develop a simple job-search model: graduates decide whether to accept a job offer depending on the utility of the different options. If graduates with better work prospects are more selective than others, it can be shown that neither the exit rate from unemployment, nor the probability of being employed at a given time, are necessarily higher for those coming from "the best" university programmes. On the other hand, under mild conditions, neither the waiting time for a job reaching given quality standards nor the probability of having a "good job" depend on individual behaviour, but only on work opportunities: this makes the latter indicators better measures of external effectiveness. Nevertheless, while cross-sectional data suffice for assessment of the working condition, evaluation of the waiting time for a "good job" requires longitudinal data.
Keywords: University educational programmes; External effectiveness; Transition from school to work; Job search models; Hazard function.

1. Introduction

The process of integrating young adults into the labour market is a common problem at the international level (OECD, 1998), as is shown by high youth unemployment rates in many countries. From a comparative standpoint, the assessment of the role of national school systems in the process of transition towards the labour market is one of the most interesting areas of investigation in this particular field (Muller & Shavit, 1998; Van der Velden & Wolbers, 2001; Iannelli, 2001; Brauns *et al.*, 2001). Other studies focus on the effects of various educational levels and programmes in individual countries (Nguyen & Taylor, 2003) or on earning returns to schooling (Angrist & Krueger, 1991; Harmon & Walker, 1995; Colussi, 1997; Checchi, 1997).

From a different perspective, occupational outcomes are considered as indicators of the *external effectiveness* of educational programmes (Gori *et al.*, 1993; Biggeri *et al.*, 2001; Rampichini & Petrucci, 2001; Bratti *et al*, 2004).

One of the aims is to rank[1] professional training courses, schools, universities or specific university study programmes, with respect to their ability to favour the entry of young people into the labour market. Ideally, the purpose should be to assess the net impact of attending a particular course: occupational outcomes are then assessed by controlling for characteristics of the individual and of the context. The main problem to face is the selection process, if the propensity to choose a particular study programme depends on attributes that would make the individuals interesting to potential employers, even in the absence of the study programme itself.

To evaluate the external effectiveness of scholastic or university education in general, various authors (Biggeri *et al.*, 2001; Rampichini & Petrucci, 2001; Nguyen & Taylor, 2003; Porcu & Tedesco, 2004) use as indicator the length of the first job-search spell after attaining a qualification. Others (Giommi & Pratesi, 2001; Bratti *et al.*, 2004) focus instead on the probability of employment at a given time. The use of these indicators is based on the (implicit) assumption that better job prospects correspond to shorter time needed to enter the labour market. As we will see, however, these suppositions may be confuted.

Purpose of this paper is: a) to show the limits of occupational outcomes related to work in itself as indicators of the external effectiveness of study programmes; b) to assess the validity of alternative indicators of external effectiveness, related to performance of a work activity that reaches a given standard of quality.

The idea is simple. Having a job means that:

(i) a job opportunity came up;

(ii) one chose to accept the opportunity.

Hence, the waiting time for first employment depends on the effective willingness of graduates to accept the jobs that are proposed to them. Using a simple job-search model, it can be shown that the waiting time to first job is not necessarily briefer for graduates with better job prospects if they have higher ambitions.

---

[1] The approach, which originated in the growing demand for accountability in public sector activities, is particularly widespread (but also frequently criticized) in Great Britain, where all levels and types of educational institutions are subject to comparative evaluation. The resulting rankings (*league-tables*) of performance indicators are easily accessible to the public (http://education/guardian.co.uk). For a critical analysis of this approach, see Goldstein & Spiegelhalter (1996).

The choice of which indicators to use depends critically on the nature of the information obtained: in particular, whether the data concern employment at the time of the interview, or rather the entire work history observed over a given period.

The present paper is divided as follows: Section 2 presents the job-search model, reproducing a “world” characterized by very simple operative mechanisms. The aim is not to propose an innovative specification of the processes regulating, at the micro level, the match between work demand and supply, but rather to show that the premises underlying the use of the waiting time for first employment as indicator of the external effectiveness of study programmes may not hold. The theoretical results, presented in Sections 3 to 5, can be illustrated by means of a simulation exercise, developed in Section 6. Working histories for graduates of different study programmes are generated under alternative hypotheses on job opportunities and choice criterion. The conclusions follow.

2. The model

The process regulating the search for work involves two categories of actors: the subjects looking for employment and that of potential employers. The latter choose whether, and to whom, they will make work available, while the former chooses whether to accept the proposals.

Our model is based on the idea, borrowed from economics literature, that people’s choices are based on comparing the values of the *utility* function[2] corresponding to the different options. In this paper, utility can assume a very general connotation, diverging from what is hypothesised in job-search models that refer to earnings (Jensen & Westergard-Nielsen, 1987; Eckstein & Wolpin, 1995). Ideally, one should refer to the variety of features that characterise job quality: type of contract, relevance to academic qualifications, need for qualifications, utilization of knowledge and skills acquired, and remuneration.

The model is characterized as follows. Every employment offer is associated with a specific value of the utility function. We assume that the utility ascribed to a job is the same for all graduates, depending only on the characteristics of the job itself. The job offers that each graduate can receive vary, however, among subjects. One and only one job opportunity is allowed for each

[2] The concept of utility is based on the notion that individuals derive satisfaction from consumption of material goods and from leisure time; utility is a measure of this satisfaction (Borjas, 1999). The decision whether to work or not (and how much to work) therefore depends on the hourly wage and on individual preferences (expressed by the indifference curve). The wage below which the individual decides not to work is called the “reservation wage”.

subject at each time t, with utility value described by the random variable $U_L^i(t)$ with distribution g. We formulate the following:

ASSUMPTION 1 $$U_L^i(t) \sim g^i\big(u_L(t)\big) = g(u_L \mid x)\ ,$$

where X is the vector of the individual characteristics which the potential employer is able to observe. X could indicate the graduate's gender, degree of motivation or intelligence, social class, and store of knowledge and skills acquired during university (represented by the disciplinary field and the particular study programme).

The $U_L^i(t)$ are i.i.d. random variables over i and t: quality and quantity of work offers do not change over time, do not depend on previous or future offers[3], and do not depend on offers received by other people.

The assumption that, at each t, one and only one work offer arrives[4] does not rule out that individuals may receive offers with varying frequency. If $P\big(U_L(t)=0\big) > 0$, the absence of offers at time t corresponds to the arrival of an offer with zero utility.

Each individual chooses whether to accept or refuse a work opportunity in relation to the threshold, u_S, the minimum level of utility that he or she is willing to accept. The choice criterion is as follows:

ASSUMPTION 2 *accept the job if* $u_L \geq u_S$;
reject the job if $u_L < u_S$.

We also assume that the threshold does not change over time within the spell[5]. The threshold is described by U_S , which is related to individual characteristics z:

ASSUMPTION 3 $$U_S^i \sim f^i\big(u_S\big) = f(u_S \mid z)\ .$$

In principle, factors X and Z are distinct (Logan, 1996) in that they are determined by different actors (X affects the choices of potential employers, who decide to whom they will propose the job offer, while Z relates to graduates' choices). However, it is plausible that the elements characterizing these factors are mostly the same. If individuals behave in a rational way, those with good

---

3 For the sake of simplicity, we do not allow for dependence on local labour market conditions, nor for structural negative duration dependence in the exit rate from unemployment, which could occur for example because of loss of skills or loosening of social networks (in these cases the i.i.d. assumption would have to be relaxed).

4 Discrete choice (Mc Fadden, 1974) and job-search models typically refer to situations in which individuals choose among a number of work opportunities (as well as the conditions of no work or of waiting for future employment). The model presented here can be related to a similar circumstance, if we hypothesize that $U_L(t)$ represents the utility corresponding to the best offer arriving at time t.

5 The condition is too restrictive if individuals become less selective as time in unemployment grows longer. This could happen because of discouragement.

employment prospects will tend to be more selective, and vice versa[6]. This could be true, for example, for subjects who are especially talented, or who belong to a high social class, or attended a university programme providing knowledge or skills that are in great demand on the market.

The dependence of U_L and U_S on common factors implies that they are generally positively correlated. We assume however that:

ASSUMPTION 4 $\quad U_L(t) \perp U_S \mid X, Z$,

implying that correlation is spurious. The reason is that work offers cannot depend in a causal sense on the individual's threshold, since such threshold is not observable by the potential employers.

Let us assume, in conclusion, that study programmes can be arranged according to the "value added"[7] which they potentially offer to the graduates in terms of marketable skills. If programme *A* provides a higher "value added" than programme *B*, offering better job prospects other things being equal, we say that *A* is preferable to *B*. We assume that, *ceteris paribus*:

ASSUMPTION 5 $\quad G^A(u_L) > G^B(u_L) \qquad \forall\, u_L$

ASSUMPTION 6 $\quad F^A(u_S) > F^B(u_S) \qquad \forall\, u_S$

where $G(u_L) = \mathrm{P}(U_L \geq u_L)$ and $F(u_S) = \mathrm{P}(U_S \geq u_S)$. The "best" programmes offer better employment prospects and lead to more selective behaviours of graduates.

Let *S* represent the body of possible study programmes, with *A* and *B* any two elements of *S*, and *A* preferable to *B* in the sense indicated above. *I* is an indicator of occupational outcome. We say that *I* is a valid indicator of external effectiveness of degree programmes if, *ceteris paribus*, one of the following relations holds:

$$E^A(I) > E^B(I) \qquad \forall A, B \in S \mid A \succ B \ , \tag{1a}$$

$$E^A(I) < E^B(I) \qquad \forall A, B \in S \mid A \succ B \ . \tag{1b}$$

That is, indicator *I* must capture the existing differences among programmes.

---

6 Jensen & Westergard-Nielsen (1987) – in a perfectly rational environment, where the graduate knows the distribution of employment opportunities – derive the optimal reservation wage, which depends explicitly on the distribution of job offers. On the other hand, Eckstein & Wolpin (1995), while observing a positive correlation between estimated reservation wage and average salary of job offers, maintain that such dependence is not theoretically necessary.

7 The term "value added" is frequently employed in the literature on league tables. Notice that we do not address here the problem of how to control over the potential confounding factors due to the selection process. The reasoning is here always "other things being equal".

If I is the length of a spell, for example the duration of first employment search after graduation, then (1) is equivalent to:

$$P^A(T = t \mid T \geq t) > P^B(T = t \mid T \geq t)\ , \tag{2}$$

where T is the duration itself and $P(T = t \mid T \geq t)$ is the discrete-time hazard function. If condition (2) holds, the speed of leaving unemployment is greater for the "better" programme A than for B, and the average duration shorter.

Employing instead a binary indicator of working condition at some time t, say $L(t)$, condition (1) corresponds to:

$$P^A\big(L(t) = 1\big) > P^B\big(L(t) = 1\big), \tag{3}$$

meaning that the employment probability should be higher for A than for B.

3. Waiting time to the first job

If T is the duration of the first employment search spell, the hazard function can be expressed as:

$$P(T = t \mid T \geq t) = P\big(U_L(t) \geq U_S \mid U_L(t\text{-}1) < U_S, \ldots, U_L(1) < U_S\big). \tag{4}$$

Starting from the particular case in which the threshold utility is a constant given Z, we obtain:

$$P(T = t \mid T \geq t) = P\big(U_L(t) \geq u_S \mid U_L(t\text{-}1) < u_S, \ldots, U_L(1) < u_S\big) = P\big(U_L \geq u_S\big) \tag{5}$$

as the utility of job offers at subsequent times are assumed i.i.d.. Note that the exit rate does not change with increasing elapsed time in unemployment.

Let us now take two subjects, identical for all other relevant characteristics, one of whom gained degree A, the other degree B. If A is preferable to B, for assumption 6 we will have $u_S^A \geq u_S^B$. Condition (2) holds if:

$$P(U_L^A \geq u_S^A) > P(U_L^B \geq u_S^B),$$

that is, $G^A(u_S^A) > G^B(u_S^B)$. However, without additional assumptions, it is not possible to derive relation (2) in the general case, as is shown in Figure 1.

Now let U_S be a random variable. From (4) we derive that:

$$P(T = t \mid T \geq t) = \int_{u_S} P\big(U_L(t) \geq U_S \mid U_L(t\text{-}1) < U_S, \ldots, U_L(1) < U_S\big)\, du_S \tag{6}$$

For t=1 we obtain that:

$$\begin{aligned} P(T = 1) = P\big(U_L(1) \geq U_S\big) &= \int_{u_S} P\big(U_L(1) \geq u_S \mid U_S = u_S\big)\ P\big(U_S = u_S\big)\, du_S \\ &\int_{u_S} P\big(U_L \geq u_S\big)\ P\big(U_S = u_S\big)\, du_S \end{aligned} \tag{7}$$

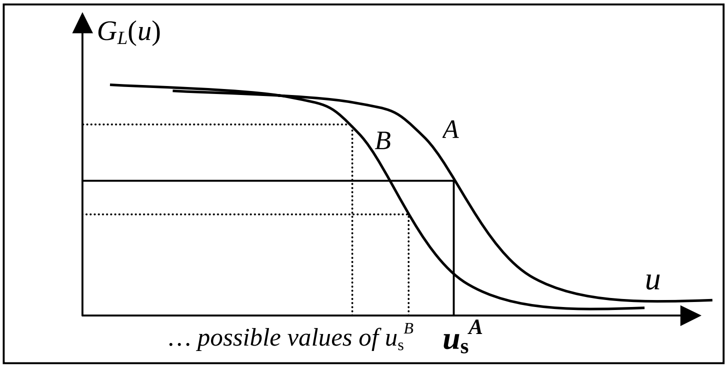


Figure 1. A priori it cannot be assessed whether $P(L(t) = 1)$ is higher for A or B

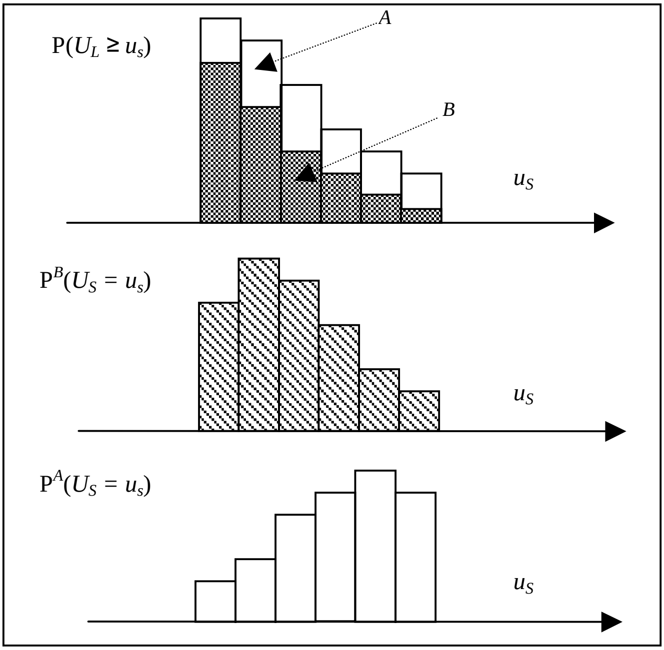


Figure 2. Factors in the integral function (7).

Separately analyzing the two factors inside the integral (see Figure 2) we observe that:

- $P(U_L \geq u_S)$ decreases with u_S. Thus, given u_S, this probability is higher for *A* than for *B* by assumption 5;
- $P(U_S = u_S)$ gives more weight to high values of u_S for *A*, to low values of u_S for *B* by assumption 6.

Without other assumptions, it is not possible to assess *a priori* which of the integrals representing $P^A(T = 1)$ or $P^B(T = 1)$ takes a larger value. Similar results hold for a generic value of t[8].

To conclude, the link between quality of education and time needed to enter the labour market is not clear-cut. The waiting time to first employment is therefore not a valid indicator of external effectiveness, in the sense indicated in Section 2.

4. Probability of employment at time t

The model described in Section 2 specifies how the first work episode after graduation begins, but does not formulate assumptions about its duration or about how the later periods of employment may start. Nevertheless, the employment condition at time *t* also depends on these factors.

Wishing to limit the model's complexity, we formulate the additional hypothesis[9]:

ASSUMPTION 7 *The first work episode has a minimum duration equal to t.*

In this framework, the graduate proves to be employed at time *t* if by that date he or she is offered at least one job with a utility higher than his/her personal threshold of acceptability. Therefore:

$$
\begin{aligned}
P(L(t)=1) &= P\{max[U_L(1),U_L(2)...U_L(t)] \geq U_S\} \\
&= \int_{u_S} P\{max[U_L(1),U_L(2)...U_L(t)] \geq u_S \mid U_S = u_S\} P(U_S = u_S) du_S \\
&= 1 - \int_{u_S} [P(U_L(1) < u_S) P(U_L(2) < u_S)...P(U_L(t) < u_S)] P(U_S = u_S) du_S \\
&= 1 - \int_{u_S} [P(U_L < u_S)]^t P(U_S = u_S) du_S
\end{aligned}
$$

For the same reasons produced with reference to (7), the general validity of relation (3) cannot be demonstrated.

---

8 Knowledge of u_s would help eliminating spurious duration dependence, but would not change the terms of the problem: relation (2) is still not generally true.

9 The hypothesis, which is not very realistic in a world where temporary or short-term jobs are increasingly common, has the effect of overestimating $P(L(t)=1)$.

5. Indicators based on quality work

Measures based on employment rate are not valid indicators of external effectiveness of study programmes because a favourable value of the indicator does not always correspond to study programmes offering better employment prospects. This occurs because subjects behave differently in the labour market.

The role of individual behaviour weakens if we take indicators based on holding a job that reaches a given quality standards. This standard may be defined with reference to the same dimensions used to connote the concept of utility: type of contract, earning, relevance of university education, need for the acquired expertise, and degree to which acquired skills and knowledge are used. A “good” job is thus a job that achieves a predefined minimum level of utility, u_0.

Two assumptions have to be added to those formulated in Section 2:

ASSUMPTION 8 $\qquad u_S^i \leq u_0 \qquad \forall\ i,$

meaning that all graduates are willing to accept jobs that reach the standard, and

ASSUMPTION 9 *when graduate i finds a job, u_S^i, changes and takes the value of the utility of the current job.*

Assumption 9 implies that, if presented with an opportunity with a utility higher than that of the work they are already doing, individuals change job (assuming that there are no job-to-job costs).

In this framework, indicating with T_Q the time before the first quality job, we obtain that:

$$P(T_Q = t \mid T_Q \geq t) = P(U_L(t) \geq u_0 \mid U_L(t-1) < u_0, \ldots U_L(1) < u_0) = P(U_L \geq u_0),$$

because $U_L(t)$ are assumed to be i.i.d.. The speed with which individuals find a “good” job, therefore, varies only in relation to work opportunities, and not to personal choice criterion. Thus, if A is “better” than B, $G^A(u_0) > G^B(u_0)$ by assumption 5, and the waiting time for A will be, on average, shorter.

Let us now consider the probability of performing a quality job at time t. Let $L_Q(t)$ be the binary random variable assuming value 1 in favourable cases. The following should hold:

$$\mathrm{P}(L_Q^A(t)=1) \geq \mathrm{P}(L_Q^A(t)=1) \tag{8}$$

Let us see. A graduate will have a good job at time t if by that date he receives a work offer with a utility higher than u_0. Therefore:

$$P(L_Q(t) = 1) = P\{max[U_L(1), U_L(2) \ldots U_L(t)] \geq u_0\}$$

$$= 1 - \left[P\left(U_L(1) < u_0\right)P\left(U_L(2) < u_0\right)\ldots P\left(U_L(t) < u_0\right)\right]$$
$$= 1 - \left[P\left(U_L < u_0\right)\right]^t = 1 - \left[1 - G\left(u_0\right)\right]^t .$$

Since $G^A(u_0) > G^B(u_0)$ by assumption 5, condition (8) is verified.

Thus, both the length of the search for a quality job and the probability of holding a "good" job at some time t lend themselves well to assessing the external effectiveness of study programmes.

6. Simulation exercise

The theoretical results described in Sections 3-5 can be illustrated with a simulation. Working histories for 1000 individuals have been generated (Table 1) under different distributional assumptions for U_s and $U_L(t)$.

Degree A is always preferable to degree B. In case 1, differences between A and B refer to working opportunities only, while in cases 2-7 they refer to the choice criteria as well.

Waiting time to first job. Average waiting times can be higher for programme A or for programme B (Table 2). A performs much better than B only in case 1, where U_S^A and U_S^B have the same distribution, i.e. the study programme does not affect the choice criterion. In all other cases, only small differences are observed.

Table 1. Distributional assumptions for U_s and $U_L(t)$. Threshold $u_0=3$.

	U_S^A	U_S^B	U_L^A	U_L^B
Case 1	$Uniform(2,3)$	$Uniform(2,3)$	$Uniform(0,4)$	$\begin{cases} 0 & with\ p = 0.25 \\ Uniform(0,3.2) & otherwise \end{cases}$
Case 2	$Uniform(2,3)$	$Uniform(1,2)$	$Uniform(0,4)$	$\begin{cases} 0 & with\ p = 0.25 \\ Uniform(0,3.2) & otherwise \end{cases}$
Case 3	$Uniform(2,3)$	$Uniform(1.5,2.5)$	$Uniform(0,4)$	$\begin{cases} 0 & with\ p = 0.25 \\ Uniform(0,3.2) & otherwise \end{cases}$
Case 4	$Uniform(2,3)$	$Uniform(1,2)$	$Uniform(0,3.5)$	$\begin{cases} 0 & with\ p = 0.25 \\ Uniform(0,3.2) & otherwise \end{cases}$
Case 5	$Uniform(2,3)$	$Uniform(1.5,2.5)$	$Uniform(0,3.5)$	$\begin{cases} 0 & with\ p = 0.25 \\ Uniform(0,3.2) & otherwise \end{cases}$
Case 6	$Uniform(2,3)$	$Uniform(1,2)$	$Uniform(0,4)$	$\begin{cases} 0 & with\ p = 0.25 \\ Uniform(0,4) & otherwise \end{cases}$
Case 7	$Uniform(2,3)$	$Uniform(1,2)$	$Uniform(0,4)$	$\begin{cases} 0 & with\ p = 0.50 \\ Uniform(0,4) & otherwise \end{cases}$

Table 2. Waiting time to the first job. Statistics based on simulated data.

	MEAN		MEDIAN		S.D.	
	A	***B***	***A***	***B***	***A***	***B***
Case 1	2.85	7.62	2	5	2.4	9.1
Case 2	2.83	2.56	2	2	2.5	2.0
Case 3	2.88	3.65	2	3	2.4	3.3
Case 4	3.86	2.51	2	2	4.1	2.1
Case 5	4.13	3.84	3	3	4.7	3.6
Case 6	2.71	2.17	2	2	2.2	1.7
Case 7	2.95	3.18	2	2	2.6	2.4

Table 3. Waiting time to the first "good" job. Statistics based on simulated data.

	MEAN		MEDIAN		S.D.	
	A	***B***	***A***	***B***	***A***	***B***
Cases 1-3	4.0	21.1	3	15	3.5	20.0
Cases 4-5	6.8	22.2	5	17	6.5	20.6
Case 6	4.1	5.4	3	4	3.6	4.9
Case 7	3.8	7.7	3	6	3.3	7.6

Table 4. Probability of being employed six time units after graduation. Statistics based on simulated data.

	% EMPLOYED AFTER SIX TIME UNITS	
	A	***B***
Case 1	92.6	60.6
Case 2	93.4	93.8
Case 3	94.6	84.5
Case 4	83.4	94.2
Case 5	83.7	85.2
Case 6	92.8	97.8
Case 7	92.4	89.5

Waiting time to first "good" job. Differences between programmes are much more marked here (Table 3). In all cases, average spells are much shorter for *A* than for *B*.

Probability of being employed at time t. The percentage of individuals holding a job six time units after graduation is not uniformly higher for *A* than for *B*, as we would expect if the indicator was a valid measure of effectiveness (Table 4). In this case, as occurs for the waiting time to the first job, the difference is markedly more favourable for *A* in case 1 only.

Table 5. Probability of having a "good" job six time units after graduation. Statistics based on simulated data.

	% HOLDING A "GOOD" JOB AFTER SIX TIME UNITS	
	A	***B***
Cases 1-3	81.3	22.2
Cases 4-5	59.6	23.6
Cases 6	81.7	67.7
Case 7	80.9	55.8

Probability of having a "good" job at time t. The percentage of graduates holding a job which reaches given standards of quality at time t=6 is everywhere much higher for those with degree *A* (Table 5).

7. Conclusions

We have argued that indicators referring to jobs reaching a given standard of quality are the most appropriate for the evaluation of external effectiveness, nevertheless we have to acknowledge that data required to construct these indicators are much more demanding than those related to the holding of "any job" (Table 6).

Table 6. Data requirements for the evaluation of external effectiveness

INDICATORS	DATA REQUIREMENTS
Probability of being employed at time *t*.	Occupational condition at time of interview.
Waiting time to first job.	(Beginning time of job-search). Beginning time of first job.
Probability of having a "good" job at time *t*.	Occupational condition at time of interview. Characteristics of current job .
Waiting time to first "good" job.	Beginning and end time of all jobs. Characteristics of all jobs.

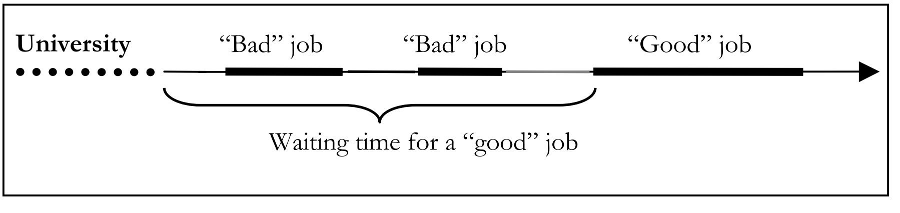


Figure 3. The assessment of time to a "good" job requires event history data.

Waiting time to the first “good” job has the highest data requirements: *event history* data on working careers are needed (Figure 3). Given the deep changes that have occurred in the labour market in recent years, this kind of information is particularly relevant, as lifelong jobs are now much less widespread among young people.

The nationwide survey on Italian graduates carried out by ISTAT (2004) does not collect event history data. Thus, the assessment of waiting time to the first “good” job is not feasible now. Nevertheless, these surveys do collect detailed information on the current job, thus the quality of the current job can be evaluated.

In conclusion, given data limitations, it seems to be a much more sensible practice to evaluate the external effectiveness of university programmes by referring to the “probability of having a job reaching a given quality standard at time *t*” rather than employing the “waiting time to the first job”.

References

ANGRIST J., KRUEGER A. (1991) Does compulsory attendance affect schooling and earnings? *Quarterly Journal of Economics*, **106(4)**: 979-1014.

BIGGERI L., BINI M., GRILLI L. (2001) The transition from university to work: a multilevel approach to the analysis of the time to obtain the first job, *J.R.S.S.* A, **164(2)**: 293-305.

BORJAS G.J. (1999) *Labor Economics*, McGraw Hill, New York.

BRATTI M., MC KNIGHT A., NAYLOR R., SMITH J. (2004) Higher education outcomes, graduate employment and university performance indicators, *J.R.S.S.* A, **167(3)**: 475-496.

BRAUNS H., GANGL M., SHERER S. (2001) Education and unemployment: patterns of labour market entry in France, the United Kingdom and West Germany. TSER Project: *Comparative Analysis of Transitions from Education to Work in Europe* (http://www.mzes.uni-mannheim.de/projekte/catewe).

CHECCHI D. (1997) *La diseguaglianza. Istruzione e mercato del lavoro*, Laterza, Bari.

COLUSSI A. (1997) Il tasso di rendimento dell’istruzione in Italia. In: ROSSI N. (ed) *L’istruzione in Italia: solo un pezzo di carta?* Il Mulino, Bologna: 255-275.

ECKSTEIN Z., WOLPIN K. I. (1995) Duration to first job and the return to schooling: estimates from a search matching model, *Review of Economic Studies*, **62(2)**: 263-286.

GIOMMI A., PRATESI M. (2001) Speranza differenziale di occupazione dei laureati e dei diplomati di scuola superiore a Firenze. In: FABBRIS L. (ed) *Captor 2000: Qualità della didattica e sistemi computer-assisted*, CLEUP, Padova: 273-284.

GOLDSTEIN H., SPIEGELHALTER D.J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance, *J.R.S.S.* A, **159(3)**: 385-443.

GORI E., MEALLI F., RAMPICHINI C. (1993) Indicatori di efficienza ed efficacia per la valutazione dell’attività di formazione professionale, *Statistica*, **3**: 501-533.

HARMON C., WALKER I. (1995) Estimates of the economic returns to schooling, *American Economic Review*, **85(5)**: 1278-1286.

IANNELLI C. (2001) School effects on youth transitions in Ireland, Scotland and the Netherlands. TSER Project: *Comparative Analysis of Transitions from Education to Work in Europe* (http://www.mzes.uni-mannheim.de/projekte/catewe).

ISTAT (2004) *I laureati e lo studio. Inserimento professionale dei laureati. Indagine 2001*, Istat, Roma.

JENSEN P., WESTERGARD-NIELSEN N. C. (1987) A search model applied to the transition from education to work, *Review of Economic Studies*, LIV(179): 461-472.

LOGAN J. A. (1996) Opportunity and choice in socially structured labor markets, *American Journal of Sociology*, **102(1)**: 114-160.

MCFADDEN D. (1974) The measurement of urban travel demand, *Journal of Public Economics*, **3**: 303-328.

MULLER W., SHAVIT Y. (1998) The institutional embeddedness of the stratification process. In: SHAVIT Y., MULLER W. (eds) *From School to Work: A Comparative Study of Educational Qualifications and Occupational Destinations*, Clarendon Press, Oxford.

NGUYEN A.N., TAYLOR J. (2003) Transition from school to first job: the influence of educational attainment, *Lancaster University Management School, Working Paper* 2003/009.

OECD (1998) *Employment Outlook*, OECD, Paris.

PORCU M., TEDESCO N. (2004) Dall'Università al lavoro: analisi dei tempi di passaggio dei laureati dell'Ateneo di Cagliari. In: AURELI-CUTILLO E. (ed) *Strategie metodologiche per lo studio della transizione Università-lavoro*, CLEUP, Padova: 281-296.

RAMPICHINI C., PETRUCCI A. (2001) La ricerca della prima occupazione: un modello di durata per i laureati dell'Ateneo fiorentino. In: FABBRIS L. (ed) *Captor 2000: Qualità della didattica e sistemi computer-assisted*, CLEUP, Padova: 257-272.

VAN DER VELDEN R.K.W., WOLBERS M.H.J. (2001) The integration of young people in the labour market within the European Union: the role of institutional settings. TSER Project: *Comparative Analysis of Transitions from Education to Work in Europe* (http://www.mzes.uni-mannheim.de/projekte/catewe).

Competence-based Compatibility between Jobs and University Curricula

Luigi Fabbris[1]

Statistics Department, Padua University, Italy

Summary. In this paper, we put forward a mathematical and statistical model for measuring the "compatibility" between jobs and university curricula based on professional competences. Our model aims at comparing work requirements and the professional counterparts achievable at school. Even if the basic concepts might be valid for the analysis of outer educational realms, the model is devoted to the analysis of highly qualified jobs, and in specific those "from technical to managerial positions" that may be taught at university. Our model is general enough, too, to frame various studies on the job market and clarify the terminological, conceptual and operational entities of occupations.
Keywords: Competences; Labour market; Jobs; Professional profile; University curricula; Formative profiles; Degree Programmes; Compatibility analysis.

1. Work activities, jobs and university curricula

The professional competences required in global-market activities change rapidly. The changes concern the technical refinement of occupations, i.e. the ability of workers in relating with customers, dealers and colleagues, communicating internally and externally with their company, working in a team, solving problems and organising their own job and that of others.

The term *competences* encompasses the knowledge, skills and attitudes required to applicants for work. This topic is dealt with in Section 2. In the following, without loss of generality, we assume that school is where the competences, as part of a person's personality causally related to superior job performances (Spencer & Spencer, 1993), are grounded.

---

[1] The author wishes to thank Professors Simona Balbi, Marisa Civardi and Corrado Crocetta for invaluable suggestions on a first draft of this paper.

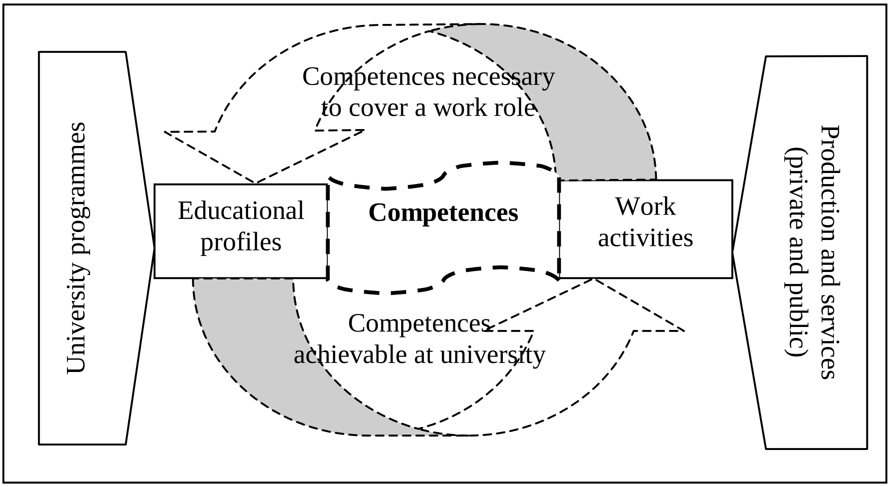


Figure 1. Type of relationships hypothesised between work activities and educational profiles in a “competence-based” society

Our relational model between work activities and competences stems from the hypothesis that the (bi-directional) links between the sites of production-reproduction of knowledge, whose edge is university, and those of economic production, represented by private companies, public bodies and professional offices[2], are knowledge and skills an educated person has to possess to cover a work role. The basic hypothesis of these links, represented in Figure 1, is a “competence-based” social system.

We can describe the links between job demand and offer in mathematical terms and quantify it by focusing upon the pertinent social segment. In particular, each work activity requires competences peculiar to the size and activity of the concerned company and its economic sector. The possibility to quantify the relationships represented in Figure 1 could allow us to predict and harmonise the needs and actions of labour and education universes.

From now on, we will use the terms “educational profile”, or “formative profile”, to indicate the competences achievable at a study programme. A formative profile is a direct output of a university programme, even if a study programme may generate more than one profile and, on the contrary, different programmes may generate the same essential profile. Sometimes, we will name “curriculum” the formative path a student can go through during his/her university studies.

The terms “professional tasks” and “activities” denote the work roles in companies. The terms “job”, “occupation” and “professional profile” are the

[2] From now on, dealing with competences, we will not distinct between the private and public sectors of economy.

work roles a person, through his/her interrelated competences and culture, can realise.

In the following, we discuss:

- *a relational model between work activity and professional competences* (Section 2); the model is specified in statistical and mathematical terms to make the relationships between entities clear,
- *criteria for identifying the current and possible work activities in the economic sectors we are concerned with, and the competences associated to the identified activities.* The criteria are extrapolated from the current literature (Section 3) and direct data collection (Section 4).

In Section 5 we summarise our analyses and suggest issues for further research.

2. A relational model between work roles and competences

Professional competences are composed of knowledge, skills and attitudes appropriate for working in a productive environment. Knowledge is the set of principles and information usable at work. Skills are the sound aptitudes that enable a person to realise a work activity and may be either job-specific or cross-occupational (Befani, 2004; Silvestri *et al.*, 2005). Attitudes relate to workers' positive relationships with the productive structure, the people working with them, and their own professional tasks.

Competences are the results of a complex personal elaboration of the educational and experiential inputs (Keagan, 2002). This process implies a continuous matching with individual values, beliefs and aims. That is why we can define the competences the professional personality, or the professional style of a person.

Cross-occupational skills and personal attitudes may give a worker five types of know-how: *(i)* facing the complexity and picking up from a situation what is new; *(ii)* focusing on situations and behaving consistently; *(iii)* selecting, with reference to values, the most correct, appropriate, desirable behaviour; *(iv)* perceiving other people's competences; *(v)* abstracting over the contingent situation and imagining alternative scenarios. This know-how has no relationship with job-specific knowledge.

The job-specific competences differ according to job. They are composed of knowledge and skills specific of a certain job. Some technical competences are basic, and, at the hiring stage, it is taken for granted that graduates who apply for certain jobs possess these general competences. For technical and managerial positions cross-occupation competences are the fluency in English, spoken and written, and the ability to use the main computer functions and to correctly use the national language (for details, see Section 4). The 'a priori' relevance attributed to these skills does not imply all graduates at work (Fabbris & Visentin, 2005) use them.

From a mathematical viewpoint, the relationship between work activities and competences may be expressed with the following data matrices:

- competence-by-job matrix (Section 2.1),
- competence-by-educational profile matrix (Section 2.2),
- job-by economic sector matrix (Section 2.3),
- curriculum-by-university matrix (Section 2.4),
- job-by-educational profile compatibility matrix (Section 2.5).

2.1 The competence-by-job matrix

A matrix of competence-by-job is a rectangular ($A \times K$) matrix, where A is the number of jobs identified within S economic sectors we are concerned with and K is the number of job-related competences:

$$_s\mathbf{Z} = \left\{ _s z_{ak} \right\} \quad (s = 1, ..., S;\ a = 1, ..., A;\ k = 1, ..., K),$$

where the element $_s z_{akv}$ is the amount of competences of k-th type necessary to realise a-th activity.

The expanded form of matrix $_s\mathbf{Z}$ is represented in Figure 2. Each row-vector represents the variety of competences used for a-th activity and each column the jobs for which k-th competence may be used. The relevance of a competence to an occupation may be estimated with Delphi-like methods (Ford, 1975).

The mean of the elements of a column vector

$$_s\bar{z}_k = \sum_a^A {_s z_{ak}} / A = \mathbf{1'}\, {_s\mathbf{Z}} / A \quad (s = 1, ..., S;\ k = 1, ..., K),$$

where **1'** is a vector of A ones, estimate the average level of k-th competence utilisation in occupations of sector s.

$$_s\mathbf{Z} = \begin{array}{c} \\ 1 \\ \vdots \\ a \\ \vdots \\ A \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & \ldots\ldots & k & \ldots\ldots & K \end{array} \\ \boxed{\begin{array}{ccccc} _s z_{11} & \ldots\ldots & _s z_{1k} & \ldots\ldots & _s z_{1K} \\ \vdots & & \vdots & & \vdots \\ _s z_{a1} & \ldots\ldots & _s z_{ak} & \ldots\ldots & _s z_{aK} \\ \vdots & & \vdots & & \vdots \\ _s z_{A1} & \ldots\ldots & _s z_{Ak} & \ldots\ldots & _s z_{AK} \end{array}} \\ \begin{array}{ccccc} & & _s z_{+k} & & \end{array} \end{array} \begin{array}{c} \\ \\ \\ _s z_{a+} \\ \\ \\ \\ \end{array}$$

Figure 2. Matrix of competences needed to realise work activities in the economic sector s (s=1,…, S)

The sum by row, weighted to account for the possible non-independence of competences relevant to a-th job (see Section 2.5), represent the average use of competences to realise that job:

$$ {}_s z_{a+} = {}_s\mathbf{Z}\,\mathbf{w}_a = \sum_k^K {}_s z_{ak}\, w_{ak} \qquad (s = 1, ..., S;\; a = 1, ..., A), $$

where $\boldsymbol{w}_a$ is a vector of K weights that vary between 0 and 1 and sum 1.

2.2 The competence-by-educational profile matrix

The competences associated to C University curricula are ordered in a (C×K) matrix

$$ {}_u\mathbf{Z} = \left\{ {}_u z_{ck} \right\} \qquad (u = 1, ..., U;\; c = 1, ..., C;\; k = 1, ..., K), $$

whose generic element ${}_u z_{ck}$ represents the amount of competences associated to k-th profile achievable by attending programmes at university u (u=1, …, U).

The expanded form of matrix ${}_u\mathbf{Z}$ is represented in Figure 3. A row-vector represents the competences associated to c-th curriculum (c=1, …, C) and a column-vector the set of programmes where k-th competence is modelled. The column mean

$$ {}_u\bar{z}_k = \sum_c^C {}_u z_{ck} / C = \mathbf{1}'\,{}_u\mathbf{Z} / C \qquad (u = 1, ..., U;\; k = 1, ..., K), $$

where **1’** is of order C, is the average level of k-th competence achievable at u-th university programmes.

Analogously, the weighted mean by row is an estimate of the level of competences achievable by attending programme c (c=1, …, C):

$$ {}_u z_{c+} = {}_u\mathbf{Z}\,\mathbf{w}_c = \sum_k^K {}_u z_{ck}\, w_{ck} \qquad (u = 1, ..., U;\; c = 1, ..., C), $$

where $\boldsymbol{w}_c$ is a column vector of K non-negative weights whose sum is one.

$$ {}_u\mathbf{Z} = \begin{array}{c|ccccc|c} & 1 & \ldots\ldots & k & \ldots\ldots & K & \\ \hline 1 & {}_u z_{11} & \ldots\ldots & {}_u z_{1k} & \ldots\ldots & {}_u z_{1K} & \\ \vdots & \vdots & & \vdots & & \vdots & \\ c & {}_u z_{c1} & \ldots\ldots & {}_u z_{ck} & \ldots\ldots & {}_u z_{cK} & {}_u z_{c+} \\ \vdots & \vdots & & \vdots & & \vdots & \\ C & {}_u z_{C1} & \ldots\ldots & {}_u z_{Ck} & \ldots\ldots & {}_u z_{CK} & \\ \hline & & & {}_u z_{+k} & & & \end{array} $$

Figure 3. Matrix of competences achievable at study programme u (u=1, .., U)

2.3 The job-by-economic sector matrix

The job by economic sector matrix is of order (S×A^*), where S is the number of sectors and A^* that of occupations of s-th sector

$$\mathbf{X} = \{x_{s\,a}\} \qquad (s = 1, \ldots, S;\ a = 1, \ldots, A)$$

where the generic element x_{as} is a 0-1 indicator of the absence-presence of a-th occupation in s-th sector. Zero denotes the absence of a-th occupation, one its presence[3]. The expanded form of matrix **X** is presented in Figure 4. The ones of a row of the matrix represent the work activities of s-th sector (s=1, …, S) and the ones of a column the sectors where a-th job is required.

$$\mathbf{X} = \begin{array}{c|ccccc|c} & 1 & \ldots\ldots & a & \ldots\ldots & A^* & \\ \hline 1 & x_{11} & \ldots\ldots & X_{1a} & \ldots\ldots & x_{1A^*} & \\ : & : & & : & & : & \\ s & x_{s1} & \ldots\ldots & x_{sa} & \ldots\ldots & x_{sA^*} & x_{s+} \\ : & : & & : & & : & \\ S & x_{S1} & \ldots\ldots & x_{Sa} & \ldots\ldots & x_{SA^*} & \\ \hline & & & X_{+a} & & & \end{array}$$

Figure 4. Matrix of jobs by economic sector

2.4 The curriculum-by-university matrix

The matrix of curriculum-by-university is of (U×C) order, where U is the number of universities and C that of different curricula:

$$\mathbf{Y} = \{y_{uj\,c}\} \qquad (u = 1, \ldots, U;\ c = 1, \ldots, C)$$

where y_{uc} is the 0-1 indicator of the c-th formative profile stemming from the homonymous programme at u-th university.

The expanded form is presented in Figure 5.

$$\mathbf{Y} = \begin{array}{c|ccccc|c} & 1 & \ldots\ldots & C & \ldots\ldots & C & \\ \hline 1 & y_{11} & \ldots\ldots & y_{1c} & \ldots\ldots & y_{1C} & \\ : & : & & : & & : & \\ u & y_{u1} & \ldots\ldots & y_{uc} & \ldots\ldots & y_{uC} & y_{u+} \\ : & : & & : & & : & \\ U & y_{U1} & \ldots\ldots & y_{Uc} & \ldots\ldots & y_{UC} & \\ \hline & & & y_{+c} & & & \end{array}$$

Figure 5. Matrix of curricula by university

3 Matrix **X** could be a *relevance* matrix of jobs to sectors, too, and its elements assume values between zero and one. In the following, we will assume, without loss in generality, that matrix elements are zeroes and ones.

2.5 The compatibility matrix

The relationship between educational profiles and occupations may be organised in a ($A \times C$) matrix of “compatibilities”:

$$\mathbf{P} = \{p_{ac}\} \quad (a = 1, ..., A; c = 1, ..., C),$$

whose element p_{ac} is the amount of competences shared by c-th educational profile and a-th occupation. The values of matrix **P** vary between zero and one, zero indicating the absolute independence of occupation and educational profile and one being its maximum. The expanded form is presented in Figure 6.

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ : \\ a \\ : \\ A \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & \ldots\ldots & c & \ldots\ldots & C \end{array} \\ \boxed{\begin{array}{ccccc} p_{11} & \ldots\ldots & p_{1c} & \ldots\ldots & p_{1C} \\ : & & : & & : \\ p_{a1} & \ldots\ldots & p_{ac} & \ldots\ldots & p_{aC} \\ : & & : & & : \\ p_{A1} & \ldots\ldots & p_{Ac} & \ldots\ldots & p_{AC} \end{array}} \end{array}$$

Figure 6. Matrix of compatibilities between formative profiles and occupations

The compatibility is an appropriate function of $_u\mathbf{Z}$ and $_s\mathbf{Z}$. If we assume competences are independent to each other, in both matrices, compatibility p_{ac} between a-th activity and c-th curriculum may be quantified as follows:

$$p_{ac}^{I} = \sum_{k}^{K} \min({}_u z_{ck};\, {}_s z_{ak})\, w_k \quad (a = 1, ..., A;\ c = 1, ..., C),$$

where w_k is a non negative weight associated to k-th competence ($\Sigma_k\, w_k$=1). Such compatibility varies between 0, if no overlap exists between a-th activity and c-th curriculum, and 1, if the use of competence is systematic and continuous.

If we assume that competences correlate either with an activity, or with an educational profile, we could transform columns of matrices $_u\mathbf{Z}$ and $_s\mathbf{Z}$ into their independent linear components. A possibility is to define a ($AC \times K$) matrix **X** whose rows are the logical product of rows of the two **Z** matrices and columns are the competences. The generic element of matrix **X** may be estimated as the minimum value of k-th competence proper to both educational and professional profiles:

$$x_{ik} = \min\left(z_{ak};\, z_{ck}\right) \quad (i = 1, ..., AC; k = 1, ..., K).$$

Another measure, p_{ac}^{II}, of compatibility between c-th educational and a.-th professional profiles may be the sum of the r factorial scores f_{ij} weighted with their eigenvalues:

$$p_{ac}^{II} = \sum_{j}^{r} f_{ij} \sqrt{\lambda_j} \quad (i = 1, ..., ac, ..., AC),$$

r being the *rank(**X**)* and λ_j the generic eigenvalue of the variance-covariance **S** matrix whose columns sum zero:

$$\mathbf{S} = \mathbf{X'X} / AC = \left\{ s_{ij} = \sum_{i}^{AC} x_{ik}\, x_{ih} / CA \right\}.$$

3. Compatibility estimation

To define the relationships between work activities and educational profiles it is necessary:

- *to represent the functional structure of companies of each economic sector*. The identification of functional areas that discriminate some types of companies, and in particular companies by size, will give us hints for understanding the specificity of competences of a given occupation;
- *to identify work activities of all functional areas of companies of a given economic sector*. We are concerned with managerial, technical and other white-collar activities, to which general clerical positions are added[4]. These work positions are those that a graduate may aspire to and, actually, are covered by graduates. To define work activities, entrepreneurs, managers and other top people in companies, and entrepreneurial associations are to be interviewed. Operatively, it is necessary to define a dictionary of activities of the examined economic sectors. We can start from the last Ateco classification (ISTAT, 2002) and add the new findings in classification research, such as the OECD (2001) classification developed after PISA project, CITP-88 international classification of jobs by ILO-BIT, ISTAT's in progress classification of jobs, ROME (1982, 1999) system developed by ANPE - *Agence Nationale française Pour l'Emploi*, the Italian Ministry of Labour-ISFOL series of publications on jobs typical of an economic sector or a functional area within companies (ISFOL, 1999; www.isfol.it/orientaonline), the repertory of jobs of Padua University (Fabbris, 2000, 2005) and other repertories fostered for guidance, statistical, or simply rational, purposes;

[4] Despite clerk, as a professional position, is generic in principle, it is often required in practice that clerical workers possess qualified competences. This is why clerk is considered as a position relevant even for graduates. Several graduates accept clerical positions while looking for first employment (Cappa & Fabbris, 2004). Blue collar positions are, instead, absent on graduates' expectations, even if they should be considered merely for first employment.

- *to identify professional competences inherent to work activities*. Competences can be ascertained according to research experience or expert evaluation. Type and intensity of competences may be correlated to structural features (economic sector, size) of companies and functional areas within them. An important issue is to realise if, in a given work environment, competences are specialised forms of a broader category, or are simply different. For this, we need a 'dictionary of competences' of each examined economic sector. The basic grid edited by the University of Padua for its repertory of jobs, the Spencer & Spencer (1993) list will be the bases to be enriched with outer experiences;
- *to define target competences of each university curriculum*. The terminology of competences must be that in use at work. We will try to harmonise the possibly different terminology of competences as defined at the labour and educational levels (see González & Wagenaar, 2005). Either competences may be achieved from the stated curriculum, or interviewing graduates about their work activities. Competences and jobs associated with a university curriculum may depend on the university that offers it, i.e. programme contents and teaching methodology of different universities may differ substantially. Nevertheless, for the sake of generality, programmes with the same name will be considered equivalent;
- *to define a one-to-one correspondence between competences qualifying work activities and university programmes*. To define the correspondence, a common terminology is to be used. We should use the terminology used in the job market;
- *to measure compatibility between work activities and programme aims*. To be able and associate a number to compatibility, we assume it is possible to quantify both the relationships between the activity and competences, and that between the curriculum aims and competences. A one-to-one correspondence between a formative profile and a job exists just for the regulated professions and some other specialised activities. Either the correspondence between non-regulated professions becomes to blur as the company size lowers[5] or its activities are new and unconventional. In micro and small companies, graduates are hired to realise work activities, some of which are consequent to his/her degree, while others may be more or less qualified.

[5] Del Favero & Fabbris (2002) found that 36% of Padua graduates work in micro and small companies (up to 19 employees). This implies that several graduates may cover rather heterogeneous professional roles.

4. Criteria for data collection on occupations

We assume that the employers, managers and experts in charge of personnel hiring within companies possess realistic information on work roles. They are informed about the company's professional needs regarding both the number of people recently hired and the company's future needs of jobs and competence.

The data may be collected face-to-face with a partially structured questionnaire administered by skilled interviewers. Normally, the data collection is realised on the company's premises.

A CATI – *Computer Assisted Telephone Interviewing* may be realised if either the companies of the analysed economic sector are heterogeneous, or a larger sample is required, or a face-to-face survey is not practicable. A CATI survey is less expensive and organisationally easier than a face-to-face interview.

The collection of data on professional roles of handicrafts and regulated professions may be realised with focus groups, where a limited number of people (8 to 12) are interviewed together by the researcher him/herself. Focus group interaction makes it easier to define roles and competences than independent interviewing.

A questionnaire should be designed to collect the following areas of content:

- Size, age, legal, organisational and network structures, and outsourcing activities of the company.
- Indicators of company development related to certification, information system, plants renewal, e-business, and so on.
- Functional organisation of company (areas with a person in charge of that function)
- Number of employee by professional position and company's functional areas. The seasonal and foreign people hiring, the frequency of parental aid and the annual frequency of internships.
- Turnover and new occupation forecast two-three[6] years apart.
- Definition of technical, intermediate and managerial positions existing within the company, or to be covered with future recruitment, by functional areas.
- Basic, cross-occupational and job-specific competences of professional roles, possibly by functional areas.

Basic competences may be surveyed with reference to the following requirements:

- the level of spoken and written English. It is convenient to inquire about the knowledge of other foreign languages, even if the other than English knowledge is to be considered a job-specific skill;

[6] It could be possible to enquire the entrepreneurs/managers and foresee their needs five years onward. Several research experiences indicate that this is really a problematic task.

- the level of computer skill. The question should be asked in a non trivial way, given that no technician can work without knowing Office package, nor can managers manifest the snobbish attitude not to use PCs. In general, we can distinguish between the abilities of using highly specific packages and that of producing computer programmes.

Cross-occupation competences are peculiar from the data collection viewpoint, since they are not taught as such at school despite their professional relevance. These competences – to be collected in relation with work effectiveness[7] and not personal effectiveness – are the capacities:

- to work in team,
- to communicate with customers and tradesmen,
- to organise his/her own and other people's work,
- to conceive and manage research and work projects,
- to write and present reports and management indicators.

The job-specific competences differ according to job. Hence, it should be decided beforehand if the survey should care about it. In general, entrepreneurs and managers leave them implicit into the occupation labels perceived as suitable for a work position. Besides, there are questions about competences used on (almost) a daily basis and others on peculiar and very qualifying professional activities. Whenever the information about job-specific competences is collected, it may be processed by means of text analysis and enters the dictionary of competences.

The attitudes that deserve to be collected are the professionally distinctive ones. We can ask about the availability to work in uncomfortable conditions, such as, working in open air, underground or in closed environment, travelling frequently or for long periods. With reference to a job, these attitudes could be thought as technical requirements.

5. Conclusive considerations

Our model aimed to quantify the compatibilities between educational and professional profiles. Compatibilities are a paradigm to qualify curricula and jobs in terms of competences, too.

Our model takes it for granted that university programmes can be compared with work needs. We doubt work needs are a parameter for quality of competences acquired at university, because otherwise cultural, non-professional, competences risk to be considered second category. We aim to define how it is

[7] Competences oriented to personal effectiveness are: the possession of formal languages, mental flexibility, leadership, self consciousness, decisional character, ability to face the unexpected, analytic and synthetic abilities, risk propensity, autonomy of judgement, emotional self-control, professional engagement, faithfulness, creativity and other competences ideal for work and life.

possible to assess relationships between the two kinds of profile whenever the decision to start this link would be taken.

The critical point of our model is the possibility to quantify the competences required for working purposes. Even the possible quantification of relationships between educational and professional profiles would be meaningless if not framed within a design concerned with the estimate of short-mean term work possibilities for the considered economic sectors.

The estimate of work chances are an aim of Excelsior informative system (www.excelsior.unioncamere.it) and Istat's surveys (Decennial population and economic activities censuses, quarterly labour force survey, European panel survey on families).

References

BEFANI B. (2004) Il dibattito sulle competenze dalla teoria alla base di dati O*Net. In: CIVARDI M. (ed) *Transizione Università-Lavoro: la definizione delle competenze*, CLEUP, Padova: 19-43.

CAPPA V., FABBRIS L. (2004) L'ambiente di lavoro come metro di qualità di una professione. In: FABBRIS L. (ed) *Capitalizzazione delle competenze dei laureati dell'Università di Padova*, Quaderno PHAROS n. 9/2004, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova: 27-42.

DEL FAVERO M., FABBRIS L. (2002) *Rappresentazione dei mercati locali del lavoro: il sistema di indicatori DOMUS*, Quaderno PHAROS n. 1/2002, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova.

FABBRIS L. (a cura di) (2000, 2005) *Il repertorio delle professioni della Facoltà di Scienze Statistiche*, CLEUP, Padova.

FABBRIS L., VISENTIN E. (2005) Tessere di competenze per mosaici di attività in aziende di varia dimensione. In: D'OVIDIO F. (ed) *Professioni e competenze nel lavoro dei laureati*, CLEUP, Padova: 21-44.

FORD D. (1975) Shang Inquiry as an Alternative to Delphi: Some Experimental Findings, *Technological Forecasting and Social Change*, **7(2)**: 139-164.

KEGAN R. (2002) Competencies as working epistemologies: Ways we want adults to know (*DeSeCo Symposium, Geneva, 11-13 February 2002*, mimeo).

GONZÁLEZ J., WAGENAAR R. (eds) (2005) *Tuning Educational Structures in Europe II*, University of Deusto, University of Groningen, Bilbao.

ISFOL (1999) *Repertorio delle professioni. Modello e metodologia*, ISFOL, Commissione Europea DG-V, Ministero del Lavoro e della Previdenza sociale – Ufficio centrale OFPL, FSE, Roma.

ISTAT (2002) *Classificazione delle attività economiche – Ateco 2002* (www.istat.it/strumenti/definizioni/ateco/).

LEVY F., MURNANE R.J. (1999) Are there key competencies critical to economic success? An economic perspective (*Definition and Selection of Competencies: Theoretical and Conceptual Foundations –DeSeCo*, mimeo).

McGAW B. (2002) From pragmatics to a framework (*DeSeCo Symposium, Geneva, 11-13 February 2002*, mimeo).

OECD (2001) *Knowledge and Skills for Life: First Results from PISA 2000*, OECD, Paris.

ROME (1982, 1999) *Répertoire opérationnel des Arts et Métiers*, ANPE, France.

SILVESTRI I., ROTA G., BOARETTO A., FABBRIS L. (2005) Trasversalità e specificità delle competenze dei laureati dell'Università di Padova. In: FABBRIS L. (ed) *Efficacia esterna della formazione universitaria: il progetto Outcomes*, CLEUP, Padova: 203-222.

SPENCER L.M., SPENCER M.S. (1993) *Competence at Work. Model for Superior Performance*, Wiley, New York.

UNIONCAMERE – MINISTERO DEL LAVORO E DELLE POLITICHE SOCIALI (2004) *Sistema informativo Excelsior. Sintesi dei principali risultati – 2004*, Unioncamere, Roma.

The Ideal Candidate.
Analysis of Professional Competences through Text Mining of Job Offers

Emilio Di Meglio, Maria Gabriella Grassia, Michelangelo Misuraca[1]

Department of Mathematics and Statistics, University of Naples "Federico II", Italy

Summary. The aim of this paper is to propose analytical tools for identifying peculiar aspects of the job market for graduates. The main objective is to reduce the complexity of the phenomenon, both on the variable side, by transforming the collected information into latent factors, and on the unit side, by classifying observations. We propose a strategy for dealing with data that have different source and nature. The dependence structure is investigated to identify potential evolutionary paths. Moreover, symbolic objects and their graphical representation are used for identifying the peculiar characteristics required by companies operating in different economic sectors.
Keywords: Text mining; Association rules; Factor analysis; Symbolic objects; Zoom-star.

1. Text mining

The huge quantity of *on-line* documents and companies' data warehouses makes necessary tools and methods for their analysis (Manning & Schütze, 2001). *Text mining* is a way to unveil the information within verbatim documents, i.e. written in natural language, by using statistical, linguistic and information technology methods. The applications of text mining are increasing at a fast rate: search engines, email filtering and automatic delivery, market analysis based on reports and papers, automatic document classification according to different queries and criteria, and case-based reasoning.

---

[1] This paper is the result of the joint research of the three authors. However, M. Misuraca was responsible for the final editing of Sections 2 and 5, whereas M.G. Grassia was responsible for Section 4 and E. Di Meglio for Section 3.

Multivariate statistical analysis gives good results in terms of synthesis and graphical representation of the information discovered. These techniques analyse the association structure in documents and create a knowledge base of the different concepts in the text.

This knowledge base can be used in different applications. For example, we can represent proximities and oppositions of different concepts on a graphical display, automatically classify documents according to some concepts and extract the information by querying the obtained index.

To analyse real situations, we can define some rules that point out situations and behaviours of the objects detectable not in an intuitive way, but only through a deep analysis of the databases using formalized methodologies. Such behavioural rules are known as Association Rules (Agrawal *et al.*, 1993).

Rules are defined as binary attributes (presence/absence,) and, if necessary, through a transformation of the data. A rule is made of a precedent and a consequent part; at the same time, it is possible to identify two distinct parts in the information contributed by a rule, called *support* and *confidence*. Support is the association strength between the considered items, confidence is the logical dependence strength expressed by the rule. The identified rules are reduced with *ad hoc* algorithms for analysing only the meaningful information.

The first order observations are generally described by classical data, while the second order observations, because of their conceptual complexity, need the use of more structured data, such as the symbolic data. The symbolic data analysis consists of a first step of collection and organisation of simple data, in a Knowledge Discovery in Database (KDD) framework.

Then, new concepts are defined in terms of complex data and analysed with statistical techniques. A *concept* is characterized by a set of properties apt to define its description, while the classes of observations that satisfy these properties, known as *objects*, represent its extent.

The objects can be created in several ways. For example, in a multidimensional dataset the categories of a variable can represent concepts to be described with the values assumed by other related variables. On the contrary, if we consider a relational database, the descriptors of the objects can be extracted with a query that expresses the properties of a set of units whose description implies the union of several relations.

Moreover, the objects are derived from the description of the classes obtained from a classification technique. In this way, it is possible to reduce considerably the first order observations. The objects are called *native* if they are the results of an expert knowledge on the phenomenon.

In this paper, we propose the joint use of multidimensional analysis techniques together with association rule building and symbolic data analysis. The aim is designing new text mining strategies, resulting in finding patterns and regularities in on-line job offer databases (Section 3), in organising the data in higher order structures and visualising them with graphic tools (Section 4), and in graphical information syntheses (Section 5).

2. The data structure: on-line job offer databases

Internet has deeply modified our approach to information. This statement applies to several aspects of common life, and to job searching.

Almost all the companies' official websites have a section dedicated to job vacancies and job offers. Nevertheless, many job thematic portals publish the announcements proposed by the selection companies and the provisional work agencies. On these websites, it is possible to insert hyper-textual links to contact the proponent or have more information, and personal documents for applying directly for job.

Among the websites specialized in the job offers we will analyse the portal <*http://www.cambiolavoro.it*>. This site contains a section dedicated to the announcements published online, a section in which the candidates can post their own curriculum vitae in four different databases (*managers*, *qualified workers*, *fresh graduates* and *unemployed*) and a section with some suggestions for improving the job search. Moreover, it is possible to subscribe to two weekly newsletters to get informed, respectively, on the proposals published by the selection companies and the provisional work agencies.

We decided to focus on the job announcements published by the selection companies because their vocabulary is fairly standardised and allows to identify, with a wider detail, competences required by the companies.

Most of the time, the documents are in some way structured. Let us think about scientific articles: they usually have an abstract, an introduction, a proposal and a conclusion. This structure can be more or less evident depending on the kind of document. We can say these documents are semi-structured.

The announcements we analysed are semi-structured. In fact, the description of the job and positions offered makes it possible to identify the characteristics and the skills required to candidates together with the information as to the economic treatment and the references for participating in the selection.

The native *corpus* is composed of more than 2000 announcements. After the deletion of the republished announcements and those related to different kinds of training courses, we proceeded to a categorisation based on the activity sector of the company and the job position offered. We analysed a collection of 726 pre-treated announcements, by normalising the data in order to reduce the risk of data splitting, and carrying out a quite in-depth lexicalisation in order to avoid ambiguity (Balbi & Misuraca, 2005).

3. Selecting the useful information in semi-structured documents

One of the main problems in treating textual data is the huge quantity of data analyzed for understanding and describing the underlying phenomenon.

Data exploration is a fundamental step in the study of natural language. Visualization is equally important because it allows the interaction with the extraction process of the significant information. In order to use the large amount of available information at best, it was organised for its subsequent processing with the recording of some meta-information, as in data warehousing systems.

The most widely used scheme to encode natural language documents is *bag-of-words*. This scheme transforms documents into *document/vectors*, a data structure to which mathematical and statistical techniques are applicable. Actually, the classical *bag-of-words* coding has some limits. Each vector/document has as many elements as the terms taken into account and so, many null values. By juxtaposing the vectors for creating a lexical table $\{terms \times documents\}$, we obtain sparse matrices whose analysis is often difficult.

Text classification is the labelling of natural language documents with thematic categories from a predefined set. It can be assimilated to a task of supervised classification and implies the task of assigning a Boolean value to each pair (d_k, c_j) in the table $\{documents \times categories\}$.

A value 'true' is assigned if a document d_k (for $k=1, \dots, n$) is classified under a category c_j (for $j=1, \dots, q$) and a 'false' is assigned if it is not under that category. This methodology exploits the structural organisation of sentences into documents in order to detect the significant ones for the following steps of the analysis. This approach, given an informative need, allows eliminating the useless information for that need. Therefore, it allows computational speed-ups, documents and term analyses targeted on the particular informative need.

Standard text classification algorithms are based on two assumptions:

- the categories do not add information to the classification procedure,
- no external information is available.

This means that documents are classified only because of their semantic content. A classifier should therefore be able to capture the semantic similarities among documents and use them in the classification procedure.

The proposed strategy (Balbi & Di Meglio, 2004), based on the principles of text classification, allows to extract interesting patterns of terms. The attempt is to go beyond the bag-of-words as we encode sentences (i.e. sequences of terms contained between two full stops) and not the whole documents. In this way, we build some sequential boundaries, as we are not interested in the general contexts in which words are used, but only in the local contexts, conveying the specific information of interest.

To 'mine' the information on a specific aspect of the document, it could be useful to consider only the sentences carrying information about the selected aspect and discard the not interesting aspects. The mining can be done by recognising the sentences that contain the information.

The first step is to eliminate non-informative terms, to obtain an indirect disambiguation of homographs through discarding the different contexts in which they might appear, to reduce the complexity and therefore the computational burdens. The sentence selection is obtained through a segmentation analysis on a training set of sentences. The generated classification rule set is validated on a test set and then applied to all the collection at hand.

Once the sentences have been identified, a text mining technique proper for the applicative domain is applied. The extracted subtexts in fact have less variability in terms of the desired information and contain less noise. This leads to better performances of visualization, retrieval and clustering techniques. The proposed strategy consists of the following steps:

[STEP 1]

- Aim: identifying sentences of interest in the document
- Tools: statistical techniques for discrimination
- Input: a training set and a test set, both consisting in sentences tagged by expert knowledge (0 = uninteresting; 1 = interesting).
- Output: logical rules for identifying interesting sentences in the document.

[STEP 2]

- Aim: eliminating uninteresting sentences in the document, by applying the logical rules identified in STEP 1
- Tools: advanced software or programming language dealing with text
- Input: the logical rules and the document to be analysed
- Output: a new document consisting only of the concerned sentences to be analysed with textual data techniques and according to predefined objectives.

Symbolic marking seems to perform well in the case of text mining, being a segmentation method with a very high performance in the case of huge data sets. Additionally, its results may be easily expressed in terms of logical operators (Balbi & Gettler-Summa, 2001).

Symbolic marking is a non-binary segmentation technique, which aims at finding the association structures in a group G_i belonging to a typology naturally defined, or obtained by a previous classification analysis. Symbolic marking takes into account logical relations, as conjunctions and disjunctions, between attributes describing the units in G_i. The result can be expressed in natural language as logical rules, connecting attributes with logical operators (Figure 1).

We applied a clustering procedure for identifying typologies of skills requirements (Table 1). The typology of reduced collection describes four skill groups: the first is mainly characterized by term *experience*. This heterogeneous class describes the skills required to experienced workers. This group also being quite large, a deeper investigation is needed. The other three classes describe more defined skill profiles, respectively, industrial relations experts, internship candidates, and salespersons.

```
------------------------------------------------------
Class C1= skill requirements weight = 168 (25.1%)
------------------------------------------------------
MARKING CORE number 1 weight = 51 (7.6%)
test-value REC DEB
weight % weight %
11.28 49 29.2 2 3.9
11.28 KNOWLEDGE
------------------------------------------------------
MARKING CORE number 2 weight = 30 (4.5%)
test-value REC DEB RECCUM
weight % weight % weight %
7.99 28 16.7 2 6.7 77 45.8
6.87 EXPERIENCE AND
2.00 NOT TENURE
------------------------------------------------------
```

Figure 1. An example of rules obtained by symbolic marking

Table 1. Classification based on the symbolic marking of sentences

Class 1 (85.5%)	**Class 2 (0.2%)**	**Class 3 (12.0%)**	**Class 4 (2.3%)**
Experience Years Title Availability Endowments	Expert Relations Industrial Milan	Word Excel Internet Internship Access Degree	Dynamic Sale Agents Aims Chemistry

We see that internship candidates are required to have a university degree and basic computer skill. To salespersons, it is required to be dynamic and work by objectives. The obtained typology describes fairly the skills required.

4. Defining professional competences

When a characteristic expressed by a category of a nominal variable is observable on all the available documents, it is possible to create homogeneous classes of documents related to that characteristic. We cannot however apply the same procedure of dimensionality reduction for the lexical table on the terms side, if we do not want to loose information. In this case, it is better to 'organise' the data in a complex structure of upper order, such as *symbolic data* (Bock & Diday, 2000).

We define a symbolic object as a triplet

$$s = (a, R, d),$$

where $d = (d_1, \dots, d_j, \dots, d_p)$ is the description of the object, based on the values assumed by a set of p descriptors $(Y_1, \dots, Y_j, \dots, Y_p)$, a is the identification function (mapping) and $R = (R_1, \dots, R_j, \dots, R_p)$ is the relation used for the comparison between the conceptual description given by d and every observation.

The descriptors of a symbolic object can be nominal, continuous or discrete and they can have several categories or values for each object. The a Boolean function assumes values {*true, false*} and allows for the identification of the elements which belong to the description set d and define the extent of the object *ext(s)*.

Let us consider a symbolic variable Y, with domain y, defined in a set E of statistical units, classes or objects, with values defined in a range $\mathcal{B}$. According to the specification of $\mathcal{B}$ in terms of y, it is possible to define the type of symbolic variable:

- if $\mathcal{B} = y$ we have the classical single-valued variable;
- if $\mathcal{B} = \mathcal{P}(y)$, with $\mathcal{P}$ function of the y non empty subsets $Y(h) \subseteq y$ (for each $h \in E$), Y assumes a set of values;
- if $\mathcal{B}$ is the set $\mathfrak{I}$ of all the intervals in y, Y is an interval variable, for each $h = E$, if $Y(h) = [\alpha, \beta]$ is an interval of values of y in the order defined in y;
- if $\mathcal{B}$ is a subset of values with $Y(h) \subseteq y$ and $|Y(h)| < \infty$, $\forall\ h \in E$, Y is a multi-value variable (categorical or numerical);
- if $\mathcal{B} = \mathcal{M}(y)$, with $\mathcal{M}$ function of the subsets of y so that $Y(h) = \pi_a$ (for each $h \in E$), where π_a is a non negative measure in y (a frequency, a probability or a weight), Y is a modal variable with domain y.

From an analytical viewpoint, it is possible to formalise the process of object construction in terms of matrices (Grassia & Misuraca, 2004). Let us consider a lexical table **T**, with p terms and n documents, and a matrix **Q** in complete disjunctive coding where a classification variable with q categories is indicated for the documents of the collection. Thus, we construct an aggregated lexical table **F** = **QT'** with the q categories of the classification variable on the rows and the p terms on the columns. The generic element is the number of times each i-th term is used by the units in the j-th category.

The **F** matrix, obtained by reducing the rows of the **T** matrix, shows the distribution of the objects created by using a classification variable among the selected terms. In order to reduce the columns of the aggregated lexical table we will organize the terms in modal variables, where each single term becomes a category of the related variable. By doing this, we will obtain a *symbolic matrix* **Z** (Figure 2), having s_q objects on rows and Y modal variables on columns, each with z_{mi} categories related to the terms.

The j-th object of the **Z** matrix is defined as:

$$s_j = \bigwedge_{k=1}^{Y} \left[Z_k = \left\{ z_{k,m}(f_{k,m}) \right\}_{m=1,2,\dots,m_i} \right],$$

where $f_{k,m}$ is the relative frequency of $z_{k,m}$, m-th category of the Z_k variable.

	Required Qualification	Language Skills	S
professional activities	ing (0.20), eco (0.40), sci (0.10), ing (0.10), sta (0.20)	ing (0.67), lis (0.33)	co
chemistry	ing (0.20), eco (0.20), ing (0.10), chi (0.10), ing (0.10), scc (0.20), sta (0.10)	ing (0.50), lis (0.25), fra (0.25)	
commerce	ing (0.17), eco (0.17), inf (0.17), dit (0.17), fis (0.17), mat (0.17)	ing (1.00)	
communication	eco (0.56), dit (0.11), sci (0.11), scc (0.22)	ing (0.50), lis (0.50)	
electronics	ing (0.21), eco (0.14), inf (0.14), dit (0.07), ing (0.21), fis (0.07), mat (0.07), ing (0.07)	ing (0.83), fra (0.17)	col (0.13), se
car manufacture	ing (0.30), eco (0.30), inf (0.10), ing (0.10), sci (0.10), ing (0.10)	ing (0.67), lis (0.33)	co
food industries	ing (0.15), eco (0.08), inf (0.15), ing (0.15), fis (0.08), chi (0.23), ing (0.15)	ing (1.00)	co
computer science	ing (0.21), eco (0.10), inf (0.21), dit (0.11), ing (0.09), fis (0.09), sci (0.09), mat (0.07), scg (0.03)	ing (1.00)	col (0.21), se
mechanics	ing (0.33), eco (0.67)	ing (0.50), fra (0.50)	sel (0.3
metal production	ing (0.67), eco (0.33)	ing (0.50), lis (0.50)	col (0.13), se
software production	ing (0.14), eco (0.14), inf (0.14), dit (0.29), sci (0.14), sta (0.14)	ing (1.00)	
research and development	chi (0.67), scg (0.33)	ing (1.00)	col (0.38), se
Tlc and logistics	ing (0.31), eco (0.06), inf (0.25), dit (0.06), ing (0.06), fis (0.06), mat (0.06), chi (0.06), ing (0.06)	ing (0.80), lis (0.20)	col (0.2

Figure 2. The symbolic matrix **Z** {*objects* × *variables*}: an excerpt

The proposed strategy enables us to visualize, in a new way, the information contained in a *corpus*, by operating a categorisation of the terms that are more interesting and a representation in terms of frequency distribution. An expert does the categorisation and so the result may be influenced by human subjectivity.

However, this *modus operandi* is necessary even in the frame of a systematisation of the term recognition process, by using a textual database manually categorised for the training of the automatic procedure. This also helps to verify the validity of the tagging operated by the expert.

The use of upper order structures makes it possible to describe the several aspects of the complex phenomenon. However, it is necessary to use specific graphical representations in order to identify the relevant information and visualize the possible similarities among the objects.

The classical two-dimensional representation is not suitable for showing the levels of the variables that characterize the objects and visualizing their composite structure.

An interesting prospective in the representation of the particular aspects of the several objects is the so-called *zoom-star* (Noirhomme-Fraiture & Rouard, 1997). These graphics derive from the "Kiviat diagrams": they are multivariate representations of "radial" type where a different variable corresponds to each axis. Instead, in the *zoom-star* it is possible to represent:

- several types of variables at the same time (interval, multinomial and modal variables);
- in the case of interval variables, the minimum and maximum limits for each symbolic object;
- in the case of multinomial variables, the respective values;
- in the case of modal variables, the values with the respective weights or the respective frequency distributions;
- the logical relations;
- the taxonomies.

The 3D *zoom-stars* procedure (Ahlberg & Schneiderman, 1994), not only give a general view and a descriptive representation of the object, but it also make it possible to visualize other kinds of information related to the distribution associated to the objects. In comparison with the factorial maps, the zoom-stars technique makes it possible to recognize correctly the forms that characterise a document. At the same time, it highlights the forms that are less frequent but that can inform on the studied phenomenon.

We can apply this strategy to analyse the language used for describing job offers, to point out the personal characteristics and the professional skills required from those that apply for a job in a specific business sector.

Table 2. Textual modal variables obtained from the *corpus* by using an expertise

Main Activity (ATPR)	ricerca; informatica; telecomunicazioni; consulenza; elettronica; information technology; commercio; manutenzione; sviluppo software; e-commerce; meccanico; sistemi informativi; assistenza tecnica; multimediale.
Required Prerequisites (RERI)	giovane; laureato; neolaureato; laurea; diplomato; curriculum vitae; professionalità; età; diploma; laureando; specializzazione; titolo preferenziale; esperienze di lavoro; master; trasferta; qualificato; preparazione; residenza; formazione universitaria.
Computer Skills (SKIN)	java; linguaggi di programmazione; unix; windows; visual basic; c++; asp; database; office; architettura; javascript; data warehouse; linux; sistemi operativi; non applicabile.
Offered Opportunities (OPOF)	formazione; inserimento; opportunità; crescita; successo; carriera; apprendere; autonomia; benefit; inquadramento; aggiornamento; realizzazione; retribuzione; incentivi; ambiente giovane; formazione continua; percorsi di carriera; training on the job; sviluppo professionale; ambiente dinamico.
Required Professionalism (PRRI)	programmatore; sistemista; manager; agente; sviluppatore; analisti programmatori; collaboratore; consulente; specialista; professionista; analista; ingegnere; figure professionali; venditore; docente; product manager; pubblicitario; project manager; stagiaire; account manager.
Required Competences (CORI)	esperienza; competenze; responsabilità; padronanza; competenze tecniche; sapere; competenze professionali; comp. tecnico-informatiche.
Business Sector (SETA)	marketing; commerciale; vendite; risorse umane; produzione; management; progettazione; comunicazione; finanza e controllo; amministrazione; logistica; ricerca e sviluppo; acquisti; customer service; engineering.
Required Qualification (LARI)	ingegneria; economia; informatica; discipline tecnico scientifiche; ing. elettronica; fisica; scienze dell'informazione; matematica; chimica e ing. chimica; ing. meccanica; scienze (com/pol); statistica; scienze (geo/bio/agr).
Language Skills (SKLI)	inglese; lingua straniera; francese.
Selection Process (SELE)	colloquio; selezionare; test psico-attitudinali; colloqui individuali.
Personal Abilities (CAPE)	capacità; dinamicità; lavorare in team; motivazione; attitudine; spirito d'iniziativa; disponibilità; spontaneità; potenzialità; flessibilità; interesse; predisposizione; forte motivazione; talento; passione; entusiasmo; rapporti interpersonali; capacità relazionali; impegno; propensione.
Contract typology (TICO)	stage; contratto; tempo indeterminato; collaborazione; assunzione.
Firm Characteristics (CARA)	italia; leadership; europa; livello nazionale; livello mondiale; dinamica; fatturato; multinazionale; certificazione.
Firm Typology (TISO)	soc a resp limitata; soc per azioni; soc di persone.
Firm Location (SESO)	nord est; nord ovest; centro; sud e isole.

With the aid of an expert, some selected terms have been used for building modal variables that characterise the objects (Table 2). The result is a symbolic matrix, graphically represented by using the zoom-stars containing the frequency distributions of the considered variables.

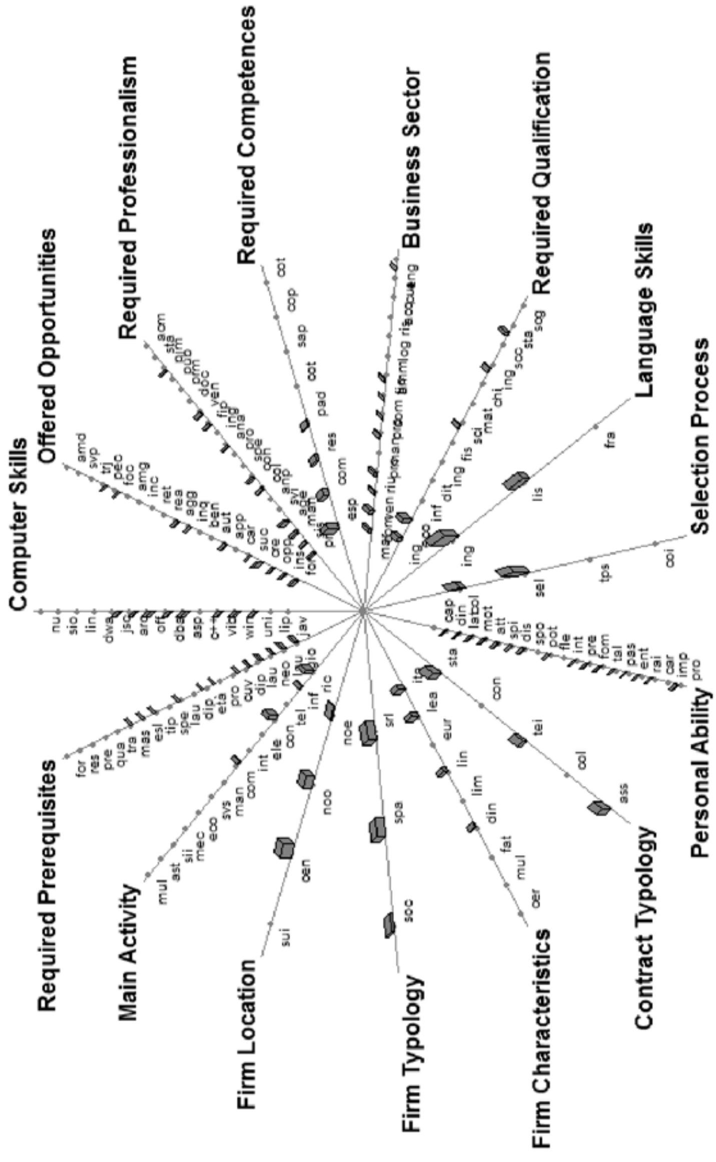


Figure 3. Zoom-star representation of Professional Activities

By using the zoom-star representation, it is possible to sketch a profile of the candidates with respect to the job positions for the different business sectors, pointing out both the frequency distribution of the skills and of the required abilities as well as the different characteristics of the firms. As an example, the representations related to *Professional Activities* (Figure 3) and *Software Houses* (Figure 4) are shown.

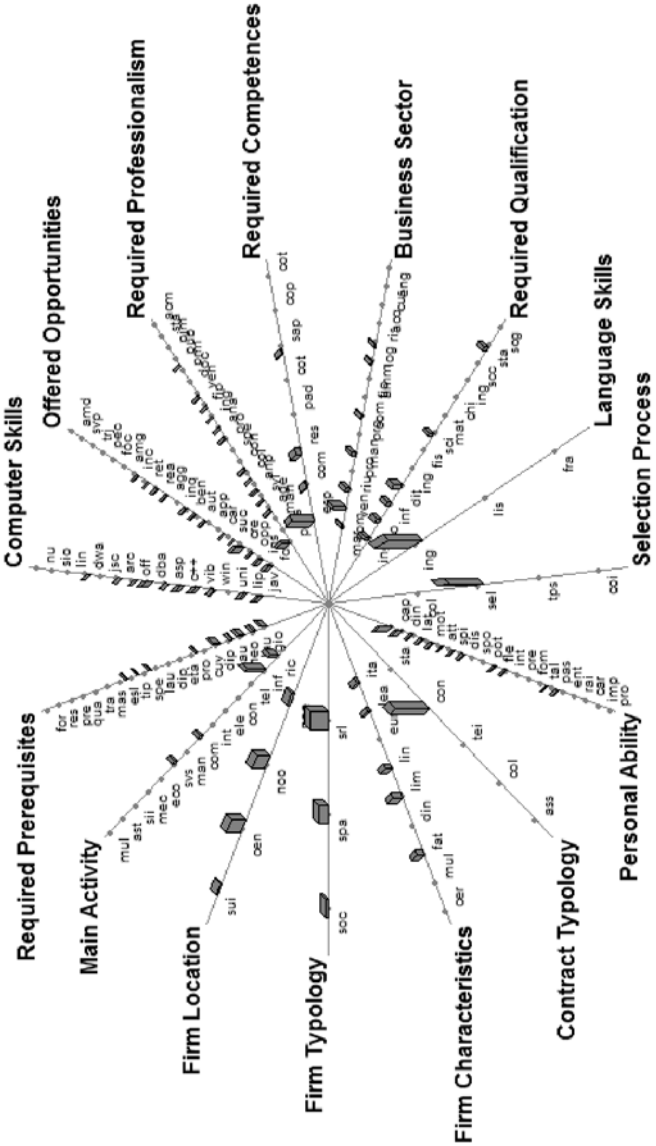


Figure 4. Zoom-star representation of Software houses

5. Lexical richness of job announcements

Many strategies of text retrieval are based on latent semantic indexing (Deerwester *et al.*, 1990) and its variations, mainly based on different weighting systems for words and documents.

Correspondence analysis and latent semantic indexing share the basic algebraic tool, the singular value decomposition, and its generalisations (Greenacre, 1984) which concern different ways of weighting the importance of each element, both in determining and representing similarities between documents and terms.

The *tf-df* family of vector based information retrieval schemes (Salton & Buckley, 1988) is very popular because of its simplicity and robustness. Some peculiarities of text analysis are the conceptual bases of the approach:

- as more frequent terms in a document are more indicative of the topic, it is important to consider f_{ij} = frequency of term i in document j;
- a normalisation of f_{ij} can be proper, by considering the number of occurrences of the most used term in each document, introducing tf_{ij}

  $$tf_{ij} = f_{ij} / max\, f_j \ ,$$

  where $max\, f_j$ is the term which occurs more frequently in j-th document;
- as terms that appear in many *different* documents are less indicative of the overall topic, it is important to measure the term *discrimination* power with the index idf_i. Naming df_i the document frequency of term i (# documents containing term i), the *inverse document frequency* of term i is given by

  $$idf_i = \log_2 (n/df_i)$$

  with n as number of documents. The logarithm may dampen the effect related to term frequency.

A typical combined term importance indicator is given by *tf-idf* weighting:

$$w_{ij} = f_{ij} / max\, f_j \cdot \log_2 (n/df_i) \ .$$

The effect of using w_{ij} is that a term i, occurring frequently in a document j but rarely in the rest of the collection, has a high weight. Many other ways of determining term weights have been proposed, but empirically *tf-idf* has been found to work properly.

The opportunity of graphically representing the similarity between documents has shown in terms of lexical richness, by using a peculiar factorial approach.

With reference to matrix **F**, our purpose is to project the cloud N_q, representing the q categories, in a lower dimensional subspace by assuming a unitary weighting system and a peculiar weighted Euclidean metric. Because of the different role played by rows and columns, we assign the same importance to all categories, but we measure the distance between categories by taking into account the different weight of the p terms, expressed in terms of *term frequency index*.

Given the i-th term frequency f_{ij}, we consider the tf_{ij} as:

$$tf_{ij} = f_{ij} / max\, f_j \,,$$

where max f_j is the number of occurrences of the most used term in the j-th category.

By considering the number of documents in each category as weights, we compute for each word the average *tf* as:

$$atf_i = 1/n \sum_j (f_{ij} / max\, f_j)\, d_j \,.$$

Let $\mathbf{\Omega} \equiv [atf_1 \ldots atf_p]^{\mathrm{T}}$ be the vector of p average *tf*, we consider as metric:

$$\mathbf{D}_\Omega \equiv \mathrm{diag}(\mathbf{\Omega}).$$

From a mathematical point of view, the method leads off with the eigen-analysis of the matrix:

$$\mathbf{A} \equiv \mathbf{F'} (\mathbf{D}_\Omega)^{-1} \mathbf{F}$$

i.e. with the generalized singular value decomposition of:

$$\mathbf{F} = \mathbf{U}\, \mathbf{\Lambda}\, \mathbf{V'}$$

$$\mathbf{U'} (\mathbf{D}_\Omega)^{-1}\mathbf{U} = \mathbf{V'V} = \mathbf{I}\,,$$

where $\mathbf{\Lambda}$ is the diagonal matrix whose elements are the singular values given by the square roots of the eigenvalues λ_α of $\mathbf{A}$, while $(\mathbf{D}_\Omega)^{-1/2}\mathbf{U}$ and $\mathbf{V}$ are, respectively, its left and right singular vectors.

The graphical representation obtained by the factorial analysis (Figure 5) helps us to evaluate the lexical richness of the document categories. The categories narrower to the origin have a wide vocabulary with respect to the others. This is readable in terms of skills required by companies with different activity sector.

The high skilled positions require, in fact, less technical competences than the lower, but more specialized, ones. Let us think for example of the duties of a front-desk clerk, especially in a small firm, with respect to those of a top manager.

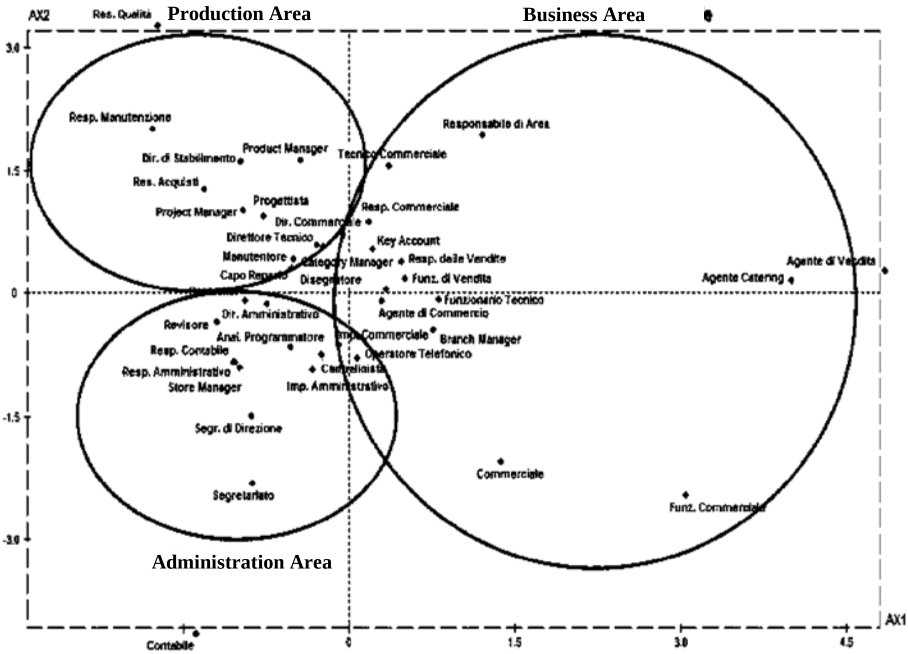


Figure 5. Factorial representation of lexical richness in job announcements

6. Conclusions

The joint analysis of texts is an interesting challenge. We have put forward a procedure for the joint use of multidimensional analysis together with unsupervised classification procedures (Symbolic marking) and association rules building. The procedure includes the development and the application of clustering complex data, for compressing and better organising elementary data in higher order structures. Our text mining strategy can find patterns and regularities in a database, organise elementary data in higher order structures, and visualise them with graphic tools.

References

AGRAWAL R., IMIELINSKI T., SWAMI A. (1993) Mining Associations between Sets of Items in Massive Databases. In: *Proceedings of the ACM-SIGMOD 1993 Int'l Conference on Management of Data*, Washington D.C.: 207-216.

AHLBERG C., SCHNEIDERMAN B. (1994) Visual Information seeking: tight coupling of dynamic query filters with starfield displays. In: *Proceedings of Conference on Human Factors in Computing Systems (CHI '94)*, ACM, Boston: 313-317.

BALBI S., DI MEGLIO E. (2004) Una strategia di Text Mining basata su regole di associazione. In: AURELI CUTILLO E., BOLASCO S. (eds) *Applicazioni di analisi statistica dei dati testuali*, Casa Editrice Università di Roma La Sapienza: 29-40.

BALBI S., GETTLER-SUMMA M. (2001) Identifying lexical profiles by symbolic marking. In: *Book of Short Papers CLADAG2001*, Palermo: 185–188.

BALBI S., MISURACA M. (2005) Visualization Techniques in Non Symmetrical Relationships. In: SIRMAKESSIS S. (ed.) *Knowledge Mining (Studies In Fuzziness and Soft Computing)*, Springer-Verlag, Heidelberg: 23-29.

BOCK H., DIDAY E. (2000) *Analysis of Symbolic Data*, Sprinter-Verlag, Heidelberg

DEERWESTER S., DUMAIS S.T., FURNAS G.W., LANDAUER T.K., HARSHMAN R. (1990) Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, **41**: 391–407.

GETTLER-SUMMA M. (1998) *MGS in SODAS: Marking and Generalization by Symbolic Objects in the Symbolic Official Data Analysis Software*, Cahier no. 9935, CEREMADE, Université Paris IX – Dauphiné.

GRASSIA M.G., MISURACA M. (2004) Il candidato ideale: Analisi delle professionalità e delle competenze. In: AURELI CUTILLO E. (ed) *Strategie metodologiche per lo studio della transizione Università-Lavoro*, CLEUP, Padova: 247-257.

GREENACRE M.J. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press, London.

MANNING C.D., SCHÜTZE H. (2001) *Foundations of Statistical Natural Language Processing, MIT Press*, Cambridge.

NOIRHOMME-FRAITURE M., ROUARD M. (1997) Zoom Star: a solution to complex statistical object representation. In: HOWARD S., HAMMOND J., LINDGAARD G. (eds) *Human-Computer Interaction - Proceedings INTERACT'97*, Sydney.

SALTON, G., BUCKLEY, C. (1988) Term-weighting approaches in automatic text retrieval, *Information Processing & Management*, **5**: 513–523.

Companies, Competences and Graduates' Selection Processes: an Attempt to Quantify

Marisa Civardi, Emma Zavarrone[1]

Department of Methods of Economics, University of Milano - Bicocca, Italy

Summary. The assessment of the employment potential of University graduates represents a direct evolution of the quality evaluation of higher education systems. This paper aims to identify competences that companies, which subscribed the Vulcano-database, require when selecting graduates for employment. Our analysis consists of two parts: one in which a quantitative score is associated through Rasch analysis to the companies' interest with respect to four classes of characteristics and competences they take into consideration when selecting candidates. In the second part, we apply a segmentation analysis to check whether there is correspondence between companies that prefer certain university degrees and the interest levels towards the characteristics and competences so quantified.
Keywords: Rasch analysis; Segmentation analysis; Rumm; C@rt; Graduates' competences.

1. A survey on companies' interests

Studies on self-assessment of work experience and career expectations of university graduates draw a general picture of graduates' transition from higher education to employment in Italy. Instead, studies on companies' interests are less frequent.

The awareness about the candidates' features companies consider relevant in the recruitment process may help universities to improve their educational offer.

This paper refers to a CATI survey carried out by Checchi & Pravettoni

[1] In this joint work, M. Civardi was responsible for the final editing of Sections 3, whereas E. Zavarrone was responsible for the other Sections.

(2004) at Milan State University[2] in 2003. The survey was conducted on a non-probability sample of 278 companies, out of the 476 that subscribed for at least one year the services of Vulcano database since its start in February 1998.

Vulcano database was set up in cooperation with Cilea[3] to link university graduates and the companies. It is a free access service for graduates, who are given a password by their own university for writing and possibly editing their curriculum vitae (CV). Companies can have either free or fee access to the database for recruitment purposes.

The survey aimed to identify companies' structural characteristics and their willingness to collect some information about the candidates' profiles and competences from the database.

"Sections" D6, D7, D8 and D9 of the questionnaire suited the objectives of this study. The aim of these questions was to measure the importance companies attributed to four classes of "characteristics and competences" of graduates. Each section consists of a set of items obtained on a 4-grade ordinal scale ranging from 0=no interest to 3 = greatest interest.

2. Methodology and application context

The following data were analysed:

- the items of section D6 derived from graduates' administrative records (age, sex, schooling, grades, dissertation topic) edited by the university services;
- graduates' curriculum (study experience abroad, foreign language skill, IT skill, internships, military service), described in section D7;
- skills evaluated by companies at graduate's recruitment stage (communicating, speaking, listening, synthesising, emotion control, concentration, understanding different points of view, flexibility, problem solving, honesty, creativity, self-confidence, study and career expectation consistency), described in section D8 of the questionnaire;
- other competences the companies take into consideration while evaluating candidates for recruitment (behavioural, relational, and organisational attitude and professional skills), described in section D9 of the questionnaire.

---

[2] We are grateful to the Authors and to COSP (Centre for Study and Career Advice) of Milan State University for consenting us to use of micro data.

[3] CILEA, established in 1974, provides Information and Communication Technology services on behalf of universities and related organizations, public organizations and enterprises. It provides also professional advice for both the planning and dissemination of advanced technologies in the fields of high performance computing, networking services and informatics.

We applied the Rasch method to estimate the relevance of each one of the four dimensions. This method creates a metric scale for each dimension and places both the characteristics of the selection process and the companies on an underlying continuum. This continuum reflects the ideal candidate's profile between the two extremes of "no importance" on the left and that of "highest importance" on the right. The Rasch results are two-fold: the *subject's ability* (i.e. his/her propensity towards the latent continuum) and the *item difficulty* that identifies the position of each item on the latent continuum (Wright & Masters, 1982).

With reference to our application, subjects are the companies, while item difficulty is the importance assigned to a specific graduate's feature by companies. The subject's score is obtained through the scores the companies assigned to the items; the item difficulty is the level of companies' endorsement of the items.

As response items are ordinal, we chose a Partial Credit model to fit the data. This model is based on the hypothesis that the ability of the i-th subject to pass from one response scores to the next can vary according to items. According to Rasch's terminology, if $Y=0, 1, 2,..., k_j$ is the ordinal response vector for the j-th item, the difficulty of indicating score y in item j for a subject is larger than the difficulty of each response placed below y in the scale.

Therefore, if δ_{jy} expresses the difficulty to indicate score y for item j:

$$\delta_{jy} = \sum_{w=0}^{y} \delta_{jw} = \sum_{w=0}^{y} (\delta_j + \tau_{jw}), \qquad [1]$$

δ_j indicates the mean difficulty of the j-th item, while τ_{jw} the threshold of the scores preceding scores y. The probability π_{ijy} of subject i to respond y to item j is obtained by:

$$\pi_{ijy} = \frac{exp \sum_{w=0}^{y} (\beta_i - \delta_{jw})}{\sum_{q=0}^{k_j} exp \sum_{w=0}^{q} (\beta_i - \delta_{jw})}, \qquad [2]$$

where β_i is the ability of subject i. For a subject i, the natural logarithm of the ratio between his/her probability to assign score y to item j and the probability to score y-1, is represented by a multi-parametric logistic model:

$$\ln\left(\frac{\pi_{ijy}}{\pi_{ij(y-1)}}\right) = \beta_i - \delta_{jy}. \qquad [3]$$

Expression [3] determines the value of ability β_i, the core dimension of the analysis. In fact, β_i indicates the position of a company on the continuum and varies between "no importance" (minimum β_i) to "very large importance" (large positive β_i).

Table 1. Ability scores descriptive statistics for the four dimensions, full sample

	D6	*D7*	*D8*	*D9*
Sample size	278	278	186	278
Mean	-0,681	-1,083	1,160	1,741
Variance	2,689	3,610	0,785	0,660
Percentile				
10	-3.590	-4.323	0.159	0.759
20	-3.590	-4.323	0.562	1.119
30	-0.744	-1.086	0.842	1.245
40	-0.233	-0.697	1.026	1.509
50	0.013	-0.508	1.221	1.650
60	0.136	-0.108	1.221	1.861
70	0.391	0.111	1.431	2.129
80	0.524	0.344	1.664	2.319
90	0.812	0.873	2.251	2.783

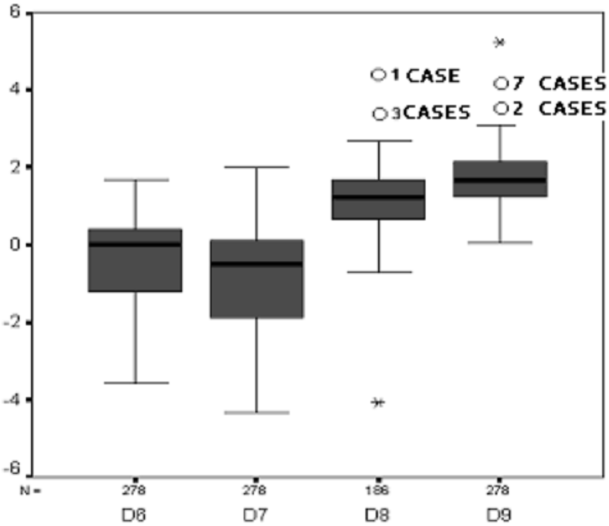


Figure 1. Ability scores statistics for the four dimensions, full sample.

Descriptive statistics of ability score for the four latent dimensions obtained with Rasch analysis, using Rumm software, are presented in Table 1 and Figure 1.

The ability scores for dimensions D6 (official graduates' CV) and D7 (general competences for work) present the greatest variability. Companies showed different levels of interest towards these two dimensions.

The mean values for these two dimensions are smaller than those for D8 and D9, which concern the features the companies regarded as highly desirable. We can conclude that the information stored in the Vulcano database are less relevant for the companies.

Nevertheless, 92 out of the 278 sampled companies did not associate any score to D8, in fact, they did not answer to these questions. Consequently, a second analysis was carried out on the subset that evaluated all items. Table 2 and Figure 2 show results of this more pertinent analysis.

Table 2. Ability scores descriptive statistics for the four dimensions, companies responding to all items

Sections	*D6*	*D7*	*D8*	*D9*
Sample size	186	186	186	186
Mean	0.121	-0.134	1.160	1.747
Variance	0.569	0.793	0.785	0.586
Percentile				
10	-0.650	-1.086	0.159	0.877
20	-0.357	-0.697	0.562	1.119
30	-0.110	-0.678	0.842	1.245
40	0.111	-0.313	1.026	1.509
50	0.260	-0.108	1.221	1.650
60	0.391	0.111	1.221	1.798
70	0.524	0.344	1.431	2.129
80	0.663	0.597	1.664	2.319
90	0.975	0.965	2.251	2.533

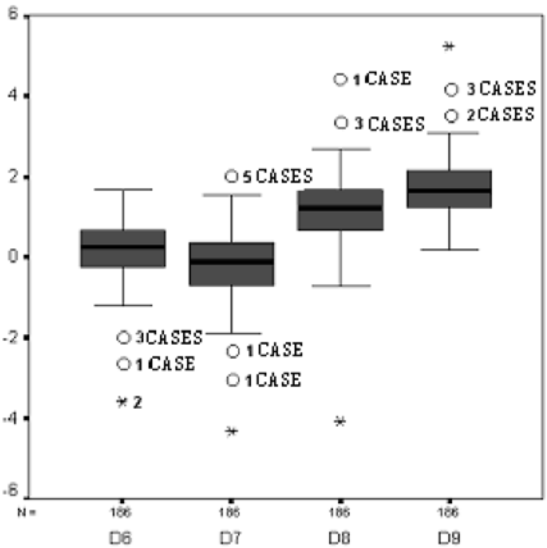


Figure 2. Characteristics/competences ability score box-plot - selected sample

Table 3. Correlation matrix, full sample

	D6	*D7*	*D8*	*D9*
D6	1	,906	0,115	-0,029
D7		1	0,217	0,006
D8			1	0,535
D9				1

Table 4. Correlation matrix, partial sample

	D6	*D7*	*D8*	*D9*
D6	1	0,487	0,115	0,145
D7		1	0,217	0,191
D8			1	0,535
D9				1

The greatest difference between the two analyses concerns dimensions D6 and D7, whose scores are much more homogeneous in the subset sample than the full one. In particular, this subset expressed the highest scores, the ones closer to the greatest importance extremity. On the contrary, there is no difference between the characteristics/competences taken into consideration at the time of the interview and the “ideal competences” expressed by D8 and D9 sections.

Tables 3 and 4 indicate values of linear correlation coefficients *r* between each pair of graduates’ features, respectively for total and partial samples. For both, the score associated to D6-D7 and D8-D9 pairs are significantly correlated. D7 and D8 are also correlated, although with a lower *r* value. In the subset sample, D6 and D9 are correlated, too, at 5% significance level.

3. Ability scores for graduates’ competences

The ability scores show the importance assigned to the four graduates’ features by companies at the recruitment stage. This raises the question whether different importance levels for each dimension correspond to the type of degree in which companies are interested. If that is the case, on the one hand, university degrees preferred by companies can be predicted. On the other, it can help the definition of an appropriate university formative offer.

We used a segmentation analysis for this purpose. This method detects groups of units characterised by maximum differences on a criterion variable according to a set of binary splits on selected covariates. In this analysis, the criterion variable was the proportion of companies that consulted the Vulcano database for selecting graduates from a specific faculty. The possible predictors were the graduates' features.

We applied the segmentation algorithms implemented on the AnswerTree 3.5 (additional SPSS software) following the C@rt procedure proposed by Breiman *et al.* (1984). Gini's impurity criterion was used as the segmentation criterion, and the minimum number of units in the parent and child nodes was set, respectively, to 5 and 2. The analysis was conducted on:

1. the courses of study of Mathematics, Physics, and Natural Sciences (Science Faculty);
2. the courses of the Faculty of Political Sciences;
3. the courses of the Humanities Faculty.

The segmentation was conducted on both the 186 companies, where the full information was available, and on the 278 companies, after the exclusion of the explanatory variable D8.

3.1 Graduates in Mathematics, Physics, and Natural Sciences

For the courses of the Faculty of Sciences, the access rate to Vulcano database did not differ substantially: 62.6% out of the 278 companies had access to it in order to select graduates of these courses, while in the subset of 186 companies the proportion was 64.5%.

Within both samples, the companies interested in selecting these graduates were characterized for a score lower than or equal to 1.309 in the dimension D9, and a D7 score greater than 0.470. With regard to the whole sample, these companies occupied the lower positions (up to 30th percentile) on D9 dimension, and the top ones (above the 70th percentile) with reference to dimension D7.

In the screened sample, the significant predictors were the same of the previous analysis, but in the latter case companies scored in the D7 dimension above the 80th percentile, far above the total sample one. In both cases, dimension D6 did not show differences between companies' propensities. Therefore, the analysis seems to indicate that the companies interested in graduates of the scientific faculty of the Milan Athenaeum, beyond being the most numerous, give the maximum importance to the cross-occupational skills achieved by the graduates while, in the direct job interview, assign relatively little importance to the behavioural, relational and organizational attitudes.

3.2 Graduates in Political Sciences

With reference to the full sample of companies, the segmentation analysis identified two main groups. The first one, which is small but well characterised, consists of the companies with D7 larger than 0.470 and $2.43<D9\leq 3.14$.

Although D6 appears in the lowest part of the tree, it does not seem to discriminate. These companies give the outmost importance to both graduates' CVs and their behavioural, relational and organisational attitude (D7 scores are above the 80^{th} percentile and between the 80^{th} and the 90^{th} percentile for D9). The second group of companies that are interested in Political Science graduates manifest a relatively small importance to dimensions D7 and D9 ($D7 \leq 0.47$ and $D9 \leq 1.579$). Again, D6 does not discriminate.

In the selected sample, the segmentation analysis identifies a small number of companies with a score higher than 1.361 for D7, while D6 scores are put together along the two distributional tails: the companies placing above the 90^{th} percentile are particularly interested on graduates' study curriculum. Amongst the companies with a D7 score lower or equal to 1.361, D6 is discriminant. Conversely, D9 does not discriminate as it excludes less than 10% of the companies (the value of this percentile being 0.877).

In conclusion, the segmentation analysis suggests that companies that are interested in Political Sciences graduates either assign high importance to D7 or, if they consider D7 less important, they take into consideration even the administrative data.

3.3 Graduates in the Humanities

The segmentation analysis carried out on the total sample shows that amongst the companies interested in Humanities graduates, a rather consistent number show D7 scores higher than -0.210 and lower or equal to 0.002 with D6 scores between -0.684 and 0.075. D9 does not seem to be a discriminatory dimension.

These results show that the importance assigned to administrative information are very close to the median, while the importance assigned to the experience is placed just above the median (60^{th} and 70^{th} percentile).

In the selected sample, segmentation analysis identified companies with D7 scores between -0,210 and 0.0015. With reference to experience, the results of the total sample are confirmed. D6 does not seem discriminant. Conversely, D8 identifies companies with scores lower than 0.075, i.e. companies that give little importance to candidates' competences emerging during the selection interview.

Hence, companies interested in selecting graduates in the Humanities give a medium–high importance to study curriculum, a medium-low importance to administrative characteristics and a very low importance to competences emerging from the selection interview. All this could indicate that graduates in the Humanities are required for general administrative positions.

4. Conclusions

The quantification of the characteristics the graduates should possess for finding a job is a complex process, because of the difficulty in identifying the interactions occurring amongst the factors involved. Undoubtedly, the administrative information "certified" by the universities leads the selection process of candidates, even if the companies do not always consider it as the decisive element for recruitment. Actually, companies pay great attention to the candidates' competences exhibited at the recruitment interview stage, but conditional to their university degree.

An in-depth analysis of the three actors involved would be needed: the university, who is responsible for the making of competences offer, the graduate, as the holder of the competences, and companies, who require it. Our approach, which is based on the quantification of the importance companies assign to graduates' competences, could be the connecting link amongst the three actors.

References

BREIMAN L., FRIEDMANJ.H., OLSHEN R.A., STONE C.J. (1984) *Classification and Regression Trees*, Wadsworth, Belmont, CA.

CHECCHI D., PRAVETTONI G. (2004) Reperimento di mano d'opera e richiesta di competenze. Una indagine sulle imprese utilizzatrici di VULCANO. In: AURELI CUTILLO E. (ed) *Strategie metodologiche per lo studio della transizione Università-lavoro,* CLEUP, Padova: 115-139.

WRIGHT B.D., MASTERS G.N. (1982) *Rating Scale Analysis*, Mesa Press, Chicago.

Tesserae of Graduates' Competences for Mosaics of Work Activities

Luigi Fabbris, Gilda Rota, Ilaria Silvestri, Anna Boaretto[1]

Padua University, Italy

Summary. In this paper, we analyse with multivariate statistical methods the frequencies of use of the basic, cross-occupational and occupation-specific competences by employed graduates. The basic idea is that competences may be *tesserae* of a mosaic of graduate's professional personality. We analyse the relationship between competences and a set of professions. We evaluated the differential effect, on competence use, of the specific degree held by each graduate, so to connect graduate's features and working environment's characteristics in terms of a selected set of competences.
Keywords: Basic competences; Cross-occupational competences; Occupation-specific competences; Multilevel analysis; Correspondence analysis; Survey on graduates; University of Padua.

1. Professional skills of employed graduates

The work activities of graduates can be characterised by the frequent and intensive use of their competences, in other words, the knowledge, skills and willingness with which they respond to professional needs. Blended together with their values, the competences define their "professional personality" or "professional style".

Knowledge is a set of principles and usable information, abilities are durable talents that make a person capable of performing work activities, and willingness is the personality trait that favours the use of abilities.

---

[1] The four authors designed and realised jointly the research work described in this paper. However, A. Boaretto wrote Sections 3 and 5 of the paper, whereas G. Rota wrote Section 2, I. Silvestri Section 3 and L. Fabbris the other sections.

The skills necessary to carry out a task are complex and multifunctional. They are multifunctional in that they help solve many tasks, and for this reason we call them "cross-occupational" (Spencer & Spencer, 1993; Isfol, 1998; Ajello & Meghnagi, 1998). They are complex because what individuals have obtained from education and professional training is processed by experience, thus enabling individuals to choose consciously their own values and goals.

Skills can be broken down into independent units. The units identified in the DeSeCo project (Befani, 2004) are:

(i) the capacity to face complexities, which consists of the capacity to break up and analyse reality by distinguishing between what fits within known parameters and what is new,

(ii) the skill of perception, which consists of the capacity to grasp the sense of situations and to vary one's own behaviour accordingly,

(iii) the self-governing skill, which consists of a guideline of one's own actions based on values determining what is right, opportune and desirable,

(iv) co-operation skills, which involve the capacity of identifying in other people's abilities the particular talents that make teamwork possible, and

(v) the skill of narration, which consists of being able to disassociate from contingency, identify alternatives and process viable future scenarios.

Many skills are not taught in school. Many remain implicit until when they are applied to activities to be undertaken. For example, it is hypothesized that individuals are competent to work in a team if they are capable of listening, presenting problems in a suitable language and style, empathizing with others, controlling their emotions, mediating between differing interests and impositions, etc. Nevertheless, the fact that they are capable of working in a team will be proven only when they are faced with a work problem requiring teamwork.

Graduates seeking employment are required to possess various types of skills and knowledge. Those that are taken for granted in job candidates are called "basic" or "strategic/cross-occupational" skills. For a graduate, basic skills include the written and oral knowledge of English and basic computer literacy. The level required for these two skills varies according to the characteristics of the company and grows in conjunction with market globalization and the rise of education and professional training in the world[2].

Basic skills are relevant to the moment of initiation of a job, but are not necessarily used in the performance of professional activities. Indeed, Fabbris & Visentin (2004) found that basic skills are used in companies less intensely

[2] Many researchers stress that a basic skill is also the knowledge of the Italian language, not taking for granted that even people who have gone through at least 16 years of the Italian educational system know it.

than one would surmise by their presence in the list of requirements necessary for recruitment.

The cross-occupational skills required of recent graduates depend on the position they are to hold. In equivalent positions, they may depend on work parcelling and on the sectoring of company activities. In large companies where production is standardized and workers' individual contribution is a part of the whole, technical skills are used with greater intensity than in small companies. The multifunctionality of the positions of graduates is, on the other hand, a peculiarity of small-to-medium companies. In these companies, graduates must perform a number of specific tasks pertaining to their position, but they are also expected to perform numerous other lower- and upper-level activities (Fabbris & Visentin, 2004).

Occupation-specific skills are the graduate's professional peculiarity. For example, to an engineer in a mechanical apparatus factory the essential skills are: knowing how to make calculations, analyses, projects, designs, knowing how to operate machinery and installations with complex equipment, etc., as well as the *forma mentis* characterizing engineers. To a marketing operator in a market-oriented company the important skills are: knowing how to develop contacts with partners, customers, sellers and mediators, participating in trade fairs and markets in Italy and abroad, knowing how to use the Internet to gain knowledge of the competition's offers, etc. as well as the *forma mentis* specific to marketing and sales personnel.

What tells cross-occupational skills apart from specific skills is not technicality, but functional generality. Indeed, Levy & Murnane (1999) call cross-occupational skills *soft skills*, identifying them as technical skills but not as *hard* as specific skills.

Graduates are usually involved in positions "from technician on up" (Del Favero & Fabbris, 2002), positions that also include middle-ranking managers, management assistants, executives and consultants. Recent graduates may even hold managerial positions in their first employment, just as they may accept to perform manual activities; these are nevertheless improbable positions.

In this paper, we intend to identify a limited number of possibly independent units in the use of graduate work skills (Section 3) and to create maps of skills to classify graduate professions (Section 4) with the units found. In Section 5, a multilevel analysis is made of the units deriving from the use of skills. Our paper ends (Section 6) with a few comments and suggestions.

The data used in the analysis was collected by the University of Padua in the course of a survey that began in 2000 and ended in 2005 (Fabbris, 2003). The sample was made up of five cohorts selected from as many undergraduate sessions.

Six months after graduating and every six months after that for a total of three years, specialized interviewers via a CATI (*Computer Assisted Telephone Interviewing*) system interviewed the sampled graduates.

2. Cross-occupation and occupation-specific skills

The analysis of the skills used by University of Padua graduates when performing professional activities (Table 1) shows that technical skills are frequently used by graduates and that the frequency of the use of skills varies according to their study programme.

Almost everyone uses the computer as a daily calculation, memorization, and communications tool within the organization (86%) and use his/her relational (82%) as well as English communication skills (77%) to maintain positive relations with customers/users and suppliers.

The computer is a functional tool in the activities of newly hired personnel in the company information network. Some employees, hired for technical-scientific related jobs, use the computer to perform complex calculations and designs. The importance of training to develop this skill is unquestionable.

Table 1. Frequency of use of basic and cross-occupational skills by graduates of the University of Padua, by type of study programme and faculty

Faculty	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Agriculture	91.5	95.7	38.7	71.7	75.5	53.2	50.0
Bachelor	*91.5*	*96.3*	*40.7*	*70.0*	*78.0*	*52.4*	*54.9*
Master	*91.7*	*91.7*	*25.0*	*83.3*	*58.3*	*58.3*	*16.7*
Economics	65.9	81.8	34.1	84.1	56.8	63.6	50.0
Pharmacy	80.3	87.7	18.2	95.5	53.0	51.5	48.5
Law	86.8	92.1	31.6	84.2	64.5	38.7	29.0
Bachelor	*86.8*	*92.6*	*32.4*	*82.4*	*63.2*	*37.3*	*27.9*
Master	*87.5*	*87.5*	*25.0*	*100.0*	*75.0*	*50.0*	*37.5*
Engineering	87.2	79.0	31.4	82.7	86.7	63.5	80.7
Bachelor	*87.2*	*78.2*	*30.9*	*80.9*	*85.6*	*60.6*	*79.3*
Master	*87.1*	*83.9*	*41.9*	*93.5*	*93.5*	*80.6*	*90.0*
Letters, philosophy	91.5	83.0	46.1	98.5	15.5	54.3	43.8
Veterinary Medicine	79.0	94.7	42.1	79.0	52.6	68.4	57.9
Psychology	75.0	74.0	13.5	79.2	59.4	33.7	38.5
Education Science	92.1	60.4	18.0	77.5	65.9	23.0	56.8
MPN Sciences	87.7	81.8	20.6	62.4	55.9	51.2	67.1
Bachelor	*86.9*	*80.0*	*20.0*	*61.4*	*54.5*	*50.3*	*66.9*
Master	*92.0*	*92.0*	*24.0*	*68.0*	*64.0*	*56.0*	*68.0*
Political Science	90.1	80.4	38.5	90.1	67.0	47.8	41.3
Statistics	88.3	44.2	15.4	69.2	60.3	41.0	47.4
Bachelor	*88.2*	*43.1*	*15.4*	*75.0*	*63.5*	*40.4*	*50.0*
Master	*88.5*	*46.2*	*15.4*	*57.7*	*53.8*	*42.3*	*42.3*
University of Padua	85.8	77.5	28.2	81.7	64.5	49.0	56.4
Bachelor	85.5	77.4	27.8	81.6	63.4	47.6	55.3
Master	88.5	79.1	32.0	82.2	78.0	66.2	70.0

1: Computer use skill; 2: English skill; 3: Other foreign languages skill; 4: Relationships with customers; 5: Working basically in team; 6: Organising other people's work; 7: Writing reports

The relational skills useful in interfacing the productive structure the graduate works for (required of 65% of employed graduates) with the outside world could be considered a personality trait that is difficult to develop with a university education. Nevertheless, the relational skills necessary for working in a team, which involve extroversion, the ability to argue convincingly, to uphold one's own theories in public and to integrate and mediate between contending positions, may be cultivated via university studies. For instance, these skills may be developed by opting for oral exams over written ones, and by examining for academic progress in terms of empirical teamwork.

More than other graduates, those holding degrees from the School of Letters and Philosophy or from the School of Political Science (98.5% and 90.1%, respectively) use relational skills with customers, users and suppliers (Table 2), whilst Engineering graduates are more likely to organize other people's work (86.7%) and to write technical reports for the company (80.7%).

Table 2. Frequency of use of occupation-specific skills by University of Padua graduates and parameters of the factorial analysis of the matches, by faculty

Skills	*Master*	*Bachelor*	*Factor loads*	*Communality*
Faculty of Agriculture				
Application of agronomic analysis techniques	44.4	8.3	0.31	0.573
Application of forestry analysis techniques	34.1	9.1	0.24	0.357
Application of zootechnical analysis techniques	15.9	0.0	0.18	0.199
Use of economics, estimates, etc.	50,0	33.3	0.25	0.375
Use of lab and technical and scientific instrumentation	31.7	16.7	0.18	0.189
Use of topological, GIS analysis techniques	44.4	16.7	0.27	0.438
Use of vegetable protection techniques	34.1	16.7	0.23	0.311
Faculty of Economics				
Accounting and budget analysis	72.8	=	0.12	0.248
Financial market analysis and decisions	31.8	=	0.13	0.304
Legal aspects of company management (commerce and taxes)	56.8	=	0.11	0.203
Management control, budgeting and reporting	59.1	=	0.16	0.448
Logistics process management	32.6	=	0.12	0.272
Company finances management	40.9	=	0.17	0.492
Marketing and communication planning/management	59.1	=	0.05	0.481
Company planning and management	59.1	=	0.16	0.459
Planning of computer systems	30.2	=	0.13	0.310
Programming of raw material production / purchasing	27.3	=	0.13	0.288
Auditing	38.6	=	0.13	0.294

Skills	*Mas-ter*	*Bache-lor*	*Factor loads*	*Com-munality*
Use of mathematical/statistical decisional models, data analysis	43.2	=	0.10	0.171
Use of organizational and process analysis tools	47.7	=	0.17	0.481
Management and development of human resources	37.2	=	0.10	0.164
Faculty of Pharmacy				
Familiarity with pharmaceutical legislation	90.9	=	-0.21	0.134
Chemical, microbiological and technological analysis	13.2	=	0.44	0.568
Development of pharmaceuticals, chemicals and cosmetics	18.2	=	0.47	0.638
Use of lab techniques and instruments	43.9	=	0.36	0.371
Faculty of Law				
The ability to interpret or apply new laws	91.2	87.5	0.23	0.328
The ability to write an act or legal opinion	75.0	75.0	0.30	0.547
The ability to orientate concrete cases to judicial categories	91.2	87.5	0.19	0.233
Composition of disputes, mediation between subjects and interests	79.4	62.5	0.25	0.419
Familiarity with contractual regulations	88.2	100.0	0.19	0.227
Search of bibliographical info on laws and regulations	77.6	37.5	0.21	0.284
The ability to hold a legal discussion	88.2	75.0	0.27	0.460
Faculty of Engineering				
Discussion of technical/scientific problems	84.5	90.3	0.33	0.345
Design/planning	48.4	38.7	0.30	0.274
Development of mathematical models	30.9	25.8	0.33	0.344
Calculation	58.8	58.1	0.35	0.385
Search of information on databases and reading material	78.6	64.5	0.29	0.27
Use of lab and technical and scientific instrumentation	52.9	74.2	0.22	0.151
Faculty of Letters and Philosophy				
Application of economics/corporate knowledge	34.1	=	0.15	0.071
Cultural, museum and environmental asset cataloguing	25.2	=	0.21	0.143
Written and oral dominion of the language	96.1	=	0.42	0.572
Search of sources and data in the Internet and reading material	72.1	=	0.38	0.471
Problem-solving	88.3	=	0.41	0.563

Skills	*Master*	*Bachelor*	*Factor loads*	*Communality*
Faculty of Veterinary Medicine				
Data management and analysis	57.9	=	0.34	0.372
Application of clinical knowledge	73.7	=	0.46	0.653
Use of technical and scientific instrumentation	79.0	=	0.49	0.750
Faculty of Psychology				
Analysis of observable and experimental data	44.8	=	0.16	0.417
Application of personality research techniques	33.3	=	0.18	0.534
Application of behaviour observation techniques	69.8	=	0.15	0.350
Analysis of group dynamics	62.5	=	0.16	0.097
Mediation between subjects, interests and cultures	71.9	=	0.08	0.119
Orientating toward study and/or work	49.0	=	0.09	0.198
Planning of educational intervention	62.5	=	0.11	0.317
Conduction of clinical interviews	29.2	=	0.14	0.227
Familiarity with and use of data sources	59.4	=	0.12	0.287
Construction of tools to measure psychological variables	20.8	=	0.13	0.229
Experimental research	23.2	=	0.12	0.384
Planning therapy	34.4	=	0.15	0.476
Administration and correction of tests	36.2	=	0.16	0.399
Faculty of Education Science				
Data analysis	64.8	=	0.12	0.337
Active listening skills	95.7	=	0.12	0.322
The application of edutainment and educational methods/techniques	74.6	=	0.15	0.476
Diagnosis of the needs/potentials of people and community	83.5	=	0.16	0.606
Interpretation of events as per education science	79.1	=	0.13	0.392
Mediation between subjects, interests and cultures	79.0	=	0.13	0.353
Orientating toward study and/or work	57.6	=	0.11	0.272
Systematic observation and use of methods and tools	79.1	=	0.13	0.364
Planning, monitoring and assessing educational events	79.9	=	0.16	0.573
Promoting interpersonal and inter-group interaction	87.1	=	0.14	0.419
Research-action skills	64.8	=	0.13	0.369
Use of mediated and direct language and techniques	89.1	=	0.1	0.236
Faculty of Mathematical, Physical and Natural Sciences				
Mediation between subjects, interests and cultures	62.1	56.0	0.09	0.033
Discussion of scientific and technical problems	74.3	72.0	0.27	0.290

Skills	*Master*	*Bachelor*	*Factor loads*	*Communality*
Development of mathematical models	33.6	36.0	0.27	0.278
Organization of lab activities	42.4	44.0	0.38	0.565
Finding, consulting and use of database information	74.5	76.0	0.25	0.254
Use of laboratory and scientific instrumentation	54.5	56.0	0.38	0.564
Faculty of Political Science				
Familiarity with budgets, accounting	54.4	=	0.31	0.511
Analysis of statistics	38	=	0.21	0.546
Familiarity with corporate economics, marketing	58.2	=	0.22	0.241
Orientating toward study and/or work	39.6	=	0.12	0.244
Composition of disputes, mediation between subjects interests and cultures	69.7	=	0.22	0.070
Knowledge of contractual, labour, corporate law	57.6	=	0.26	0.260
Knowledge of civil law	44.6	=	0.11	0.356
Knowledge of economics and financial science	64.8	=	0.32	0.660
Faculty of Statistics				
Budget analysis, management control	42.3	38.5	-0.02	0.003
Familiarity with and use of official data sources	51.9	42.3	0.2	0.347
Construction and interpretation of decision-making indicators	51.9	46.2	0.18	0.265
Population and social forecasting	11.5	8.0	0.19	0.297
Economic forecasting	34.6	38.5	0.14	0.162
Computer simulation (Montecarlo method)	34.6	26.9	0.17	0.252
Interpretation of miscellaneous analysis results	32.7	30.8	0.24	0.469
Planning and assessment of services	30.8	50.0	0.21	0.360
Survey, market research and product testing planning	27.5	26.9	0.21	0.379
Experiment planning and interpretation of results	34.6	30.8	0.21	0.377

Graduates make use of skills acquired during their university studies in varying measures depending on whether the job they hold is consistent with the studies they did. Almost all occupation-specific skills used at work were acquired during the university years. School of Letters and Philosophy graduates are an exception, as 34.1% apply business-corporate skills not obtained through their university studies.

Some technical skills are “shared” by graduates of different majors. Graduates from the Schools of Education Science, Political Science and Mathematical, Physical and Natural Sciences share the ability to mediate between subjects, interests and culture. This proves that the frequency with which technical skills are used ties with the profession, context and company size.

It is also evident that the basic skills used by those who hold a technical position could be skills specific to those who hold administrative or marketing positions. For example, linguistic skills, which are taken for granted in any newly hired employee, are specific to those who work in import-export departments or in trade fair or market organization. The basic nature of the skills is also connected with the tasks performed and the operating context.

3. Main factors of skills

We applied factor analysis to search for latent factors (Fabbris, 1997) with the SAS package (SAS Institute Inc., 1994, 2000) on data regarding the use of skills by employed graduates. Various types of analysis were conducted:

- one on the use of seven cross-occupational or basic skills common to all University of Padua graduates;
- one for each school, concerning occupation-specific skills.

The factor loadings found with the analysis of non-specific skills are shown in Table 3. The solution shows the presence of a main factor, which grades the presence *versus* the absence of basic cross-occupational skills used by graduates.

The second factor compares two different types of non-specific skills. On the one hand there is the professional knowledge of languages, by now expected as a requisite for most jobs. On the other hand, there are cross-occupational skills, such as the ability to organize the work of others and perform teamwork.

Table 3. Factor loadings and common variables for cross-occupational and basic skills

Skills	*Factor loadings*		*Communality*
	I factor	*II factor*	
Working in team	0.180	0.453	0.380
PC use	0.237	0.128	0.263
Relationships with customers	0.208	0.119	0.205
Read and speck in a foreign language	0.307	-0.500	0.701
Organising other people's work	0.241	0.471	0.512
Reading and writing in English	0.338	-0.433	0.716
Writing reports	0.285	0.174	0.389

Every sample unit corresponds to two scores, one concerning the "use of specific skills" factor, and the other to the use of non-specific skills. The average score by school is an estimate of their level of use by employed graduates (Table 4).

The correlation coefficient between the two factor scores for Paduan graduates is 0.38. Clearly, a high correlation indicates that when performing their job, graduates use cross-occupational skills jointly with technical skills, whilst

Table 4. Average score value on factors combining cross-occupational skills and correlation coefficients with the occupation-specific factor, by faculty

Faculty	*Average score on cross-occupational skills*	*Correlation between technical-specific and cross-occupational scores*
Agriculture	0.144	0.36
Economics	0.004	0.50
Pharmacy	-0.194	0.46
Law	-0.060	0.50
Engineering	0.260	0.20
Letters and Philosophy	-0.009	0.48
Veterinary medicine	0.148	0.28
Psychology	-0.266	0.47
Education Science	-0.199	0.49
Sciences (Math, …)	-0.051	0.48
Political Science	0.068	0.34
Statistics	-0.295	0.50
Padua University	0.000	0.38

graduates who do not use cross-occupational skills also do not use technical skills. This shows that the work activities performed by graduates are not so characterized by the type of skills used as by the intensity of the use of skills.

Engineering graduates recorded the lowest correlation between the intensity of use of technical and cross-occupational skills. In fact, engineers hold heterogeneous professional positions, from those who require the use of organizational-management skills where the approach to work activities is less technical in the traditional engineering sense, to those that require complex occupation-specific skills, such as those necessary for planning and research, in which few cross-occupational skills are used. The highest correlation was recorded for activities in the social and economics sectors, that is, for graduates from the schools of Economics, Law and Statistics (50%).

4. Jobs on skill maps

A number of 118 professions was recorded among working graduates, some of which, being specific or with low market demand, are filled by a meagre amount of graduates. A factor analysis (Fabbris, 1997) was performed on the usage of seven cross-occupational and basic skills. The factorial solution is described in Figure 1 and in Table 5.

The graph juxtaposes (along the first axis, vertical) an area of usage versus the non-usage of skills and (along the horizontal factor) of skills requiring public relations *versus* those that envisage activities performed individually.

Table 5. Factor coordinates of active variables that determined the solution of analysis of correspondences and some supplementary variables

	Code	F(1)	F(2)	*Code*	F(1)	F(2)
Active variables						
Computer skill	Yes=c1	-0,26	-0,01	No=c1_n	0,88	0,01
English skill	Yes =c3	-0,71	0,37	No=c3_n	0,73	-0,39
Other languages	Yes =c4	-1,02	0,87	No=c4_n	0,39	-0,34
Relationship with customers	Yes =c5	-0,20	0,05	No=c5_n	1,01	-0,27
Working in team	Yes =c6	-0,16	-0,14	No=c6_n	0,87	0,78
Organising others' work	Yes =c7	-0,37	-0,44	No=c7_n	0,69	0,82
Writing reports	Yes =c9	-0,48	-0,42	No=c9_n	0,56	0,49
Supplementary variables						
Company size	Artig.	0,17	0,07	Piccola	0,05	-0,05
	Media	-0,11	0,01	Grande	-0,23	-0,02
Job	a2	-0,16	-0,20	a3	0,36	-0,06
	a4	0,03	0,42	a5	1,12	0,47
Income	<600	0,64	0,36	600-1000	0,22	0,11
	1000-1600	-0,10	-0,08	>1600	-0,46	-0,09
Satisfaction for own job	s1	0,49	0,18	s2	0,09	0,06
	s3	-0,17	-0,09	s4	-0,39	-0,06
Value professional skills	Val	-0,08	-0,05	No_Val	0,56	0,36
Use methods and techniques	Mt	-0,16	-013	No_mt	0,21	0,16
Forma mentis	Fm	-0,08	-0,08	No_fm	0,37	0,35
Work environment	S	0,30	0,30	Eq_u	-0,03	-0,11
				Eq_d	-0,17	-011
Faculty	Fac01	-0,18	0,06	Fac02	0,07	0,44
	Fac03	0,05	0,13	Fac04	0,14	0,08
	Fac05	-0,31	-0,42	Fac06	0,01	0,97
	Fac08	0,19	0,46	Fac09	0,30	0,24
	Fac10	0,57	-0,24	Fac11	-0,01	-0,17
	Fac12	-0,05	0,13	Fac13	0,70	0,07
Substitutability of degree	laurea_spec	-0,11	-0,16	Diploma	0,41	0,31
				Other degree	-0,07	0,02

Setting the professions on coordinated axes (Figure 2, Table 6), four sections can be identified:

- *the first section* (top right), characterized by a meagre use of skills and interaction with customers, work carried out independently, low starting salaries, irregular labour contracts or precarious status, and low job satisfaction. Professions are: teacher and tutor of pre-university educational institutions, sales clerk, pharmaceuticals representative, journalist, analytical psychologist and production and logistics head with subordinate duties and activities not related to university studies. These graduates are predominantly from the schools of Humanities, Psychology, Pharmacy, Veterinary Medicine and also from the schools of Law and Statistics;

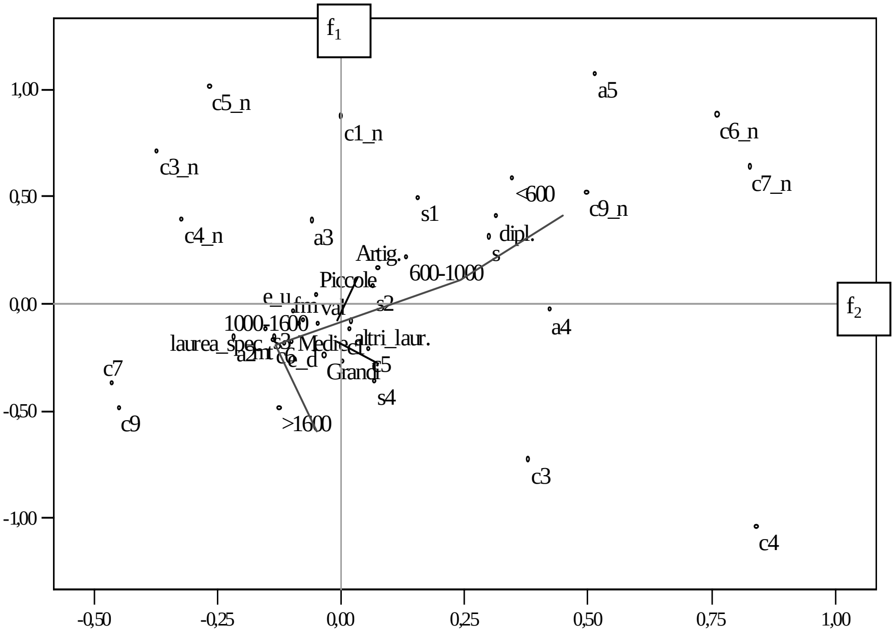


Figure 1. Active and illustrative variables over the layout of correspondence analysis

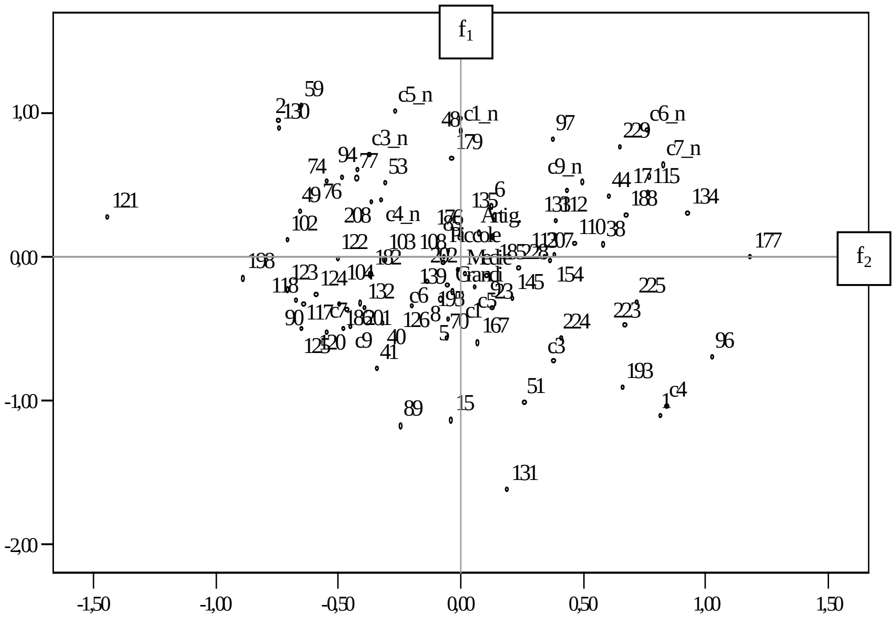


Figure 2. Active modes and professions represented on the resulting graph from the analysis of correspondences

Table 6. Factorial coordinates of professions

	Profession	D1	D2		Profession	D1	D2
1	Purchases	-0.75	0.82	118	Civil engineer	-0.28	-0.65
2	Insurance agent	0.95	-0.73	120	Telecommunications engineer	-0.48	-0.46
5	Agronomist	-0.40	-0.03	121	Construction engineer	0.28	-1.43
6	Finance administration	0.38	0.12	122	Electric engineer	0.00	-0.48
8	Financial analyst	-0.10	-0.15	123	Electronics engineer	-0.33	-0.62
9	Analyst programmer	-0.23	-0.02	124	Management engineer	-0.21	-0.57
15	Quality controller	-1.11	0.00	125	Computer engineer	-0.60	-0.54
17	Production assistant	0.60	0.77	126	Mechanical engineer	-0.31	-0.15
23	Lawyer, legal practitio-ner	-0.31	0.14	130	Electronics engineer-ing technician	0.71	-0.79
38	ISO quality certifier	-0.03	0.47	131	Computer engineering technician	-0.74	-0.45
40	Chemist	-0.43	-0.20	132	Mechanical engineer-ing technician	-0.41	-0.38
41	Industrial chemist	-0.75	-0.31	133	Instructor	0.50	0.44
44	Clerk, sales clerk	0.47	0.60	134	Primary school teacher	0.34	0.97
48	General consultant	1.05	0.06	135	Second. school teacher	0.32	0.16
49	Labour consultant	0.21	-0.87	139	Chemical analyst	0.03	-0.09
51	Financial consultant	-0.97	0.31	145	Marketing analyst	-0.26	0.23
53	Legal consultant	0.53	-0.30	154	Veterinary small anim.	0.02	0.38
59	Management control	1.08	-0.65	167	Training operator/org.	-0.56	0.11
70	Technical-scientific documentation	-0.53	-0.04	176	Prod. planning, pur-chases/sales	0.49	0.42
74	Educator	0.55	-0.55	177	Customer care	0.04	1.21
76	Pre-school teacher	0.58	-0.47	179	Production & logistics	0.45	0.02
77	Community educator	0.57	-0.41	182	Development & design	-0.10	-0.35
85	Pharmacist	0.17	0.00	185	Clinical psychologist	0.17	0.18
89	Applied physicist	-1.15	-0.20	186	Labour psychologist	-0.30	-0.49
90	Electronics physicist	-0.31	-0.63	188	Neuropsychologist	0.33	0.70
94	Trainer	0.63	-0.42	193	Public relations	-0.86	0.71
96	Trainer-counsellor	-0.66	1.08	195	Quality controller	-0.54	-0.11
97	Tutor	0.84	0.39	198	Laboratory head	-0.13	-0.89
102	Geologist	0.14	-0.71	201	R & D	-0.30	-0.39
103	Environment geologist	0.01	-0.29	202	Researcher	-0.10	0.06
104	Geologist -construction	-0.10	-0.32	207	Secretary & staff	-0.03	0.36
108	Human res. manager	0.03	-0.06	208	Personnel selection	0.42	-0.37
110	Journalist, press-room	0.14	0.46	223	Marketing	-0.43	0.70
112	Admin. clerk/officer	0.33	0.36	224	Quality assessment	-0.06	0.62
113	Paralegal	-0.19	0.10	225	Sales management	-0.27	0.75
115	Pharmaceut. promoter	0.47	0.80	228	Other, related major	-0.02	0.27
117	Chemical engineer	-0.52	-0.65	229	Other, unrelated	0.75	0.64

- *the second section* (top left) has to do with jobs that do not involve interaction with customers. These types of occupations include that of corporate, legal or labour consultant, community and pre-school teacher, insurance agent, personnel selector, company administration management control head, building designer, electronics technician and professional geologist;
- *the third section* (bottom left) consists of positions such as laboratory head, research and development and quality control technician and designer, almost all types of engineers and engineering technicians, applied physicist, chemist, computer engineer, scientific document collector, quality assurance inspector, labour psychologist, financial analyst, corporate finance and administration head, agronomist and field geologist. These jobs are characterized by high salaries, the feeling that the worker's professional skills are appreciated in the workplace, daily use of cross-occupational skills except for linguistics skills, and by *forma mentis*, teamwork, often together with experts of various professions in fairly sizeable companies that may overtop a workforce of 250;
- *the fourth section* (bottom right) is comprised of occupations with a high use of skills geared toward clientele and toward international contacts (linguistic, dialectic and relational). They involve purchasing, marketing and post-sales (customer care), public relations, guidance, certification and quality assurance, secretarial work and administration staff, financial consulting, training course organization, law firms, social and economic research, as well as other professions generically relevant to the major studied. Performing the job gives a certain satisfaction in monetary terms, not always accompanied by as much professional satisfaction. This is the area with the most graduates with degrees in Economics, Political Science and Agriculture.

5. Multilevel analysis of use of skills

In order to study the variables associated to the use of skills shorn of the influence of the degree, the relationship between the factor scores in Section 3 can be examined from a multilevel standpoint. The analysis consists of an estimate of a multilevel linear regression model in which the first level units, the graduates, are embedded within the second level units, the faculties:

$$y_{ij} = \gamma_{00} + \sum_{h=1}^{H} \gamma_{h0} x_{hij} + \sum_{k=1}^{K} \gamma_{0k} z_{kj} + U_{0j} + R_{ij}$$

Factor scores on the first cross-occupational factor and on the first specific factor are dependent variables. There being no second level explanatory vari-

Table 7. Multilevel regression models on factor scores calculated over the first cross-occupation factor and on the first occupation-specific factor (* p=0.15)

Variables	Cross-occupation Factor			Occupation Specificity Factor		
	γ_{h0}	$s(\gamma_{h0})$	p	γ_{h0}	$s(\gamma_{h0})$	p
Intercept	-0.59	0.48	0.237	-0.49	0.71	0.497
Manager	0.17*	0.10	0.087	0.30*	0.16	0.059
Regulated profession	-0.05	0.10	0.642	0.09	0.16	0.575
Atipical contract	0.06	0.06	0.358	-0.06	0.10	0.532
Other	-0.16*	0.11	0.142	-0.02	0.17	0.898
Private sector	-0.01	0.05	0.886	-0.14*	0.09	0.109
Income >1100 Euro	0.20*	0.09	0.030	0.08	0.15	0.588
Income between 850 and 1100 €	0.07	0.07	0.376	0.06	0.12	0.607
Unemployed at graduation	-0.01	0.04	0.804	0.07	0.07	0.298
Firm 50-250 workers	-0.02	0.05	0.765	0.02	0.09	0.814
Firm 250-1000 workers	0.10*	0.06	0.091	-0.06	0.09	0.523
Firm more than 1000 workers	0.21*	0.10	0.035	-0.06	0.16	0.698
Age at graduation	-0.01	0.01	0.159	-0.01	0.01	0.511
Graduation mark	0.01*	0.01	0.029	0.01*	0.01	0.140

ables, the variability between academic degrees is attributed to the intersection's random component. The results[3] are in Table 7.

Significant predictors in the model in which the dependent variable is the cross-occupational skills score are the professional position, the salary, the size of the company and the graduation mark. The use of these skills is greater when the company is larger and the higher the grade-mark, the higher it is for executives and the lower for blue-collar workers, apprentices and merchants (all from the 'other' category) with respect to employees.

As regards regression on occupation-specific skills scores, the multilevel model perforce yields the same results as a simple linear regression: in fact, since scores are normalized in each school, divergence between groups is null. As per fixed effects, the use of technical skills is higher in executives than in clerks and for the more brilliant academic careers.

This analysis confirms the hypothesis that highly qualified jobs involving a high degree of responsibility require greater use of both cross-occupational and specific skills. It comes as no surprise that company size is important, since in companies with a high workforce the tasks of a recent graduate are technical but also orientated toward management and organization.

---

[3] The SAS (SAS Institute Inc., 2000) MIXED Procedure, which allows for the estimate of linear hierarchy models with a method of the closest reflection of reality (Wolfinger, 1999), was used for statistics analysis. The predictors' levels of reference are: clerk/staff member as the average professional position; salaries under 850 euros as the average pay; less than 50 employees as the average company size.

6. Conclusions

We found that cross-occupational skills are used massively in the workplace, more so than basic skills. Indeed, jobs requiring computer use and knowledge of the English language, considered necessary requirements for hiring, as these skills are overall used less than cross-occupational skills.

A person hired to perform a professional activity needs skills specific to the sector. These skills are acquired almost exclusively through higher studies, at university level, in particular. The intensity with which graduates use them when working depends on the specific position they hold. Nevertheless, the speed with which technology, techniques and the skills necessary for work are to be updated requires the development of the capacity to learn throughout one's entire working life with an outlook toward constant updating. The basic skill of workers is therefore the capacity to learn.

Indeed, jobs and professional positions may also vary considerably depending on company size. Graduates working in smaller companies must be more versatile and more willing to cover a wide spectrum of tasks. The larger the company, the more probable it becomes that the type of service rendered will be exclusively of a technical nature. Therefore, not only the level of use of skill varies, but also the mix necessary for different types of companies.

The possibility of associating to each graduate values representative of the intensity of use of occupation-specific, cross-occupational and basic skills to a graduate is analysed in the note. Thus, a sample of employed graduates was used to measure their use of skills.

The solution of estimating the intensity of the use of specific skills separately from cross-occupational skills and cross-occupational-strategic skills results in the definition of skills specific to a particular profession. If they are specific, these skills are absent in most of the other professions. The method of estimating the technical level of a graduate's work activities is therefore heavy under a computational standpoint and sparing in general indications.

The quantification of the intensity of use of specific, cross-occupational and basic skills by each graduate is nevertheless the technical condition used to qualify each position or profession according to the technical and cross-occupational levels of the skills used. The types and levels of professional qualification to associate with educational profiles emerging from university careers can be discerned by mapping these skills.

The intensity of use of cross-occupational skills is discriminating from a social and professional standpoint. On one side of the spectrum, there are professions openly deemed as satisfying, consistent with the university career graduates invested in, and that make them practically irreplaceable in their position. On the other side, there are more general jobs characterized by a high graduates' offer and by the frustration of those employed regarding the moral and material rewards obtained.

From among the positions that graduates consider the most frustrating, and which we hand over to the attention of university study strategists, we extrapolate the teacher.

Another profession that merits the attention of academics is that of administrative staff. It is located in the area of little use of occupation-specific and cross-occupational skills. However, it is better positioned, in terms of competence use, than numerous other specific occupations (cultural tourism, professions involving marketing and post-sales services, and even professions such as microbiologist and some engineering positions).

This corroborates the idea that numerous "generalist" study programmes that converge in this position can be enhanced in the future, given the demand for more technical training for this type of professional.

References

BEFANI B. (2004) Il dibattito sulle competenze dalla teoria alla base di dati O*Net. In: CIVARDI M. (ed) *Transizione Università-Lavoro: la definizione delle competenze,* Cleup, Padova: 19-43.

DEL FAVERO M., FABBRIS L. (2002) *Rappresentazione dei mercati locali del lavoro: il sistema di indicatori DOMUS*, Quaderno PHAROS n. 1/2002, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova.

FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.

FABBRIS L. (2003) *Il monitoraggio dell'inserimento occupazionale e professionale dei laureati e dei diplomati dell'Università di Padova*, Quaderno PHAROS n. 6/2003, Osservatorio sul mercato locale del lavoro dell'Università di Padova, Progetto FORCES, Cleup, Padova.

FABBRIS L., VISENTIN E. (2005) Tessere di competenze per mosaici di attività in aziende di varia dimensione. In: D'OVIDIO F. (ed) *Professioni e competenze nel lavoro dei laureati*, Cleup, Padova: 21-44.

GOLDSTEIN H. (1995) *Multilevel Statistical Models*, Arnold, London/ Halsted, New York.

ISFOL (1998) *Unità capitalizzabili e crediti formativi. I repertori sperimentali*, Franco Angeli, Milano.

LEVY F., MURNANE R.J. (1999) Are there key competencies critical to economic success? An economic perspective (contribution to project *Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo).*

SAS Institute Inc. (2000) *SAS/STAT Software: Changes and Enhancements, Release 8.1*. SAS Institute Inc., Cary, NC.

SPENCER L.M., SPENCER M.S. (1993) *Competence at Work. Model for Superior Performance,* Wiley, New York.

WOLFINGER R. D. (1999) *Fitting Nonlinear Mixed Models with the New NLMIXED Procedure,* SAS Institute Inc. Cary, NC.

Competences Offered to Statisticians by the Italian Universities and Required by the Job Market

Enrica Aureli*, Domenica Fioredistella Iezzi**

* *Department of Statistics, Probability and Applied Statistics, University "La Sapienza" of Rome, Italy*
** *Department of Economical and Financial Studies and Quantitative Methods, University of Rome "Tor Vergata" , Italy*

Summary. The development of the European higher education system, within the frame of the Bologna process, requires that higher education be more and more integrated with the wider economic strategies (Lisbon objectives) and that employability be taken into account by the study programmes. Higher education institutions are invited not only to answer to the needs of the national labour markets but also to incorporate into their perspectives the European one. The Italian university reform provides for a first triennial cycle with educational and training purposes that give graduates an immediate access to the labour market. The labour market requires, on the other hand, graduates to have not only a theoretical education, but also professional competences. This paper aims to build cognitive maps of the study programmes in Statistics and rank the competences required by the job market to graduates in Statistics. We apply for that various methods of multivariate and textual analysis.
Keywords: Competences; Statistics; Programme statements; K-means cluster analysis; Multiple correspondence analysis; Rasch analysis; Textual analysis.

1. Competences and the Italian University reform

The reform of the Italian university system (decree of 3 November 1999, no. 509) has radically transformed the formative objectives of the study programmes (Aureli & Iezzi, 2004, 2005; Iezzi, 2005). Since 2001, the Italian university programmes are organised on 3 cycles: Three-year cycle (CDL1 = Bachelor); Two-year cycle (CDL2 = Master); Third cycle of variable length (CDL3 = Ph.D.).

The CDL1 aims at guaranteeing undergraduate students with an adequate command of general scientific concepts and methods, as well as professional skills. General admission requirement is the Italian high school certificate awarded to those who pass a national exam, after completing 13 years of schooling; foreign comparable certifications may be accepted as well. Admission to individual degree programmes may be subject to specific requirements. First level degrees are awarded to students who have earned 180 credits.

The CDL2 includes highly qualified activity in specific areas. Access to the second level is through the Italian CDL1 or a foreign comparable degree. This degree is awarded to students who have earned a global amount of 300 credits, including a maximum of 180 of the first level that have been recognised for access to the CDL2. The drawing up of an original dissertation is compulsory.

A limited number of programmes regulated by EU directives (in dentistry, human medicine, veterinary medicine, pharmacy) share the following different features: access is free to those who have an Italian high school graduation diploma or a foreign comparable certification. Admission is always subject to an admission exam; normal length is 5 years (human medicine takes 6 years).

The postgraduate studies consist of the following typologies of degree programmes:

A) Ph.D. programmes; B) 2nd level specialisation programmes;
C) 2nd level university master degree programmes.

Competence, or *competency* (Spencer & Spencer, 1993) is “an underlying characteristic of an individual that is causally related to criterion-referenced effective and/or superior performance”. In our case, a competence may be defined "an enduring characteristics of an individual that causes and foresees his/her work-hunting and professional behaviour".

Competences are a mix of knowledge, skills and attitudes that make it possible graduates to be successful in their interactions with others at work, school, home, and in the community at large. To match the university educational offers and the job market requisites we assume the so-called ‘iceberg model’ of competences (Spencer & Spencer, 1993, see Table 1).

The concept of competence became influential initially in business organisations, more specifically in the field of recruiting and selecting new employees. The McClelland’s (1973) goal of “testing for competence rather than for intelligence” is the starting point of the competence movement (Barrett & De-

Table 1. Iceberg model of competences

Measurement	Performance	Skills Knowledge	Behaviour
Hidden ***Guess***	Competence Ability	Self-concept Stable personal attributes Motives & intentions	Competence Convictions/ beliefs Identity Spirituality

pinet, 1991). McClelland argued that commonly used I.Q. and personality tests were poor predictors of successful performance, and that competence assessment had to be developed instead. He suggested the Behaviour Event Interview (BEI), an interview method that combines Flanagan's Critical Incident Method (Dailey, 1971; Boyatzis, 1982) with the Thematic Apperception Test (McClelland, 1989).

The BEI is often applied to discover differences between a person, who has been selected by knowledgeable judges as outstanding, and a reference person. The underlying assumption is that it is less difficult to deciding who is competent than what makes them competent (McClelland, 1998). Starting from several studies conducted in various types of organisations, Spencer & Spencer (1993) wrote a dictionary of competences that distinguishes superior from average performers in middle to upper-level jobs.

Consoli & Benadusi (1999) point out that the concept of competence emerges together with that of learning, formative credits, project organisation, continuous evaluation. There is an ongoing national and international debate (Ajello, 2002; Boyatzis, 1982; Capaldo *et al.*, 1996; Carretta *et al.*, 1992; Civelli & Manara 1997; Cerase, 2002; Consoli, 1998; Consoli & Benadusi, 1999; Fabbris, 2004; Iezzi, 2003; Zan, 1988) on competences required by the job market from new graduates.

The aim of this paper is to verify points of contact between the university offer and the demand of the job market. The structure of the paper is as follows: in Section 2 we present the data at hand; in Section 3 the results of programme statement analysis; in Section 4 the ranking of competences required by job market from new graduate in Statistics, and in Section 5 the conclusions of our analysis.

We will use *partial credit* method (Wright & Masters, 1982) to rank the competences required from graduates. Partial credit is a Rasch model that transforms categorical data in continuous data, providing a complete solution to almost every measurement problem encountered in social science studies.

2. The data

Our study examines the Italian university system and the new role of competences in CDL1. In this way, we can match the competences achievable through university studies and those required by the labour market.

The functional aspect of this new system favours the development of skills rather than just knowledge as it happened in Italy in the days before the reform. That is why we tried to verify if and how the university educational offer is shaped according to the labour market perspectives, and which are the skills a graduate should posses to get a job.

We realise our analysis in two steps: in the first, we examine the reformed university system; in the second, we identify the requests of the market. The point of contact between the first and the second phase is the graduate.

In the first phase, we will examine the stated intentions in the programme's arrangements (Ministry decree 4th August 2000). The reform provides for each disciplinary class with a "declaratory", i.e. a programme statement. We will use the 41 statements of "Statistics". Each statement is divided into two parts: the aims of this discipline and the job opportunities for its graduates.

The programmes of the Statistics class represent a small part in comparison with the amount of programmes activated after the reform (1.5% of the triennial programmes). All programmes were organised by five universities: Milan "Bicocca", Padua, Bologna, Rome "La Sapienza", and Messina. We considered the ministerial arrangement as a reference for the 41 programmes.

In 2004, we interviewed 137 people who graduated in Statistics at the University "La Sapienza" between March 2000 and March 2001 through a CATI – *Computer Assisted Telephone Interviewing* survey (Aureli & Ottaviani, 2004). We collected data on graduates' knowledge, technical and "soft" skills and confronted this with their work needs. Knowledge was composed of questions on "*how is it important for your current job*" each one of the 13 following disciplines: Insurance, Demography, Law, Economics, Finance, Operational Research, Computer Science, Mathematics, Probability, Sociology, Economical statistics, and Social statistics.

3. A model to analyse the statements

The first phase of the analysis was the transformation of the verbatim information into a data matrix. We obtained a matrix of 42 rows, 41 CDL1s and the ministerial arrangement. A pre-processing step was required to eliminate meaningless words (Bolasco, 1999). After that, we applied lexical analysis and identified m matrixes one for each of two themes: "programme aims" and "job opportunities".

We have built two different data matrices because we wanted to test each of two analytical strategies:

i) a matrix **T** of $[(n+1) \times p)]$ order; where n is the number of statements, 1 represents the model and p is the number of meaningful words;
ii) a matrix **S** of $[(n+1) \times q)]$ order, where n is the number of statements, 1 is the model and q the number of chosen keywords ($p>q$).

For the first strategy, we applied principal component categorical (PCC) to identify and show the association structures on the table. This statistical technique is useful when it is not possible to explain the relationships among objects because of the number of variables.

The principal factors obtained by PCC showed m text maps and m keywords. We used an integrated approach to classify the texts by starting the results of PCC. We searched the best partition of texts and the words from initial configuration of group centroids.

For the second analytical strategy, we applied multiple correspondence analysis (MCA) to build a map. The **S** matrix was transformed in *m* blocks of matriz $\mathbf{Z} = [Z_1, ..., Z_m]$ with full disjunctive codes. The dimension of **Z** is $[(n+1) \times q)]$. The application of MCA detected the latent factors (Greenacre, 1984).

Even in this case, we applied an integrated approach to classify the texts. We used three different algorithms to identify the number of groups (single linkage, centroid, and Ward' method). We identified three groups with Ward's method and four with single linkage and centroid methods.

Since the university reform started, the Italian Ministry of Education, University and Research organised a web site for reporting all the activities of the Italian Universities (http://offertaformativa.miur.it/corsi/). We searched the information on first and second study levels and acquired the 41 statements related to Statistics.

The wealth of words depends on topic, the descriptive analysis showed a major wealth for "aims" than for "job opportunities" (Table 2). A detailed list of aims means a good specificity of a programme that justifies various educational offers in a given faculty.

Table 2. Descriptive analysis of graphical forms

Key topic	*no. units*	*no. words*	*graphical forms*	*% graphical form*	*most frequent words*
Aim (Model)	1	175	111	63	TO POSSESS
Job opportunities (Model)	1	193	131	68	ACTIVITY
Aim (CDL1)	41	9718	1465	15	STATISTICS
Job opportunities (CDL1)	41	4598	1041	23	FIRM

The more frequent word is "to possess" for job opportunities topic and "activity" for aims. Those words are also in declaratory statements, but the more common graphical form is "firm" for job opportunities and "statistics" for aims.

We noticed that a change of outlook is starting in university statements, both for knowledge taking and technical and "soft" skill training.

The ministerial purpose is the possession of analytical skills: "graduates must possess adequate knowledge and tools to analyse data". The statements promote the professional role of Universities and define a graduate in Statistics as a problem solver. This young professional is able to analyse and manage complex systems (Table 3).

An exploratory analysis of the model shows that job opportunity keywords focus on educational activity, study and learning, while the statements introduce professional aspects. In the last ones, we find a direct connection to job

Table 3. Keyword of study programme aims

MODEL	Number of words	STATEMENTS	Number of words
TO POSSESS	7	STATISTICAL	148
SUITABLE	3	DATA	83
KNOWLEDGE	3	ANALYSIS	78
STATISTICS	3	MANAGEMENT	72
TREATMENT	2	STATISTICS	61
DATA	2	TOOLS	57
STATISTICAL	2	COMPETENCES	54
TOOLS	2	KNOWLEDGE	52
DISCIPLINES	2	STATISTICIAN	42
COMPETENCES	2	SYSTEMS	41

* A star marks the words used both by the model and by statements.

Table 4. Keyword of job opportunities in statements and model

MODEL	Number of words	STATEMENTS	Number of words
ACTIVITY	6	FIRMS	58
STATISTICIAN	4	ACTIVITY	41
EDUCATIONAL	3	STATISTICS	36
STUDY	2	STATISTIC	35
KNOWLEDGES	2	MANAGEMENT	34
TO UNDERSTAND	2	ANALYSIS	32
CASE	2	PROFESSIONAL	28
CAPABILITY	2	CONTROL	25
FIELD	2	SYSTEMS	21
LEARNING	2	SERVICES	19

* A star marks the words used both by model and by statements

market: firms, activity, management, analysis, professional, control, systems and services (Table 4).

The ministerial arrangement is oriented to curriculum vitae, while the statements are focused on job context and the projection of present students into future graduates.

Maps, obtained with ACM, mark latent factors of aims and job opportunities.

The first factor juxtaposes theory and application words. In fact, we observe that on the left hand side there are knowledge, skills and discipline, and on the right one tools, data and systems. The second factor juxtaposes knowledge and

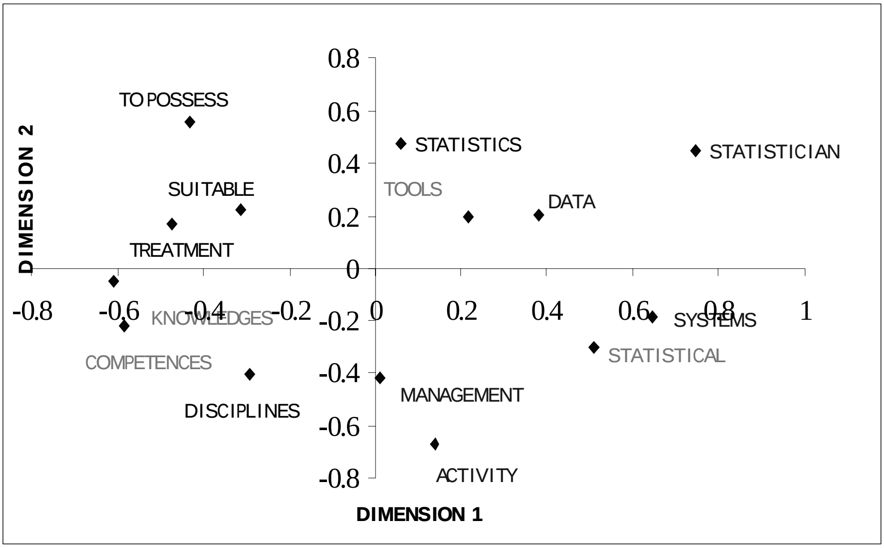


Figure 1. Keyword map of declaratory statement aims

ability; we observe that the verb “to possess” is in the opposite position to management and analysis (Figure 1).

The map of job opportunities shows one principal latent factor that marks prerequisites for work: learning, capability, analysis and management in the private and public sectors. In Figure 2 the two sectors are in an antithetical position. About 60% of the university statements focus on knowledge.

The employment issues classify the CDL1 into four groups that move along a continuum: from public administration oriented programmes, where the words “services”, “study” and “educational” prevail, to market-oriented ones, where the words “firm”, “activity”, and “control” dominate.

Inside this continuum, we have crossed public and private-oriented matter. In the public-private oriented group, we have words like “knowledge”, “professional” and “management”; in the private-public oriented group, we found words like “system”, “capacity” and “analysis” (Table 5).

Based on MCA, we classified the CDL1. Programme aims divided CDL1 into four groups: theoretical, professional, theoretical-professional, professional-theoretical (Table 6).

Within the theoretical group, knowledge dominates all the aims. In the opposite position, we find the professional group that prefers to teach the tools of the subject matter. The groups may be disposed along a *continuum*: it starts with a theoretical approach (“to possess”, “adequate”, “treatment”) and ends up to a professional planning (“tools”, “data” and “statistics”. Along this continuum, we have a theoretical-professional and a professional-theoretical approach in which “knowledge”, “skill”, “ability” and “disciplines” link together.

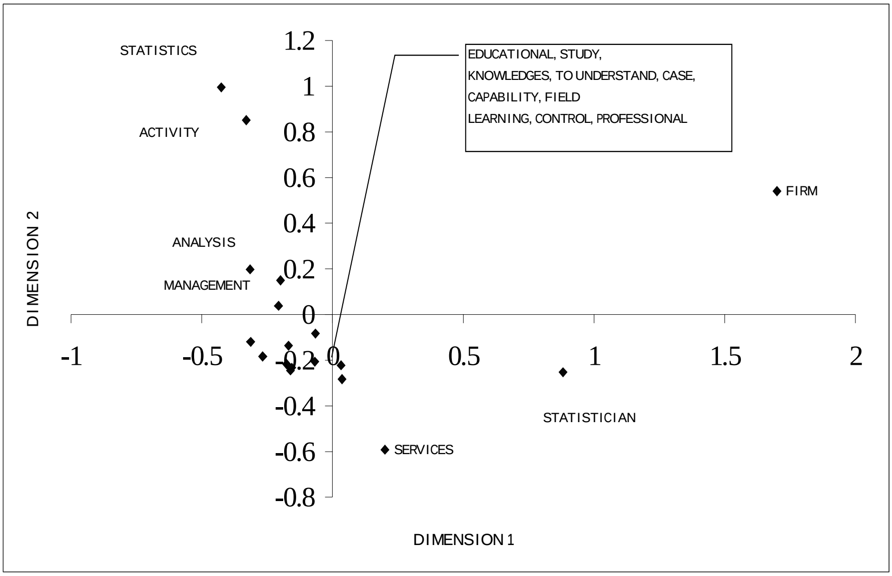


Figure 2. Keyword map of job opportunities

Table 5. Classification of CDL1 based on job opportunities

Key topic	*GROUP*			
	1	***2***	***3***	***4***
CDL1 characteristics	Public oriented	Public-private oriented	Private-public oriented	Private oriented
Keywords	Services Study Educational	Professional Management Knowledge	Systems Capacity Analysis	Firm Activity Control

Table 6. Statement classification of CDL1 based on programme aims

Key topic	*GROUP*			
	1	**2**	**3**	**4**
CDL1 characteristics	Theoretical (cdl=knowledge)	Theoretical-professional (knowledge+skill)	Professional-theoretical (skill+knowledge)	Professional (cdl= skill)
Keywords	To possess Adequate Treatment	Knowledge Competences Discipline	Systems Management Analysis	Tools Data Statistics
Number of CDL1s	7	14	14	7

Moreover, there is cohesion between objectives and job opportunities of the classes of CDL1 (last row of Table 6). The lexical analysis of the *corpus* highlights that it prevails programmes in which theoretical and practical aspects are balanced. However, if we divide the *corpus* into “aims” and “job opportunities”, in the second we can find a clear partition between traditional and modern vision of University.

4. A model for competence ranking

We converted the ordinal data into values applying a Rasch partial credit model (RPCM) in order to analyse the data on knowledge, technical and soft skills collected on ordinal scales. RPCM is an extension of the Rasch one-parameter ‘item response’ model proposed by Rasch (1992) and extended by Wright & Masters (1982), Wright & Mok (2000) and Linacre (2004).

We measured distance and determined the rate of change to define scale units and measures using verbal scale of four items (much, enough, little, not at all). Rasch methods are to obtain objective, fundamental, linear measures from stochastic observations of ordered category responses.

For a given item with *m* score categories, the probability a statistical unit scores *x* on item *i* is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^{x} (B_n - D_{ij})}{\sum_{k=0}^{mi} \exp \sum_{j=0}^{x} (B_n - D_{ij})},$$

where x=0, 1, 2, …, m-1, B_n is the individual competence level and D_{ij} is a step difficulty, and $\sum_{j=0}^{0} (B_n - D_{ij}) \equiv 0$.

The RPCM estimates the probability that a graduate scores *x* on the *m* steps of question *i* as a function of his or her competence level B_n and step difficulty D_{ij} out of the *m* steps in prompt *i*.

We used WINSTEPS program to implement Rasch's JMLE (UCON) (unconditional maximum likelihood, joint maximum likelihood). The JMLE (UCON) method was iterated to obtain more precise estimates, standard errors and fit statistics. We used proportional curve fitting for finding improved estimates.

We examined technical skills through the assessment of 19 items aggregated in 5 classes of topics:

1) writing documents; 2) looking for sources and methods;
3) analysing data; 4) organising and checking trials;
5) speaking a foreign language.

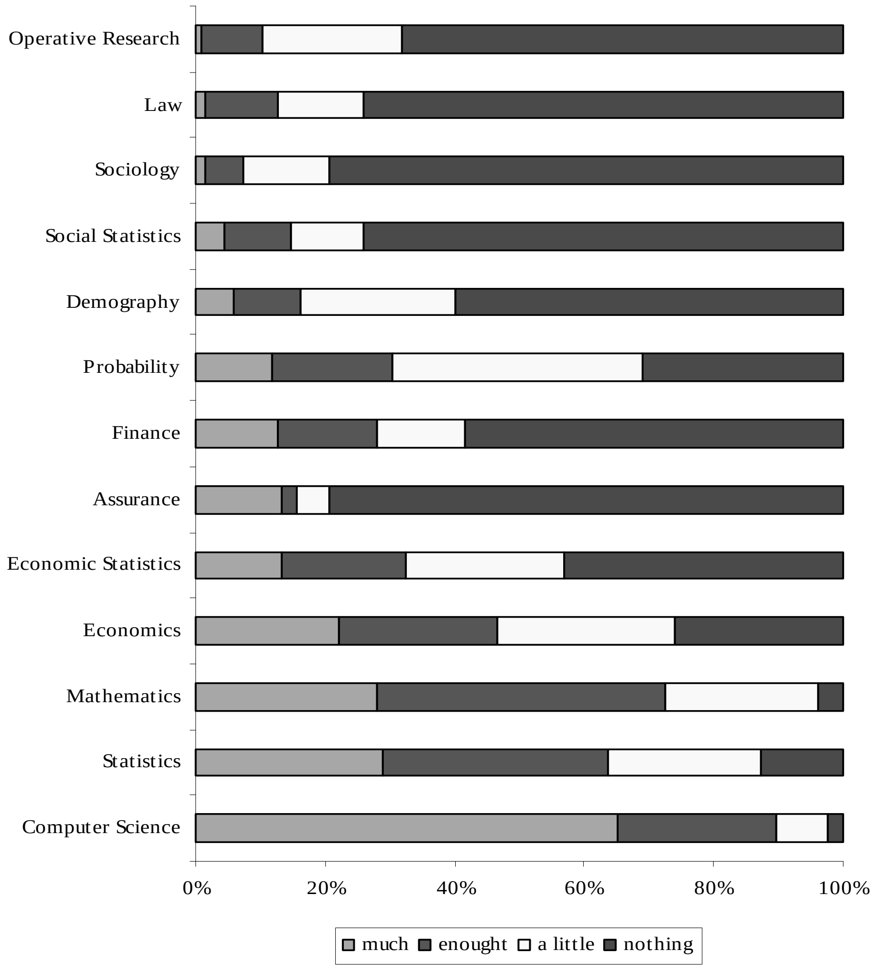


Figure 3. Frequency distribution of the knowledge dimension

We verified soft abilities through the examination of 18 items separated in 5 classes of individual propensity: 1) decision taking; 2) communicating with others; 3) working in a team; 4) setting and solving problems; 5) inclination to long life learning. We have to highlight that there is no clear cut between 'technical' and 'soft' competences.

In knowledge dimension (Figure 3), 60% of graduates' responses (Sociology, Economical Statistics, Social Statistics, Operational Research, Insurance, Law) are concentrated on the negative side.

Technical skills (Figure 4) are more variable than the other dimensions, with the exception of the use of personal computer, which is widespread (Figure 5). Soft ability dimension items are positively oriented (Figure 5).

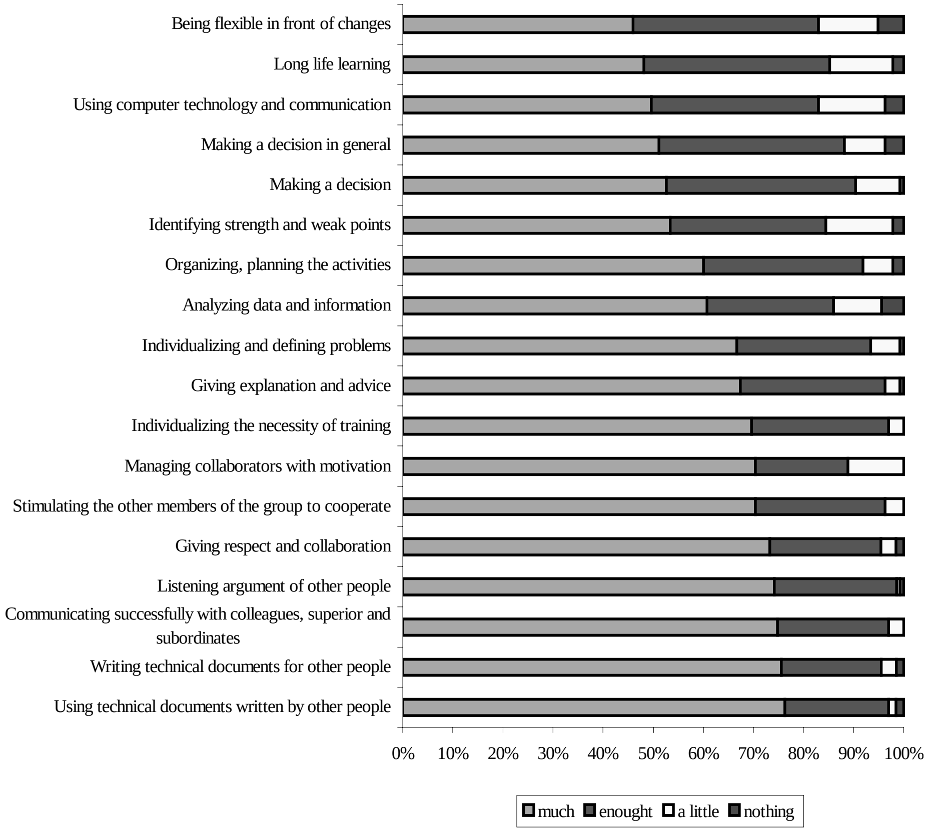


Figure 4. Frequency distribution of the ‘technical’ skills

We used Cronbach’s Alpha (Cronbach, 1951) to measure if and how a set of items for each ‘competence dimension’ (knowledge, technical and soft skills) describe a uni-dimensional latent construct. The resulting reliability is good for the three subsets; in fact, the alpha value equals 0.76 for knowledge items, 0.86 for skill ones and 0.82 for the ability ones.

The knowledge item map (Figure 6) partially confirms the results of the descriptive analysis: the *continuum*’s top includes the least used disciplines (Sociology and Law) and the bottom includes the widely used ones (Computer Science). We can observe some changes of position.

During the initial stages of employment, freshly graduates in Statistics make use of the basic subjects, such as Mathematics, Statistics and Economics, and a little specialist knowledge.

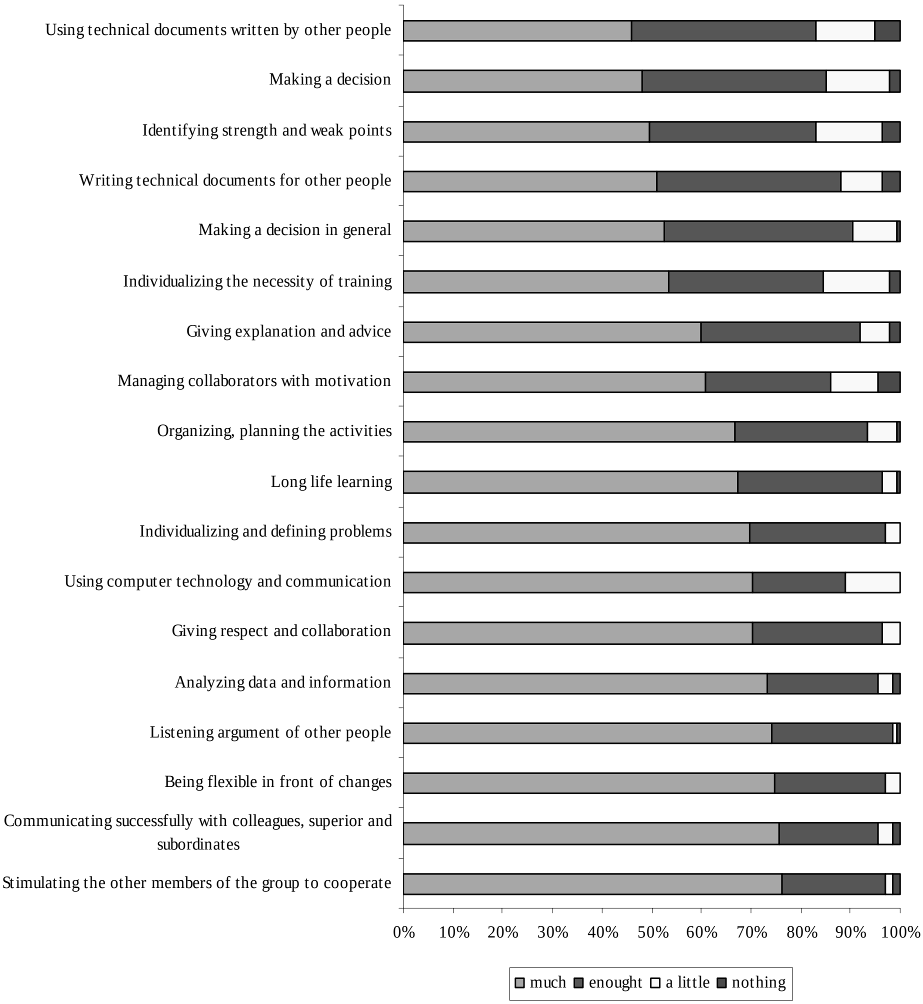


Figure 5. Frequency distribution of the soft abilities

It is worth noticing that some positions, such as Probability, moved and we would not have been able to observe it if we stopped at a descriptive analysis. Moreover, some distances among the disciplines differ.

Computer Science is far away from all other subjects; while Insurance, Social Statistics, Law, Operational Research are close to each other.

Technical skill dimension (Figure 7) is consistent with the knowledge scaling: working with a personal computer is the most important skill used at work by young statisticians, far above any other subject. All other subjects are close to each other: writing technical documents for other people, using technical documents written by other people, using data to evaluate.

```
     <more>||<frequ>
90                ++
                  ||
                  ||
               .  ||
                  ||
                  ||
                  ||
80             .  ++
                  ||
                  ||
               #  ||
                 T||
           .####  ||
                  ||
70           ###  ++T
                  ||
              ##  ||
              .# S||
              ##  ||
              ##  ||  Sociology
                  ||
60          ####  ++S Law
       #########  ||  Operative Research   Social Statistics
           ##### M||  Assurance
             .##  ||  Demography
             .##  ||
         #######  ||  Finance
         .######  ||
50          .###  ++M Economic Statistics
             .## S||  Probability
             .##  ||
              ##  ||
             ###  ||  Economics
                  ||
               .  ||
40               T++S Statistics
                  ||
                  ||  Mathematics
                  ||
                  ||
                  ||
                  ||
30                ++T
                  ||
                  ||  Computer Science
                  ||
                  ||
                  ||
                  ||
20                ++
            <less>||<rare>
```

Figure 6. Map of graduates’ knowledge items

```
<more>|<rare>
   70                +  Analyzing capital markets
                     |  Preparing insurance and financial contracts
                     |S Checking financial balance of firm
   60                +
                  .  |  Studying new statistical methodologies
                     |  Choosing the method to apply
                        Planning the pattern of research
                        Speaking a foreign language
                        Utilizing statistical packages
   50             .  +M Collecting data
                        Looking for sources of statistical data
                        Using data to forecast
                        Using data to manage
                        Using data to plan
                    T|  Organizing data-bank
                        Using data to describe
                        Using data to evaluate
                  #  |  Using technical documents written by other
people
                        Writing technical documents for other people
   40             .  +
            ####### S|S
              .####  |
   30   .########## M+
          #########  |
             .#####  |T
   20           .## S+
                 .#  |
                  #  |
   10               T+
                  .  |  Working with personal computer
                  .  |
    0                +
                     |
                  .  |
  -10                +
               <less>|<frequ>
```

(EACH ' #' IS 3 persons).

Figure 7. Map of graduates’ technical skills

In the cross-occupational ability map (Figure 8), the five areas of the propensity to assume decisions, communicate with others, work in a team, solve problems and be incline to long life learning are very close to each other.

This may mean that during first years of job activity the newly hired graduates in Statistics operate more as technicians than decision makers. This result may depend on the possibility that the survey reached graduates working in research activities together with their dissertation supervisor. These precarious activities are a sort of extension of the final dissertation research activities that graduates accept waiting for job offers.

```
              <more>|<rare>
   80           ###  +
                     |
                     |
                     |
                     |
                     |
               ####  |
   70                T+
                     |
                     |
                     |
               .###  |
                     |
                     |
   60        .##### S+
                     |
               .###  |
                     |
              #####  |
                .##  |
                     |
   50          .### M+
           ########  |
               ####  |
              #####  |
               .###  |
             ######  |
                  #  |
   40           .## S+T
                 ##  |  Using technical documents written by other
people
                  #  |  Identifying strength and weak points
                        Making a decision
                  .  |S Individualizing the necessity of training
                        Writing technical documents for other people
                 .#  |  Making a decision in general
                        Managing collaborators with motivation
                  .  |  Giving explanation and advice
                  .  |M
   30             . T+  Organizing, planning the activities
                        Using computer technology and communication
                     |  Long life learning
                     |  Analyzing data and information
                        Giving respect and collaboration
                        Individualizing and defining problems
                     |S Communicating successfully with colleagues,
superior and subordinates
                     |  Being flexible in front of changes
                        Listening argument of other people
                        Stimulating the other members of the group to
cooperate
                  .  |
                     |T
   20                +
               <less>|<frequ>
```

Figure 8. Cross-occupational item ability map

5. Conclusions

The aim of this study was to check if and how the university offer for higher education in Statistics had assumed the competence approach to outline the educational profiles, and if graduates' curricula and job market's requirements matched. We realised that changes for statisticians' education and employment are slowly taking place.

From the methodological point of view, we highlighted that, within a short document containing a homogeneous and standardized text, it is sufficient to analyse keywords. In fact, sparse matrices gave the same results, but needed much more computational burden than traditional approaches. We put forward the proposal to classify key topics and the entire *corpus* with a consensus algorithm, not to waste resources (Vichi, 1997).

From a substantive point of view, the analysis of the statements showed that, although universities have adopted the ministerial model, they modulate their statements taking into account the local market requirements, and tend to define programme statements where typology, knowledge, skill requirement and ability demand prevail in a functional way. In fact, even if the ministerial arrangement remains adherent to substantive knowledge, the universities propose programmes targeted at the development of cross-occupational competences and other job-specific competences on students.

The interviewing of newly graduated statisticians about competence use at work gave a surprising outcome: the graduates, at the very beginning of their career, apply the knowledge of basic disciplines (mathematics, probability and mathematical statistics) and in very limited cases they use the so-called 'applied statistics'. This is right the opposite of the result of similar studies and of commonsense.

Moreover, the labour market appreciates cross-occupational competences that are not explicit on graduates' curricula. This is open to a dual interpretation: either the companies prefer to shape themselves professional competences, and expect solid and flexible foundations from the educational system, or the universities should develop even students' soft skills, not only the technical-specific ones.

References

AJELLO A.M. (2002) *La competenza,* Il Mulino, Bologna.

AURELI E., IEZZI D.F. (2004) Classificazione di testi omogenei. Il problema delle matrici sparse. In: AURELI E., BOLASCO S. (eds) *Applicazioni di analisi statistica dei dati testuali,* Casa Editrice "La Sapienza", Roma: 55-68.

AURELI E., IEZZI D.F. (2005) Conoscenze, abilità e altre competenze nei profili professionali dei laureati in Scienze Statistiche. In: FABBRIS L. (ed) *Efficacia e-*

sterna della formazione universitaria: il progetto OUTCOMES, CLEUP, Padova, **7**: 129-144.

AURELI E., OTTAVIANI M.G. (2004) Dal disegno concettuale allo strumento operativo: la rilevazione delle competenze degli statistici nel mondo del lavoro, In: D'OVIDIO F. (ed) *Professioni e competenze nel lavoro dei laureati,* CLEUP, Padova, **6**: 45-68.

BARRETT G.V., DEPINET R.L. (1991) A reconsideration of testing for competence rather than intelligence, *American Psychologist,* **46 (10)**: 1012-1024.

BOLASCO S. (1999) *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*, Carocci, Roma.

BOYATZIS R.E. (1982) *The Competent Manager: a Model for Effective Performance,* John Wiley & Sons, New York.

CAPALDO G., VOLPE A., ZOLLO G. (1996) Competenze individuali e sviluppo organizzativo, *Sviluppo & Organizzazione*, **155**: 69-79.

CARRETTA A., DALZIEL M., MITRANI A. (1992) *Dalle risorse umane alle competenze,* Franco Angeli, Milano.

CERASE F.P. (2002) *L'analisi delle competenze nel lavoro amministrativo,* Franco Angeli, Milano.

CIVELLI F., MANARA D. (1997) *Lavorare con le competenze. Come conoscerle, gestirle, valorizzarle*, Guerini e Associati, Milano.

CONSOLI F. (1998) Le ragioni forti della metodologia delle competenze, *Scuola Democratica*, **1-2**: 88-107.

CONSOLI F., BENADUSI L. (1999) L'emergenza della metodologia delle competenze nel pensiero teorico e pratico manageriale, *Osservatorio ISFOL,* **XX(5-6)**: 30-89.

CRONBACH L.J. (1951) Coefficient alpha and the internal structure of tests, *Psychometrika*, **16**: 297-334.

DAILEY C. (1971) *Assessment of Lives,* Jossey-Bass, San Francisco.

FABBRIS L. (2004) *Capitalizzazione delle competenze dei laureati dell'Università di Padova*, Quaderno PHAROS, n. 9, CLEUP, Padova.

GREENACRE M.J. (1984) *Theory and Applications of Correspondence Analysis*, Academic Press, London.

IEZZI D.F. (2005).A new method to measure the quality on teaching evaluation of the university system: the Italian case, *Social Indicators Research*, **73(3)**: 459-477.

IEZZI D.F. (2003) L'analisi delle competenze nel lavoro amministrativo. In. CIVARDI M. (ed) *Transizione Università-lavoro. La definizione delle competenze nella pubblica amministrazione*, CLEUP, Padova: 44-58.

LINACRE J.M. (2004) *Facets Rasch Measurement Computer Program*, Winsteps.com, Chicago.

MC CLELLAND D.C. (1973) Testing for competence rather than 'intelligence', *American Psychologist*, **28**: 1–14.

MC CLELLAND D.C. (1989) *Human Motivation*, Cambridge University Press, Cambridge, UK.

MC CLELLAND D.C. (1998) Identifying competences with behavioral-event interviews, *Psychological Science*, **9**: 331-339.

RASCH G. (1992) *Probabilistic Models for some Intelligence and Attainment Tests,* Mesa Press, Chicago.

SPENCER L.M.Jr., SPENCER S.M. (1993) *Competence at Work: Models for Superior Performance*, John Wiley & Sons, New York.

VICHI M. (1988) Two-way data matrix representative syntheses of a three-way data matrix, *Statistica*, **48**: 91-106.

VICHI M. (1997) Principal Classification Analysis: a method for generating consensus dendrograms and its application to three-way data, *Computational Statistics and Data Analysis*, Springer, **27**: 311-331.

WRIGHT B. D., MASTERS G. N. (1982) *Rating Scale Analysis*, MESA Press, Chicago.

WRIGHT B.D., MOK M. (2000) Rasch models overview, *Journal of Applied Measurement*, **1**: 83-106.

ZAN S. (ed) (1988) *Logiche di azione organizzativa*, Il Mulino, Bologna.

Jobs and Competences of Graduates in Statistics

Maria Cristiana Martini, Luigi Fabbris[1]

Statistics Department, University of Padua, Italy

Summary. The aim of this paper is to analyse the competences used at work and the professional profiles for people who graduated in Statistics at the University of Padua. Our study, accomplished by means of correspondence analysis and cluster analysis, highlights a general consistency between the formative profiles modelled at the Statistics Faculty and the tasks performed by graduates. Moreover, our analysis points out that several graduates are working in clerical jobs that do not allow them to use the competences they achieved during their university studies.
Keywords: Jobs; Competences; Educational profiles; Graduates in Statistics; University of Padua; Correspondence analysis; Cluster analysis.

1. Why an analysis of competences?

In Italy, tertiary education in statistics is traditional at any programme level, bachelor, master and PhD. The placement of graduate statisticians on the job market has always been easy, at least in comparison with other non-technical types of study programmes (Fabbris *et al.*, 2002).

Statistics is not a regulated profession in Italy[2]. Nevertheless, statistical skills are strongly required by the market and statisticians are required for a variety of jobs. Statisticians work in agencies of services supply, craft and

---

[1] This note has been edited by L. Fabbris for Section 1 and by M.C. Martini for the other sections.

[2] In fact, that of actuarial statistician is a regulated profession in Italy, but the Paduan Faculty of Statistics does not supply the specific degree required for the admission to this position. A graduate in Statistics can take an exam to qualify in Statistics for the public administration, but this qualification is not really valued.

industrial companies of any size and activity, agencies for public opinion surveys, central and local public administrations, professional offices, and cooperative societies of services.

The knowledge and skills in demand concern marketing research, information systems management, industrial production and provision of quality control, demographic, social and economic forecasts, different types of counselling, technical support to activities of certification, goods and services quality control, technical offices, research and experimentation for goods and processes development.

However, statisticians are at a disadvantage because the poor knowledge of their professional potential: entrepreneurs and managers perceive the importance of statistics, but are not fully aware of the work duties a statistician can realize. Moreover, the extent of statistical activities rarely requires a “whole” statistician. This means that, in general, the statistical tasks are associated with other tasks and graduates in Statistics are often employed as computer scientists, business economists, and other capacities.

The variety of roles and contexts where the statisticians operate makes it more difficult to identify the professional profile of graduates in Statistics. The functions that a statistician exerts, in fact, are common to almost all the companies and institutions and, with few exceptions, to the functional areas where statisticians are engaged.

The condition of statisticians is common to most qualified workers. In small and medium companies, and in particular in professional offices, graduates are hired to carry out a multiplicity of functions, one of which – hopefully but not necessarily the main – referring to his or her educational subject.

The key to define the work activities carried out by statisticians, and to shape their professional profiles, are the competences they use. Competences are cross-occupation and job-specific kinds of knowledge and skills, and attitudinal features useful to work in a complex working environment, such as the capability to work in team, to feel him/herself part of a whole, to detect and concretely solve company’s problems.

Technical competences can be taught at school. Attitudes are believed to be a sort of personal heritage. Nevertheless, with suitable educational methods, it is possible to develop even professional attitudes. For example, working in team is an attitude that can be instilled by studying in a team, so that students understand the value of a so-called “professional democracy”, the necessity of reaching a shared solution while being able to argue in favour of their own ideas. Analogously, problem-solving abilities can be developed by focusing on problems instead of methods at some points during their university studies.

Competences represent the parameters of a professional profile, its suitability for working purposes. Knowledge, skills and attitudes, whose achievement starts at school, can be used to satisfy work needs. Hence, competences are the elementary units for comparing professional profiles and work needs and, too, for describing the contents of formative curricula and jobs.

In the following, we will describe the competences used by graduates in Statistics (Section 2). We will also try to understand if either the competences used at work are the same, apart from content specificities, or the occupations of statisticians are so culturally and technically different that we cannot define a single professional figure of statistician but distinct ones (Section 3).

We will analyse the answers given by 145 bachelor and master graduates in Statistics at the University of Padua. The graduates were interviewed by telephone six months after graduation (Fabbris, 2003). A number of 99 out of 145 had a job and answered about their work activities.

The use of competences in their working experience has been assessed by means of 23 closed questions. The jobs in which the graduates in Statistics were engaged have been collected by means of a partially closed question based on a list of typical jobs and the possibility to specify the job in a verbatim form. In several cases, instead of directly coding the job, interviewers preferred to record it in a verbatim form.

2. Competences of the Paduan graduates in Statistics

Martini (2003) found that graduates in Statistics find a job in a relatively short time, get a fair income and are fairly satisfied for their job, but their jobs and study curricula are weakly consistent. Hence,

- which competences do graduates really use, and
- what are the ones they feel they miss?

The given answers (Table 1) show that new graduate in Statistics use some specific competences acquired at university, but are employed mostly as computer scientists. It is not surprising that the employed statisticians use social and demographic competences less than the economic ones, since the frequency of economics and business courses overwhelms that of socially oriented courses.

In order to study the relationships between pairs of competences, we built a square symmetric matrix whose general element, for each crossed pair of competences, is the frequency of graduates who reported the use of these competences. This matrix can be considered a complete social network (Scott, 1991) whose nodes are the competences, strengths and the intensities of common use in a given work activity. Ideally, the intensity is inversely proportional to the Euclidean distance between paired competences.

We conducted a hierarchical agglomerative cluster analysis on competences with Ward’s, complete linkage and single linkage methods (Fabbris, 1997). The cophenetic correlation coefficients (Sokal & Rohlf, 1962) were, respectively, 0.83, 0.91 and 0.86. This indicates complete linkage as the most adequate method to represent our data and, consequently, that competences tend to cluster.

Table 1. Percent distribution of competences use levels for people who graduated in Statistics at the University of Padua and work six months after graduation (n=99). Within brackets the label of competences represented in Figure 1

	Not at all	*Scanty*	*Fair*	*High*	*Total*
Use of official data `(sources)`	42.4	20.2	23.2	14.2	100.0
Survey design `(survey)`	57.6	24.3	14.1	4.0	100.0
Sample design `(sample)`	52.1	26.5	14.3	7.1	100.0
Experimental design `(exper)`	52.5	24.2	18.2	5.1	100.0
Experimental results analysis `(exp_res)`	38.4	19.2	27.3	15.1	100.0
Computer programming `(program)`	26.3	10.1	29.3	34.3	100.0
Use of statistical packages `(package)`	49.5	13.1	21.2	16.2	100.0
Simulations `(simulation)`	46.5	24.2	21.2	8.1	100.0
Multivariate analysis `(multiv)`	48.5	19.2	20.2	12.1	100.0
Socio-demographic analysis `(social)`	60.2	16.3	14.3	9.2	100.0
Economic analysis `(economic)`	39.4	14.1	26.3	20.2	100.0
Demographic forecasting `(demography)`	76.5	13.3	7.1	3.1	100.0
Economic forecasting `(forecast)`	49.0	13.3	24.5	13.2	100.0
Social services design `(social_prog)`	76.8	12.1	7.1	4.0	100.0
Budget analysis `(balance)`	43.9	14.3	23.5	18.3	100.0
Management control `(manage)`	43.4	19.2	24.3	13.1	100.0
Service quality evaluation `(quality)`	43.9	21.4	25.5	9.2	100.0
Market research `(research)`	63.7	21.2	11.1	4.0	100.0
Indicators for decision design `(indices)`	39.4	15.1	27.3	18.2	100.0
Writing research reports `(report)`	40.2	15.5	29.9	14.4	100.0
Internal communication `(experts)`	31.3	18.2	25.3	25.3	100.0
Use of multi-media tools `(audiovis)`	36.4	28.3	15.1	20.2	100.0
English and other languages `(language)`	34.3	28.3	23.2	14.2	100.0

The analysis of the resulting dendrogram (Figure 1) highlights three groups of competences:

- one related to economics and management curricula,
- another to socio-demographic curricula
- an independent group of basic competences.

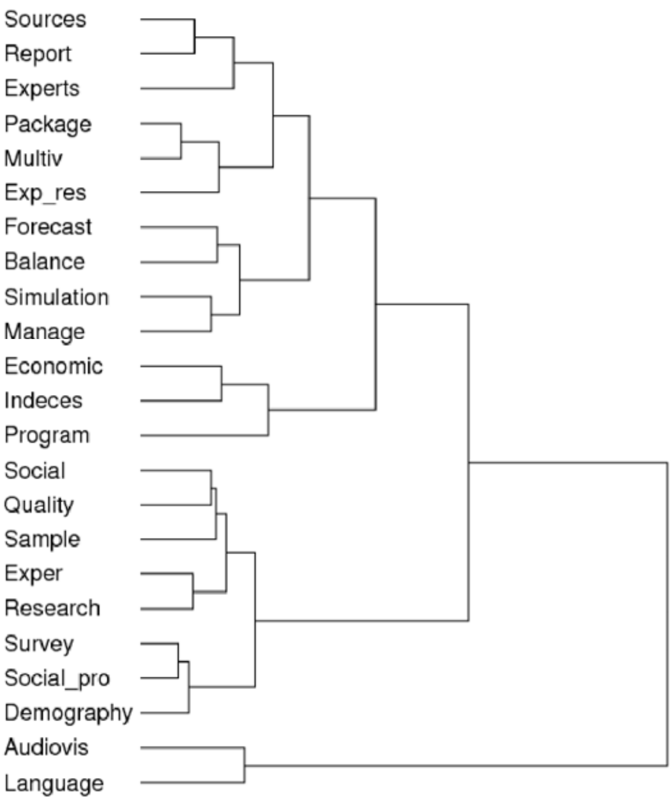


Figure 1. Dendrogram from the cluster analysis on competences (between-competences Euclidean distance; complete linkage analysis)

3. Jobs as combinations of competences

The long list of collected professional figures made it necessary to aggregate the jobs of statisticians into 11 categories (Figure 2). It is clear that non-statistical jobs prevail among those graduates. They relate mainly to computer science, general clerical works, teaching, and to a residual category that cannot be logically associated to any of the other categories (among the others, waiter, call-centre operator, building surveyor, worker, but also consultant, census coordinator, and so on).

Either such a large number of graduates with vaguely consistent jobs may depend on the fact that some of them already worked before graduation, and remained in the same position after it, or that other graduates, as newcomers, were assigned a marginal professional position.

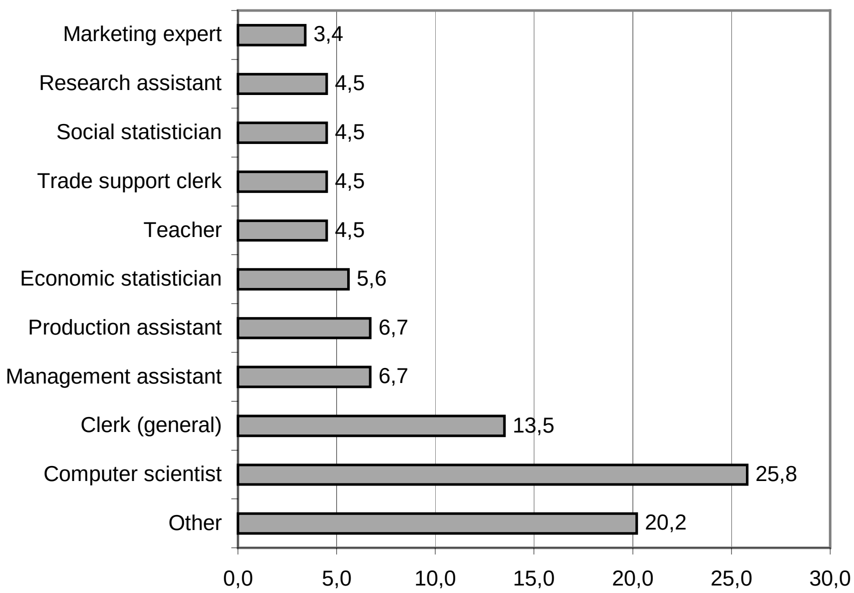


Figure 2. Percent distribution of Paduan graduates in Statistics jobs

In order to define the relations between competences and jobs, we carried out a correspondence analysis (Benzécri, 1979) on the 23 dichotomized competences[3]. Then, we projected over the factorial axes the identified jobs and some general characteristics of graduates (gender, type of programme) and of their work (satisfaction for job, consistency between job and curriculum, use of *forma mentis*, work status before graduation).

The data were processed with CORRESP procedure of package SAS (SAS Institute, 1994, 2000). In Table 2, we report the factor coordinates for each answer category of the active variables, and in Table 3 the coordinates of the supplementary variables. The scree test shows the existence of a unique factor[4]. Nevertheless, to ease the analysis, a second dimensions is represented in Figure 3.

The first dimension represents the general use of competences acquired during university studies, and shows juxtaposition between the use and non-use of competences. The main contributions for the use of competences are given by the use of official data, socio-demographic and economic analyses, economic forecasts, research reports, and decisional indicators, all competences that suggest qualified professional activities. Opposite we see the non-use of spe-

[3] In order to avoid the unwanted influence of the ordinal items on correspondence analysis, the answers on competences have been dichotomized into “Use” (“Fair” + “High”) and “Non use” (“Not at all” + “Scanty”).

[4] The first two dimensions explain 81.6% and 11.9% of Benzécri (1979) adjusted inertia.

Table 2. Factor coordinates on competences used by Paduan Graduates in Statistics

	Coordinates			*Coordinates*	
	1	*2*		*1*	*2*
Use of official data sources	0.77	0.03	*No demographic forecast*	-0.10	-0.19
No use official sources	-0.46	-0.02	*Economic forecasting*	0.77	-0.54
Survey design	0.86	1.27	*No economic forecasting*	-0.48	0.33
No survey design	-0.18	-0.27	*Social services design*	0.81	1.74
Sample design	0.63	0.58	*No social services design*	-0.10	-0.21
No sample design	-0.17	-0.15	*Budget analysis*	0.63	.0.61
Experimental design	0.77	0.28	*No budget analysis*	-0.43	0.42
No experimental design	-0.24	-0.09	*Management control*	0.64	-0.34
Experiment analysis	0.60	0.10	*No management control*	-0.40	0.21
No experiment analysis	-0.47	-0.07	*Service quality control*	0.65	0.40
Computer programming	0.52	-0.17	*No services control*	-0.30	-0.19
No computer programming	-0.91	0.30	*Market research*	0.75	-1.20
Use statistical packages	0.64	0.40	*No market research*	-0.11	0.18
No use statistical packages	-0.40	-0.25	*Decisional indicators*	0.68	-0.11
Simulations	0.64	-0.71	*No decisional indicators*	-0.56	0.09
No simulation	-0.27	0.30	*Research reporting*	0.77	0.15
Multivariate analysis	0.67	0.31	*No research reporting*	-0.55	-0.11
No multivariate analysis	-0.33	-0.15	*Internal communication*	0.54	.0.09
Social data analysis	0.94	0.83	*No inter. communication*	-0.53	0.09
No social data analysis	-0.28	-0.25	*Multimedia tools use*	0.70	-0.13
Economic analysis	0.69	-0.45	*No use multimedia tools*	-0.33	0.06
No economic analysis	-0.59	0.38	*Foreign languages*	0.60	0.12
Demographic forecasting	0.95	1.77	*No foreign languages*	-0.36	-0.07

cific competences and especially the non-use of computer skills. This may be considered the negative pole of this dimension. Because the non-use of computer skills is at the extreme of the non-use of statistical competences, we can say that computer skill is the professional *minimum quantum* for a graduate in Statistics.

The second dimension contrasts the socio-demographic and economic kinds of competences. Along this axis, the main contributions come from survey design and social services design, demographic forecasts and socio-demographic analyses on the one side, and marketing research, simulations and budget analyses on the other.

We observe that the activities of clerks and teachers lie in the area of non-use of specific competences, together with the residual category. Graduates employed as computer experts do not use but basic competences.

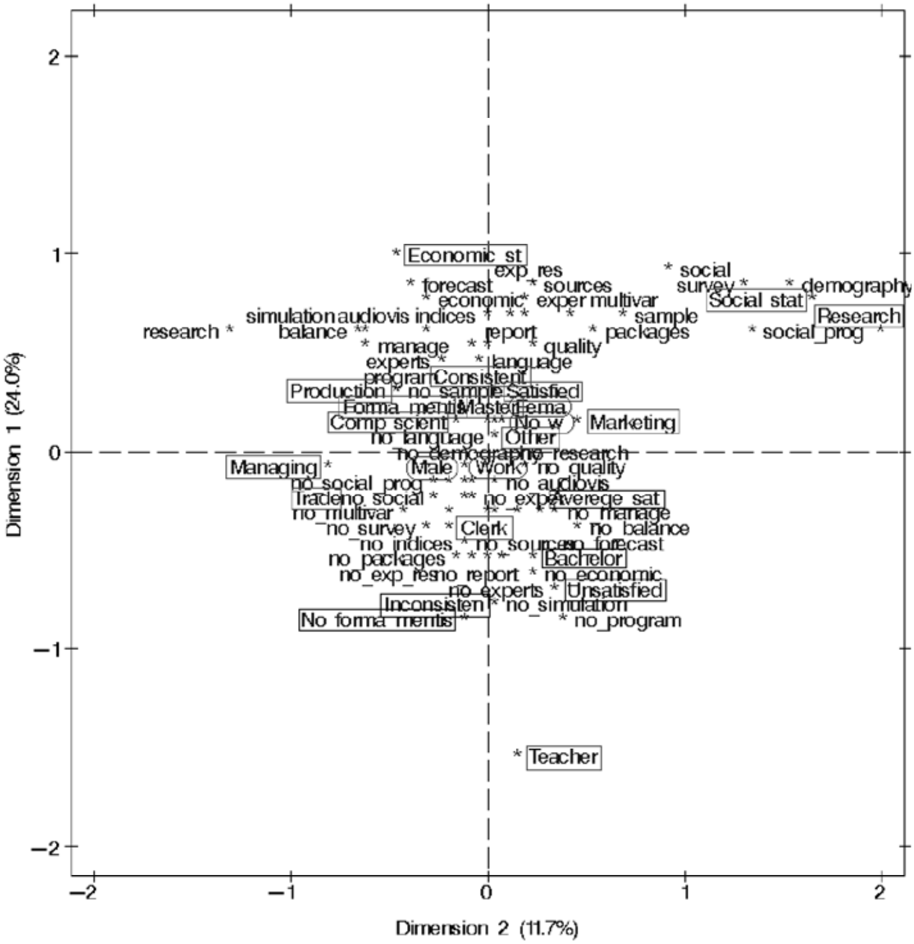


Figure 3. Correspondence analysis on use of competences for Paduan graduates in Statistics with profession, gender, type of programme, working status before graduation, job satisfaction, consistency, use of forma mentis, and need for educational qualification as supplementary variables.

Among the jobs related to the socio-demographic field we find the social and bio-statisticians that use the acquired competences to the largest extent, the assistant to research, and the assistant to marketing[5].

---

5 The distance between the professions in the marketing area, in the socio-demographic part of the display, and the competence of marketing research design, in the extreme side of the economic zone, is rather curious: although marketing research is an activity for economic statisticians, in practice it is realized with competences formed in the social study programmes.

Table 3. Factor coordinates and quality (sum of partial contributions) for each answer category of the supplementary variables

	Coordinates			*Coordinates*	
	1	*2*		*1*	*2*
Research area	0.19	1.95	*No work before graduate*	0.12	-0.02
Social, health statistician	0.67	1.53	*Master programme*	0.19	-0.06
Marketing expert	0.17	0.40	*Bachelor programme*	-0.36	0.11
Economic statistician	0.95	-0.38	*Unsatisfied with job*	-0.87	0.21
Production assistant	-0.03	-0.70	*Fairly satisfied with job*	-0.24	-0.09
Commercial assistant	0.03	-0.63	*Satisfied with job*	0.39	0.04
Management assistant	0.22	-1.04	*Job consistent curriculum*	0.54	-0.02
Computer scientist	0.25	-0.02	*Job inconsistent*	-1.01	0.03
Teacher	-1.45	-0.06	*Use forma mentis*	0.27	0.05
Clerk (general)	-0.75	-0.25	*No use forma mentis*	-1.13	-0.19
Other	-0.29	-0.12	*Compulsory school required*	-1.18	-0.16
Male	-0.25	-0.13	*Secondary school required*	-0.45	-0.38
Female	0.19	0.10	*University degree required*	0.55	0.30
Worked before graduation	-0.32	0.06			

On the economic side, we find both jobs associated with a massive use of competences, such as the economic statistician – the highest position on the first factor – and others associated with a limited use of competences, like the clerical positions in trade, management and production activities.

We observe a tendency for females to work in the social area and use more competences than males in consequence of the larger frequency of socio-demographic courses by female students.

Those who worked before graduation use competences less than those who entered the job market after graduation. This may depend on the likelihood that, six months after graduation, most of them may have maintained the same job they had before, which obviously did not require a university education.

The extreme positions on the first factor regard job-curriculum consistency, job satisfaction and systematic use of *forma mentis*. The non-use of competences is associated with unsatisfactory and inconsistent jobs, which exclude the use of a "statistical *forma mentis*". The second factor has no extreme values on any of the supplementary variables, apart for types of job.

In general, the professional activities of graduates in Statistics are consistent with the underlying competences, so that we can imagine jobs as combinations of competences. Hence, jobs of newly recruited graduates in Statistics can be considered as a "basket" whose content is the required knowledge, skills and attitudes.

Finally, we identified similarity between jobs according to the type and intensity of competence use. We carried out a hierarchical agglomerative cluster

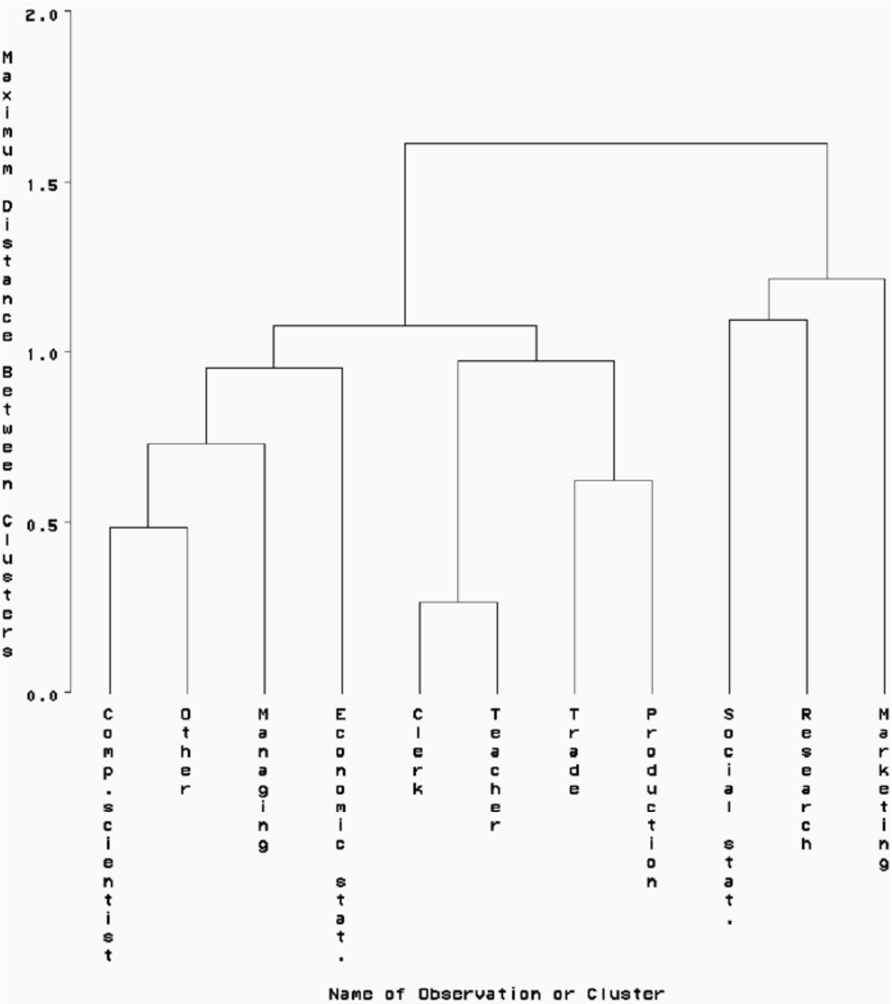


Figure 4. Dendrogram from cluster analysis on jobs of Paduan graduates in Statistics (Euclidean distance, complete linkage)

analysis on this matrix, using the complete link and the single link methods. The cophenetic correlation coefficient computed for the three applications indicates complete link as the preferable method, and this allows us to identify clusters of jobs.

The dendrogram analysis (Fig. 4) suggests the existence of two groups of jobs, one typical of socio-demographic jobs and another typical of economic and business activities. The economic area includes also jobs that make scarce use of statistical competences, such as teaching and general clerical positions.

4. Final remarks

The analysis of competences of graduates in Statistics highlighted a heterogeneous situation where, together with small groups of graduates who use refined statistical competences, a majority of graduates applies just a little of the knowledge acquired at school.

The competences graduates use the most are cross-occupation and we put forward they should be cross-curriculum: foreign language skills, computer skills, communicational, relational and organizational skills. It is relevant to be able to communicate with experts on other-than-Statistics subjects and write research reports.

Highly qualified statistical competences are the skill of identifying indicators for decision-making, that of analysing economic phenomena and company's budget and the skill of interpreting the results of multivariate analyses and experiments.

We stressed the relevance of computer skills for a graduate in Statistics to find a suitable job. Nevertheless, the graduates employed as computer scientists may get frustrated in the end because this occupation is barely compatible with statistical activities.

For those who use the acquired competences, the professional activities substantially reflect their formative curricula. A professional profile made of organizational and substantive competences in the field of company management, economic analyses, computerized support to production, R&D and commercialization activities, echoes the courses in the economics and business fields.

On the other hand, the socio-demographic curricula corresponds to professional profiles mainly employed in support-to-research activities, health and community services, marketing research, where competences on survey methodology and data analysis, design and evaluation of social policies, projects and services, demographic and qualitative analysis are needed.

The framework does not change if we concentrate on jobs instead of competences: in fact, we observe a tri-partition of graduates according to the type of work activity that is graduates employed in a-specific jobs, in social field jobs and in the economic and management field jobs.

The tricky aspect is the set, quite wide in terms of size, of graduates who do not use specific-to-statistics competences, cover a position that is rarely consistent with their expectations, show a low level of job satisfaction, and whose position does not require a university degree. A thorough analysis of the performed activities, of the used competences, and of the missing ones, could be useful to identify possible educational gaps and plan courses adherent to the job market needs.

References

BENZECRI J.-P. (1979) Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum à [BIN-MULT], *Cahiers de l'Analyse des Données*, **4**: 377-378.

FABBRIS L. (1997) *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.

FABBRIS L. (2003) *Il monitoraggio dell'inserimento occupazionale e professionale dei laureati e dei diplomati dell'Università di Padova*, Quaderno PHAROS n. 6/2003, Osservatorio sul mercato locale del lavoro, Università degli Studi di Padova, CLEUP, Padova.

FABBRIS L. (ed) (2005) *Il repertorio delle professioni della Facoltà di Scienze Statistiche, II ed.*, CLEUP, Padova.

FABBRIS L., MARTINI M.C., ROTA G. (2002) Il valore di mercato dei titoli di laurea e diploma di Padova a sei mesi dal conseguimento. In: DELVECCHIO F., CARLI SARDI L. (eds) *Indicatori e metodi per l'analisi dei percorsi universitari e post-universitari*, Cleup, Padova: 171-194.

MARTINI M.C. (2003) Professional benchmarking of degrees at the University of Padua. In: MRVAR A., FERLIGOJ A. (eds) *Proceedings of the Seventh Young Statisticians Meeting, Metodološki zvezki, vol.21*, Ljubljana.

SAS Institute Inc. (1994) *SAS/STAT User's Guide, Volume 1 (version 6)*, SAS Institute Inc., Cary, NC.

SAS Institute Inc. (2000) *SAS/STAT Software: Changes and Enhancements, Release 8.1.*, SAS Institute Inc., Cary, NC.

SCOTT J. (1991) *Social Network Analysis: A Handbook*. Sage, London.

SOKAL R.R., ROHLF F.J. (1962) The comparison of dendrograms by objective methods, *Taxon*, **11**: 33-40.

Academic Training and Competence Analysis of Social Work

Laura Carli Sardi

Department of Quantitative Methods, Siena University, Italy

Summary. In this paper, we present a quantitative and qualitative analysis of the competences and employability of graduates in Social Work. Our approach follows the methodological lines of the Credits project, which includes the statistical analysis of interviews with experts, such as teaching staff and entrepreneurs, mail questionnaires, focus groups, and the textual analysis of documents. Our target is the entire population of social workers in Tuscany. We expect to create indicators of the competences required by the job market, verify the efficacy level of education and training, and create an interactive systematic relation between academe and the labour market.
Keywords: Academic training, Competences, Social work.

1. Competence and performance

In the following, we refer to *competence* as a new concept and not as a derivative of the ancient latin term *competentia,* which applies to the sphere of jurisdiction. The English word *competence*, as Chomsky (1969) used it in linguistic epistemology, is to be considered a developmental process of knowledge, while *performance* names what is put into action. *Competence* is the acquisition through experience, on a knowledge basis, of the ability to act (Smelser & Baltes, 2001).

The transition from university to work has been the object of several scientific studies (Dolton *et al*., 1993; ISFOL, 1994; Mariani & Tronti, 1994; ISTAT, 2000; Biggeri *et al*., 2001; Fabbris 2003) and the relationship between graduates' competences and professional roles has been the object of a number of empirical research projects (La Rosa, 1984; Carli Sardi, 2002). These studies helped to interpret and link individual, institutional and market factors with the elaboration and experimentation of particular learning models.

The need of worker's knowledge and the interactive relationships with the organisation take us back to Likert & Murphy (1938) definition of the worker as a person with 'emotional maturity' so that it is necessary to take into consideration a set of psycho-social processes which may change his or her cognitive structure, motivation and capacity for action. The question concerns all formative agencies, from school to university, from profession to workplace, be it a public institution, a private company or a professional office.

Wittorski (1998) has broken down the competence as a function of work into components of the training process. The components are cognitive, affective and social, each of which concerns actions which imply a number of dimensions which are cross-occupational, subjective, inter-subjective, objective, thus defining a complex concept of competence, measurable through a set of indicators and evaluation procedures.

The core components of competences have been sufficiently theorised (Boyatzis, 1982), but there is a certain fragility going from description to definition and to identification[1].

From the theoretical point of view, competences are not just behaviours, even though they may be expressed through individual behaviour. Competences are the expression of knowledge, abilities, and attitudes that express a potential capacity of the individual that can be realised and powered within a work context. Therefore, knowledge, experience and motivation refer to specific contexts.

We want to draw the attention to the competences that are necessary to realise a "regulated profession". They are different from the professional business skills because the work contexts are different and vary according to economic activity and size of the company.

In the following, we will refer to the professional profile of the *social worker*. This profile has been under consideration since the first normative law of study programmes in Social work[2]. A syllabus was first organised during the period from 1960 to 1970 in six "special university schools". In the last years, a ministerial norm regulated the training of social workers with rigid curricula, until the university diplomas and the current three-year first-level programmes were born (Carli, 1996). The common denominator in all these phases has always been the distinction between knowledge (*savoir*), expertise or skills (*savoir faire*) and attitudes (*savoir être*).

The relationship between education and competences has generated a vast multidisciplinary literature with an applied psychology perspective.

---

1 The links between professions and the market have been studied by Sarchielli (1993), Levati & Saraò (1998), Ajello (2002). Advanced statistical methods for the analysis of competences have been experimented by Buscema *et al*. (1999), Authier & Levy (2000), Fabbris (2003), Bolasco (2003).

2 A programme in Social work was first formulated by Senator Ossicini, professor of Psychology at the University of Rome "La Sapienza".

We can identify three basic models:

a) an individualistic approach (Spencer & Spencer, 1993), which refers to the cognitive sphere and personal abilities, and also takes into account "achievement";
b) a systemic–relational approach (Barney 1991), which refers to the job system, and so to skills and technologies, with a strong base of competences;
c) a mixed model, which usefully combines personal and environmental characteristics with experience, similar to Piaget (1972) pedagogic conceptualisation.

In the present analysis, we adopted the last model. We will test it by examining both the social worker's route from university to the workplace and documents that identify, on the one hand, the training system and, on the other, standards of recognition and testing.

We used the data gathered by CREDITS-EST (Schmidt & Michelotti, 2002) project that examines the subject matter in the Social work study programmes. In order to detect the subject contents on the syllabus, we analysed official programmes and consulted various documents, such as the norms for the thesis production and project work, and the English language requirements. The aim was to gain an understanding of the competences (basic, technical-specific and cross-occupation) and the partition into academic subject fields.

The result of such an analysis was a list of contents that we submitted to the judgement of some teachers and people responsible for other training experiences. They were asked to rate – on a 1 to 5 point scale – the items according to the importance for future social workers (see the following scheme).

> Please, assign a number to each according to the following scale:
> 1=totally irrelevant 2=not relevant; 3=relevant; 4=very relevant; 5=extremely relevant)
>
> *Principles and foundations of social service*
>
> Social service as a multidimensional profession
> The relationship between Welfare State and the development of social service in Italy
> The profession of social worker: social and institutional mandate,
> Objectives and functions
> The philosophical and value basis, ethics and attitudes
> Ethics and attitudes
> Theoretical and methodological basis of work: models
> Working instruments and techniques: relationships, working with groups, documentation, and administrative work.

We applied the same procedure with recent graduates who did not have work experiences as social workers, again with the aim of collecting their relevance judgement about competences. A group of working professionals was sampled for comparative reasons from the Tuscany Register of Professionals. As a whole, the register contained 1,600 names in the year 2003.

A mail questionnaire was sent to on a random sample of 600 graduates, which produced a 70% return without any reminder[3]. After a check for validity and completeness, an exploratory study of 200 randomly selected subjects was carried out.

In order to maintain a correspondence with the academic classification, both traditional and general areas were grouped according to competence classification into the basic, technical-professional and cross-occupational. The analyses carried out on the different viewpoints are presented in Section 2.

2. The exploratory analysis

The competences that the teachers felt they offered, those that the graduates felt they had acquired, and those that the professional social workers believed they should posses are presented in Table 1.

Table 1. Average scores given to competences by new graduates, university teachers and social workers.

Subjects	New graduates	University teaching staff	Working graduates	**Total**
Principles and fundamentals of social services	4.20	3.84	3.25	3.87
Public institutions legislation	3.53	3.54	4.06	3.69
Private institutions legislation	3.26	2.80	1.38	2.63
Political economics	3.07	2.92	1.88	2.69
Sociology	3.40	3.08	4.06	3.55
Hygiene	4.15	3.50	1.94	3.40
General psychology	3.52	3.59	5.00	3.97
Social statistics	4.20	3.75	3.44	3.91
Social history	3.40	3.75	1.56	2.89
Computer science	4.05	3.50	4.06	3.97
English (as a foreign language)	2.93	4.50	3.44	3.31
Administrative law	3.44	3.33	2.63	3.19

[3] The mail questionnaire contained 180 items relating to socio-demographic and family items, school and university curriculum, previous experience and present working situation, specific knowledge, updating, competences used and deemed necessary, evaluation of training etc., and standard scale items relating to personality, scales measuring attitude (Krech *et al.*, 1962) such as willingness to work in a public body, values and other psychological factors.

The teaching staff and the graduates' evaluations are close to each other. One can discern a slight tendency of new graduates to rate lower than teachers the individual components of the knowledge/competences, and, on the other hand, higher scores are given by the teachers to general principles such as the juridical aspects of social work and research methodology.

The professionals, on the other hand, rate either higher or lower the competences. One can see convergence on the juridical-economical issues and the fundamental technical-specific areas, with little standing out for the competences relative to specialist subjects (e.g. historical approaches).

The question raised by the specific analysis refers to how much the competences indicated by those working in the field as being necessary or important reflect the necessities of the social work and in which proportion they are a factor of personal utility which could be less representative.

Table 2. List of identified *transversal* competences and related academic subjects

Computer skills	▪ Knowledge of and ability to use the PC for activation or use of information systems ▪ Knowledge of and ability to use word-processing software	Information technology Methodology of social research
Foreign language	▪ Threshold level of English or French for the understanding of documents and articles or for interacting with the public or colleagues when an interpreter is not available ▪ Exchange of simple letters with particular reference to the English language	P.E.T. Language test Erasmus programme.
Dissertation	▪ Methodological abilities and research skills using a variety of sources ▪ Knowledge of a specific sector ▪ Ability to create a document to report objectively and correctly expressed ▪ Abstraction skills	Dissertation
Practicum	▪ Specific knowledge of the institutional sector ▪ Ability to relate to colleagues and outside institutional skills ▪ Communication skills on a number of levels ▪ Carrying out of administrative tasks ▪ Prioritisation of problems presented by members of the public ▪ Use of information technology ▪ Use of specific professional techniques (interviews, reports, meetings, assessments, projects) ▪ Reading the community and local area ▪ Awareness of attitudes and behaviour of individuals and groups with reference to professional values and principles (tolerance, democracy, etc.)	Practicum

Table 3. List of identified *professional* competences and related academic subjects

Methodologies	▪ Recognition of the rights of citizens as being protected by essential levels of help (UEA/UVEAS) and ability to define them ▪ Ability to organisational innovation of traditional services and of expansion ▪ Use of planning methodology ▪ Knowledge of and correct use of staff and line of the Service ▪ Ability to express and communicate on a number of levels of the institutional organisation and to manage outside information ▪ Ability to analyse and evaluate the needs of groups through direct research processes ▪ Ability to use methods of quality control	Organization of Social services Methodology of social research Thesis
Principles Ethics	▪ Activation of promotional functions with respect to individuals and the community ▪ Use of interdisciplinary work and team work ▪ Use of networks ▪ Respect for the ethical code with respect to the public and colleagues.	Methods and techniques of the Social services Principles and fundamentals of Social services Social policy
Work with people Relationship skills	▪ Ability to recognise "the person" and support with individualised programmes of aid ▪ Sharing of the objectives of efficacy of the services and interventions with adequate documentation and checks ▪ Ability to share with other services efficiency and economy with respect to institutions with documented and checked use of resources ▪ Capacity to produce administrative documents taking direct responsibility; knowledge of own and those depending on administrative management ▪ Ability to produce specific professional documents for the use of the public and the services ▪ Ability to work with projects with scientific planning and methodology	Methods and techniques of Social Services Organization of Social services General and developmental psychology Practicum

After this explorative analysis, we included in the structure of the questionnaire the contents that are shown in Tables 2, 3 and 4.

Table 4. List of identified *basic* competences and related academic subjects.

Social	▪ Awareness of the socio-political context of the country and with reference to Europe ▪ Ability to identify social phenomena to be found in physiological and pathological forms in a specific local area (community), including social-health problems ▪ Acquisition of and adaptation to new principles (for example, recently the idea of "federalism", "universalism" of the services)	General sociology Sociology of the family General psychology Social medicine Political systems
Legal	▪ Public institutional contexts ▪ Pluralism of the service providers ▪ Updating of Health and Social service legislation at the national and regional levels ▪ Acquisition and updating of general welfare norms through use of published material or more competent colleagues ▪ Acquisition of and adaptation to new principles	Public institutions legislation Private institutions legislation Penal law Administrative law
Statistical and economic	▪ Ability to trace statistical documentary sources in order to become familiar with demographic and socio economic characteristics of an area ▪ Ability to produce database spreadsheets and/or to carry out directly surveys to document a report, a plan, to produce summarising data	General economics Applied economics Statistics

The correlations between the judgements expressed by the workers about their training and those concerning their work situation proved to be helpful (Table 5).

One can see a very high value assigned to the adequacy and satisfaction for technical skills with reference to the standard of the university training. One reason is that many of the teaching staff also have roles as management in the social work and thus corroborate the transfer of the academic knowledge into expertise by the graduates in their practical experience, with the help of supervisors and the integration of seminars and experimental theses.

For the same reason, it is possible to explain why the basic and cross-occupational competences received lower consideration from those at work, even though they were judged as important and useful for work.

This may depend on the fact they are fragmented, often of single value and above all characterise "the other" so much, so that the common response from all graduates was that any other graduate could not carry out their job.

Table 5. Matrix showing the correlation between the evaluation of training and the evaluation of competences

		Basic training and acquired technique	Acquired specialised competences	Practical-professional acquired experience	Importance of the basic subjects for work	Importance of professional subjects or work	Adequacy of the professional training received from University	Average of basic competences	Average of professional and characteristic competences	Average of the transversal competences
Basic training and acquired technique	***Pearson***	1.000	0.815	0.707	0.281	0.185	0.786	0.162	0.222	0.090
	p-value	.	0.000	0.000	0.036	0.169	0.000	0.225	0.094	0.500
Acquired specialised competences	***Pearson***		1.000	0.873	0.365	0.242	0.817	0.137	0.163	0.062
	p-value		.	0.000	0.006	0.072	0.000	0.308	0.226	0.647
Practical-professional acquired experience	***Pearson***			1.000	0.237	0.158	0.698	0.108	0.146	0.069
	p-value			.	0.079	0.242	0.000	0.420	0.273	0.605
Importance of the basic subjects for work	***Pearson***				1.000	0.402	0.381	0.278	0.195	0.092
	p-value				.	0.002	0.005	0.038	0.151	0.500
Importance of professional subjects or work	***Pearson***					1.000	0.376	0.298	0.233	0.238
	p-value					.	0.005	0.025	0.081	0.075
Adequacy of the professional training received from University	***Pearson***						1.000	0.233	0.261	0.139
	p-value						.	0.087	0.054	0.310
Average of basic competences	***Pearson***							1.000	0.705	0.586
	p-value							.	0.000	0.000
Average of professional and characteristic competences	***Pearson***								1.000	0.764
	p-value								.	0.000
Average of the transversal competences	***Pearson***									1.000
	p-value									.

3. Conclusions

What has emerged from this first analysis illustrates a clear transfer of the academic training into professional action.

By examining groups of basic, cross-occupational and specific competences, we identified a professional profile of the social worker, within a public and institutional dimensions, with a strong orientation towards the rights of citizens, and responsible for the governance of the care processes in terms of social and health services. He or she receives the support of specific professional methodologies, but also of procedures, technical instruments such as the computer, of professional skills and finally the ability to plan and evaluate scientifically.

The profile outlined does indeed correspond to the aims of the training with advances in some competences for example the use of computers with information systems, administrative procedures, basic foreign language skills for communication, attention paid to research methods for all documentary sources and their production; for other dimensions they appear somewhat restricted compared with multidisciplinary training curricula.

On this topic, it seems that the economic dimension is not clearly defined across the competences, despite a need to manage budgets and estimate economies.

Limited are also the specialist skills and useful further training for particular target populations, e.g. immigrants.

References

AJELLO A.M. (ed.) (2002) *La competenza*, Il Mulino, Bologna.

AUTHIER M., LEVY P. (2000) *Gli alberi di conoscenza*, Feltrinelli, Milano.

BARNEY J. (1991) Firm resources and sustained competitive advantages, *Journal of Management*, **17**: 99-120.

BIGGERI L., BINI M., GRILLI L. (2001) The transition from University to work: a multilevel approach to the analysis of the time to obtain the first job, *Journal of Royal Statistical Society*, **164**: 293-305

BOLASCO S. (2003) *Applicazioni di analisi testuale*, Ed. Universitaria, Roma.

BOYATZIS R. (1982) *The Competent Manager: a Model Effective Performance*, Wiley, New York.

BUSCEMA M., SEMEION GROUP (1999) *Reti neurali artificiali e sistemi sociali complessi*, Angeli, Milano, Vol I, II.

CARLI SARDI L. (1996) La scuola di Servizio Sociale di Siena nella sua storia ed evoluzione, *Atti Convegno Formazione e operatività del Servizio sociale*, Tipografia Senese, Siena: 16-68.

CARLI SARDI L. (2002) Formazione universitaria e mondo del lavoro, *Quaderni del Nucleo di Valutazione*, Università degli Studi, Siena, **7**: 21-68.

CHOMSKY N. (1969) *Aspetti della teoria della sintassi*, Boringhieri,Torino.

DOLTON P.J., MARKPEACE G.H., TREBLE J.G. (1993) *The Youth Training Scheme and the School to Work Transition*, Research Paper, University of Wales, Bangor.

FABBRIS L. (2003) *Il monitoraggio dell'inserimento occupazionale e professionale dei laureati e dei diplomati dell'Università di Padova*, Quaderno Pharos, 6, CLEUP, Padova.

ISFOL (1994) *Competenze trasversali e comportamento organizzativo*, FrancoAngeli, Milano.

ISTAT (2000) Inserimento professionale dei laureati, *Informazioni 10*, Istat, Roma.

KRECH D., CRUTCHFIEELD R.S., BALLACHEY E.L. (1962) *Individual in Society*, Mc Graw Hill, New York.

LA ROSA M. (ed.) (1984) *Il lavoro che cambia*, Angeli, Milano.

LEVATI W., SARAO M.V. (1998) *Il modello delle competenze*, Angeli, Milano.

LIKERT R., MURPHY G. (1938) *Public Opinion and the Individual*, Harper, New York.

MARIANI P., TRONTI L. (1994) La transizione Università - Lavoro in Italia: una esplorazione delle evidenze dell'indagine ISTAT sugli sbocchi professionali dei laureati, *Economia e Lavoro*, **25**: 3-26.

PIAGET J. (1972) *Problèmes de psychologie génétique*, Dunot-Gonthier, Paris.

SARCHIELLI G. (1993) Nuove competenze per il lavoro che cambia: il ruolo della formazione professionale nel processo di socializzazione adulta, *Atti Convegno "Formation professionnelle en mutation", OFIAMT, Berne*: 4.

SMELSER N., BALTES P. (2001) *Social and Behavioral Sciences - International Encyclopedia*, University Press, Cambridge, IV: 2433.

SPENCER L.M., SPENCER M.S. (1993) *Competence at Work*, Wiley, New York.

SCHMIDT R., MICHELOTTI S. (eds) (2002) *Manuale per l'integrazione dei sistemi formativi in Toscana* , TRIO, Regione Toscana, Firenze.

WITTORSKI R. (1998) De la fabrication des compétences, *Education permanente*, **135** : 57-69.

Profiling and Labour Market Accessibility for the Graduates in Economics at Naples University

Simona Balbi, Maria Gabriella Grassia[1]

Department of Mathematics and Statistics, University of Naples "Federico II", Italy

Summary. In this paper, after defining a pseudo-panel of groups observed at subsequent times, we propose a strategy for the construction of a set of association rules related to different survey occasions. First, we measure the similarity between systems built at different times for understanding the stability of the phenomenon. We apply a procedure developed for symbolic data analysis for this purpose. The procedure consists of two phases: the definition of the pseudo-panel and that of a system of rules referred to the semantic marking technique. Then, the agreement between the systems is measured. We applied such a strategy for studying the labour market accessibility for graduate in Economics, the University of Naples "Federico II", and the market evolution during an eight-year time span.
Keywords: Semantic marking technique; Pseudo-panel; Association rule; Symbolic objects.

1. The pseudo-panel definition

The Faculty of Economics at the University of Naples "Federico II" has been carrying out for over twenty years recurrent sample surveys on its graduates in order to evaluate their labour market accessibility. We will examine the last three surveys for evaluating the evolution of the phenomenon and apply a coherent policy.

The questionnaire and the survey methods for these surveys are constant in time. Therefore, it is possible to examine the evolution of the graduates' be-

---

[1] This paper is the result of the joint effort of the two authors. S. Balbi was responsible for the final editing of Sections 1, 6 and 7, whereas M.G. Grassia was responsible for the other Sections.

haviours and destiny by constructing a *pseudo-panel* formed by *cohorts,* that is to say, sets of individuals identified according to characteristics that do not vary in time according to the studied phenomenon.

The analyses will be carried out on higher order units, formed by aggregating the elementary units, which are present at each survey occasion. For instance, it is possible to study if the selection devices used for graduate women who have obtained the maximum final score have changed in a given period.

The literature on symbolic objects has produced statistical methods for the analysis of complex structures. The complexity relates both to the characteristics of the units and the membership relationship linking each elementary unit to its own object (Section 2).

In the following, we put forward a strategy that, taking an advantage from the tools developed within the analysis of symbolic objects, makes a pseudo-panel approach feasible in the described context. At each survey occasion, the proposed strategy defines the constitutive elements of a pseudo-panel in terms of association between descriptors.

We propose a data-driven strategy suitable to set up a pseudo-panel according to the data association structure. The rules (called *implication rules* or *logical rules "if-then"*) will be referred to the survey waves and associated to measures of the rule authenticity.

The comparison between the rules may give a measure of the structural stability of the phenomenon. The rules refer to a symbolic data analysis frame, the *symbolic marking*, and the comparison between rules will be carried out by means of a similarity measure between symbolic objects.

2. The symbolic objects

A symbolic object, *s*, is defined as a triplet:

$$s = (a, R, d)$$

where: $d=(d_1, ..., d_j, ..., d_p)$ is a set of values on p descriptors, $Y=(Y_1,...,Y_j,...,Y_p)$, of the object,

a is a *recognition* function ,

$R=(R_1, ..., R_j, ..., R_p)$ is the type of relation applied for the comparison between the description provided at a conceptual level, in *intention*, from d, and an observation.

The descriptors of a symbolic object can be on a nominal, continuous or discrete scale and can have several categories for each object. The Boolean function a has categories *true* and *false* and identifies those elements which belong to the d description set and that are the *extension* of the s object, $ext(s)$.

The construction of conceptual models described in terms of symbolic objects (Bock & Diday, 2000) may be based on:

1) the expert opinion,
2) the knowledge acquired from repeated surveys.

We will merge the two approaches by using the implied longitudinal nature of the data. In order to construct a *pseudo-panel* we have to identify the structural characteristics that allow a partition of the time-related samples. Then, we will interpret the partition through the associated symbolic object.

Let us consider a set of units $E = \{1, 2, \ldots, Q\}$ to which a questionnaire was administered T times. This set can be partitioned in t subsets E_t, with $t = 1,\ldots, T$, composed by the units who participated to wave t. If, for example, a graduate participated in several waves, he or she is considered each time as a different individual. So, $E_1 \cap E_2 \cap \ldots \cap E_t = \varnothing$ and $\cup E_t = E$, for $t = 1, \ldots, T$.

Let us assume that the E elements are described by the same P variables $Y = \{Y_1, \ldots, Y_j, \ldots, Y_p\}$ and that each Y_j variable has m_j response categories. Continuous variables are made discrete with the same scale at all times. If the elementary units of the E set are groups of individuals that possess common characteristics, the data structure will be a symbolic matrix whose generic element is the marginal frequency distribution of the Y_j modal variable.

The s_k symbolic object is then defined as:

$$s_k = \mathop{\wedge}_{j=1}^{P} \left[Y_j = \left\{ y_{jm}, p_{jm} \right\}_{m=1,2,\ldots,m_j} \right], \tag{1}$$

where p_{jm} is the relative frequency of y_{jm}, m-th category of Y_j .

In defining a symbolic object, it can be useful to consider the object implications, assuming relations that can be expressed as logical rules (*if-then*, see*:* Agrawal *et al.*, 1993). The symbolic object s_k is defined as:

$$s_k = \underbrace{\mathop{\wedge}_{a=1}^{A} \left[Y_a = \left\{ y_{am}, p_{am} \right\}_{m=1,2\ldots m_a} \right]}_{(A)} \Rightarrow \underbrace{\mathop{\wedge}_{c=1}^{C} \left[Y_c = \left\{ y_{cm}, p_{cm} \right\}_{m=1,2\ldots m_c} \right]}_{(C)} \tag{2}$$

$$(A) \Rightarrow (C)$$

with $A, C \subseteq Y$ and $A \cap C = \varnothing$,

where A is the set of the antecedent categories (whenever possible, independent and exclusive variables) and C is the set of consequent categories.

With reference to T times, it is possible to define s_{kt} in each E_t. The use of complex structures allows us to measure the similarity between the objects collected in different times, as well as to value the stability of the structures. The comparison is done among the implication logical rules: for instance, two generic objects s_{kt} and $s_{k't'}$ with the same expression for A and a different expression for C, imply a change of the individual behaviour going from time t to time t'.

3. Selection of descriptors that define the objects implication

For the selection of the descriptors apt to define the intension of the symbolic object, having chosen the *consequence* variable, we use the semantic marking technique (Gettler-Summa, 1998; Grassia e Muratore, 2001), considering all the remaining variables that are part of the set of *antecedent* characteristics within the *implication* of a logical rule.

The semantic marking is a non-binary segmentation technique aimed at pointing out the characteristics of a class *K* (it may be a natural partition, or derived from a cluster analysis), considering the conjunctions and disjunctions logical links among the attributes that describe the units. The procedure determines some *marking cores*, that is to say, groups of individuals that are identical according to a set of "traits":

$$mc_g : [Y_1 = y_{1m}] \wedge ..... \wedge [Y_r = y_{rm'}] \quad \text{with} \quad r \leq P . \tag{3}$$

The union of the *G marking cores* mc_g (expressed in terms of logical AND), based on the OR disjunction operator, forms the description of the K class:

$$K : mc_1 \vee mc_2 \vee ...... \vee mc_g \vee ...... \vee mc_G \ . \tag{4}$$

By using the semantic marking, we build abstractions based on the two criteria of: homogeneity of the *K* elements and difference with the *NOT-K* elements. The parameters of the algorithm are the indexes:

1. Rec = Card $[\, ext_K(mc_g)\,]$,
2. Deb = Card $[\, ext_{\overline{K}}(mc_g)\,]$,

where mc_g is a generic marking that is a subset of *K*, characterized by the same categories of one or more descriptors. The *Rec* index is the percentage of elementary units belonging to *K* that satisfy the conditions defined by the mc_g marking. The *Deb* index is the percentage of elementary units that satisfy the marking conditions, but do not belong to *K*.

The semantic marking is a procedure for constructing symbolic objects, because its output is a symbolic matrix of smaller dimensions than the input matrix, with the same variables expressed in modal form, that is to say, with the respective frequency or probability distributions.

4. The comparison among objects

We compare two symbolic objects in times *t* and *t'* with a dissimilarity measure (Bock & Diday, 2000; Bruzzese & Davino, 2002) based on the Minkowski L_1 distance:

$$d(s_{k_t}, s_{k'_{t'}}) = \sum_{j=1}^{P} \frac{1}{P} d(Y_j(s_{k\,t}), Y_j(s_{k'_{t'}})) \tag{5}$$

where $d(Y_j(s_{k\,t}), Y_j(s_{k'_{t'}}))$ compares the frequency distributions so that:

$$d(Y_j(s_{k\,t}), Y_j(s_{k'_{t'}})) = \frac{1}{m_j} \sum_{j=1}^{m_j} \frac{\left| p_{y_j}(s_{k_t}) - p_{y_j}(s_{k'_{t'}}) \right|}{\max \left| p_{y_j}(s_{k_t}) - p_{y_j}(s_{k'_{t'}}) \right|}. \tag{6}$$

The dissimilarity index varies between 0 if the two objects have the same frequency distribution for each variable, and 1 if they are completely different, that is to say, if the dissimilarity is maximum for all variables.

If we introduce the logical relationships (2), the dissimilarity between two symbolic objects can be the average of the dissimilarities among the frequency distributions of the variables in antecedent (A) and consequent (C) categories:

$$d(s_{kt}, s_{k't'}) = \frac{1}{2}\left(d(A_{kt}, A_{k't'}) + d(C_{kt}, C_{k't'}) \right), \tag{7}$$

where $d(A_{kt}, A_{kt'})$ e $d(C_{kt}, C_{kt'})$ are the dissimilarities among the objects s_{kt} and $s_{kt'}$ obtained with formulas (5) and (6), considering in the first case only the variables in antecedent and in the second only the variables in consequence.

5. A synthetic index for rule evaluation

Let us consider a single modal consequent variable. The extension of this variable to other consequent variables is immediate, through the construction of composite variables. By using the semantic marking, which characterizes a natural partition obtained from the categories of the consequent variable, we can have objects that, compared in *T* times, can assume the same or different consequence.

If the Y_c consequent variable has m_c response categories ($y_{c1}, \ldots y_{cm}, .. y_{cm_c}$), the dissimilarity between times *t* and *t'* will be the mean obtained for each category by averaging the means of the dissimilarities among the antecedences of the objects at times *t* and *t'* that have an identical consequence and averaging the similarities among the antecedences of the objects at times *t* and *t'* having different consequences:

$$d(t,t') = \sum_{y_{cm}=1}^{m_c} \frac{1}{2m_c} \left\{ \frac{1}{ng} \sum_{k=1}^{n} \sum_{k'=1}^{g} d(A_{s_{k_{y_{cm}}t}}, A_{s_{k'_{y_{cm}}t'}}) + \frac{1}{nh} \sum_{k=1}^{n} \sum_{k''=1}^{h} (1 - d(A_{s_{k_{y_{cm}}t}}, A_{s_{k''_{\bar{y}_{cm}}t'}})) \right\}, \tag{8}$$

where n is the number of symbolic objects at time t having y_{cm} category for the answer variable (consequent), g is the number of symbolic objects at time t'

having the same modality m_j for the response variable, h is the number of symbolic objects at time t' having any other category for the variable Y_c.

6. Labour market accessibility for Economics graduates

We applied the strategy set forth in the previous chapters in order to study the labour market accessibility for students who graduated in Economics at the University of Naples "Federico II".

The data at hand were collected in three repeated surveys performed in 1997, 2000 and 2002. The questionnaire was structured into a set of common questions and specific modules with each one relating to a survey occasion. The comparisons concerned the common parts of the questionnaires (Table 1).

We analysed 1030 units: 385 from the 1997 survey, 397 from 2000 and 248 from 2002 (Table 2).

The *consequence* variable is the employment status with three categories: *unemployed*, *not permanently employed*, *permanently employed.* The aim of our analysis is to study the evolution of the labour market accessibility from 1997 to 2002. There are 29 *antecedent* variables in the construction of logical rules. By using the semantic marking, 43 marking cores have been pointed out for the different periods (13 in 1997, 18 in 2000, and 12 in 2002). A measure

Table 1. Common variables of the questionnaire

1	Gender	16	Number of job interviews
2	Residence during studies	17	Job conditions
3	Diploma degree	18	Time spent searching for job
4	Age	19	Job position
5	Type of secondary school	20	Type of job
6	Secondary school degree	21	Company's economic sector
7	Type of studies	22	Channels used for job finding
8	University attendance	23	Work site
9	Subject of diploma	24	Second job
10	Years spent getting degree	25	Job sector condition
11	English knowledge	26	University satisfaction
12	Informatics knowledge	27	Job satisfaction
13	PhD	28	University education
14	Job qualification	29	Job mobility
15	Worked during studies	30	Research channels

Table 2. Dataset structure

	ID	*Gender*	*Study Residence*	*Diploma degree*	*Age*	*…*	*Job satis-faction*	*University education*	*Research channels*
1997	*I1*	1	3	4	2	….	1	4	4
	I21	2	2	2	3	….	2	2	2
	….	….	….	….	….	….	….	….	….
	I254	2	1	1	2	….	1	1	1
2000	*I1*	1	2	3	1	….	2	3	3
	I2	1	2	1	1	….	2	1	1
	….	….	….	….	….	….	….	….	….
	I332	1	1	2	2	….	1	2	2
2002	*I1*	2	3	1	2	….	2	1	1
	I2	2	1	1	2	….	3	1	1
	….	….	….	….	….	….	….	….	….
	I220	1	2	1	4	….	1	1	1

of importance related to the previously described *Rec* and *Deb* indexes has been associated to each *marking core*.

For example, let us consider the first marking for the *not permanently employed* category during the years 1997, 2000 and 2002 (Tables 3, 4 and 5). By using the semantic marking, the individuals who answered not to have a permanent job constitute pseudo-panels represented by 7 symbolic objects in 1997 and in 2000 and by 5 in 2002. All the non-marked individuals form a residual symbolic object for each year.

Within the wide group of interviewees who declared not to have a permanent job, in 1997 there was a subgroup of individuals who had found a job thanks to family ties. The persons of this subgroup were *self-employed*, *did not work* during their university studies and *were unsatisfied* of the university education received.

So, the logical rules were expressed as:

IF Channels Used = *family ties* **And** Job Position = *self-employed* **And** University Satisfaction = *no* **And** Worked during studies = *no* **And** Company's economic sector = *n.a.* (*not applicable*)
THEN Job Position= *not permanently employed*.

This rule was good for the 13% of the *not permanently employed* persons and was bad for the 2% of the other persons.

Table 3. First marking of the not permanently employed category in 1997

Category: not permanently employed category year 1997 (119)					
	Weight	**Percentage**	**T-Value**		
Marking 1	**18**	**1,75**	**7,443**	**Category**	**Variable**
			8,118	*Family ties*	Channels used
REC	16	13,45	7,933	**And** *Self-employed*	Job position
REC Correct	16	13,45	7,299	**And** *No*	University satisfaction
REC Cumulated	16	13,45	3,623	**And** *No*	Worked during studies
DEB	2	11,11	2,048	**And** *not applicable*	Company's economic sector

For the year 2000 survey, we have the following rule:

IF University Satisfaction = *partial* **And** Company's economic sector = *services for firms* **And** Channels Used = *family ties* **And** Job Position = *self-employed person* **And** Type of Diploma = *technical diploma* **THEN** Job Position= *not permanently employed.*

This rule was good for the 12% of the *not permanently employed* persons and was bad for the 1% of the other graduates.

Table 4. The first marking of the not permanently employed category in year 2000

Category: not permanently employed category year 2000 (181)					
	Weight	**Percentage**	**T-Value**		
Marking 1	**24**	**2,33**	**7,411**	**Category**	**Variable**
			9,248	*Partial*	University satisfaction
REC	21	11,60	7,552	**And** *Services for firms*	Company's economic sector
REC Correct	21	11,60	7,247	**And** *Family ties*	Channels used
REC Cumulated	21	11,60	4,324	**And** *Self-employed*	Job position
DEB	3	12,50	1,598	**And** *Technical Diploma*	Type of diploma

For the year 2002 survey, we have the following rule:

IF University satisfaction = *missing* **And** Job position = *self-employed* **And** Job satisfaction = *No*
THEN Job Position= *not permanently employed.*

This rule was valid for 50% of the *not permanently employed* and, in any case, it was false.

Table 5. The first marking of the not permanently employed category in year 2002

Category: not permanently employed category year 2000 (181)					
	Weight	**Percentage**	**T-Value**		
Making 1	**76**	**7,38**	**18,015**	**Category**	**Variable**
REC	76	50,33			
REC Correct	76	50,33	8,018	*Missing*	University satisfaction
REC Cumulated	76	50,33	7,473	**And** *Self-employed*	Job position
DEB	0	0,00	6,578	**And** *No*	Job satisfaction

We highlight the radical modifications occurred in the last period for the necessary qualifications requested to enrol at the Register of Graduates in Economics and Commerce, a regulated profession that represents one of the most important employment opportunities for graduates in Economics in Naples.

Therefore, this event can be measured, by applying the proposed strategy, considering the characteristics in the antecedent part of the rules referred to different periods.

The aim of the proposed strategy is an evaluation of all the modifications registered in the structure of the rules that represent the phenomenon in each period. The three previous marking cores represent only one of the 7 observed subsets (5 for 2002). The global evaluation must then consider all the rules that have the same consequence and all the situations when the antecedences have produced different consequences in different times. We have also considered, while computing this index, all the variables pointed out on the studied object.

By using the marking cores, we have built the matrix of the symbolic objects in modal form, composed by 43 rows and 30 columns (Table 6).

In Table 7 we show, for each compared pair of years, the value of the first term of (8), $\sum^{m_c} \frac{1}{n\,g\,m_c} \sum_{k=1}^{n} \sum_{k'=1}^{g} d(A_{s_{k_{y_{cm}} t}}, A_{s_{k'_{y_{cm}} t'}})$, that is to say the dissimilarity of the rules that have identical consequence.

The similarity among the rules in each time t and the rules of the other years having different consequences, that is to say the value of the second term $\sum_{y_{cm}=1}^{m_c} \frac{1}{n\,h\,m_c} \sum_{k=1}^{n} \sum_{k''=1}^{h} (1 - d(A_{s_{k_{y_{cm}} t}}, A_{s_{k''_{\bar{y}_{cm}} t'}})$ of (8), is shown in Table 8.

Table 6. Symbolic matrix of data

	Gender		Residence			Diploma Degree				....	To move				
	Male	Female	Naples	Province of Naples	Other Province	- 95	96 - 105	106 - 110	110 with Lode	... ...	Yes	Yes only in Italy	Yes only in my region	No	NR
1997 unemployed 1	35,00	65,00	45,00	35,00	20,00	0,00	100,00	0,00	0,00	... ...	40,00	25,00	25,00	5,00	5,00
1997 unemployed 2	52,38	47,62	52,38	23,81	23,81	23,81	33,33	23,81	23,81	... ...	42,86	33,33	14,29	0,00	9,52
1997 unemployed 3	46,15	53,85	38,46	46,15	15,38	0,00	100,00	0,00	0,00	... ...	38,46	7,69	23,08	30,77	0,00
1997 unemployed 4	53,66	46,34	50,00	26,83	23,17	21,95	37,80	21,95	20,73	... ...	45,12	30,49	10,98	9,76	3,66
1997 unemployed without marking	100,00	0,00	17,65	41,18	41,18	5,88	70,59	5,88	0,00	... ...	23,53	17,65	35,29	23,53	0,00
1997 not permanently 1	38,89	61,11	27,78	38,89	33,33	16,67	72,22	16,67	5,56	... ...	33,33	11,11	5,56	11,11	38,89
1997 not permanently 2	46,15	53,85	38,46	23,08	38,46	15,38	46,15	15,38	23,08	... ...	38,46	23,08	0,00	0,00	38,46
1997 not permanently 3	42,86	57,14	7,14	35,71	57,14	21,43	71,43	21,43	0,00	... ...	35,71	7,14	14,29	7,14	35,71
1997 not permanently 4	0,00	100,00	27,27	18,18	54,55	0,00	100,00	0,00	0,00	... ...	27,27	0,00	9,09	9,09	54,55
1997 not permanently 5	12,50	87,50	37,50	37,50	25,00	25,00	50,00	25,00	0,00	... ...	12,50	25,00	0,00	12,50	50,00
1997 not permanently 6	40,00	60,00	100,00	0,00	0,00	0,00	60,00	0,00	20,00	... ...	20,00	20,00	20,00	0,00	40,00
1997 not permanently without marking	53,49	46,51	34,88	37,21	27,91	16,28	32,56	16,28	18,60	... ...	25,58	0,00	4,65	4,65	65,12
1997 permanently without marking	55,56	44,44	38,89	38,89	22,22	5,56	27,78	5,56	27,78	... ...	22,22	44,44	16,67	5,56	11,11
2000 unemployed 1	66,67	33,33	16,67	16,67	66,67	0,00	100,00	0,00	0,00	... ...	50,00	0,00	0,00	0,00	50,00
2000 unemployed 2	50,00	50,00	16,67	16,67	66,67	0,00	100,00	0,00	0,00	... ...	0,00	0,00	33,33	0,00	66,67
2000 unemployed 3	57,14	42,86	0,00	100,00	0,00	14,29	42,86	14,29	14,29	... ...	28,57	14,29	28,57	0,00	28,57
2000 unemployed 4	60,00	40,00	40,00	60,00	0,00	0,00	100,00	0,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 unemployed 5	25,00	75,00	100,00	0,00	0,00	0,00	100,00	0,00	0,00	... ...	25,00	0,00	25,00	0,00	50,00
2000 unemployed without marking	43,94	56,06	46,97	30,30	22,73	22,73	31,82	22,73	22,73	... ...	13,64	21,21	7,58	9,09	48,48
2000 not permanently 1	70,83	29,17	33,33	45,83	20,83	16,67	45,83	16,67	12,50	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 2	50,00	50,00	37,50	37,50	25,00	18,75	43,75	18,75	18,75	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 3	25,00	75,00	37,50	50,00	12,50	12,50	43,75	12,50	12,50	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 4	60,00	40,00	60,00	30,00	10,00	50,00	30,00	50,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 5	33,33	66,67	66,67	16,67	16,67	16,67	50,00	16,67	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently 6	0,00	100,00	50,00	37,50	12,50	12,50	25,00	12,50	25,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 not permanently without marking	55,66	44,34	48,11	28,30	23,58	23,58	38,68	23,58	17,92	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 1	0,00	100,00	25,00	62,50	12,50	25,00	25,00	25,00	25,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 2	37,50	62,50	25,00	75,00	0,00	12,50	50,00	12,50	12,50	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 3	40,00	60,00	60,00	40,00	0,00	20,00	80,00	20,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently 4	100,00	0,00	25,00	75,00	0,00	0,00	75,00	0,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2000 permanently without marking	44,90	55,10	48,98	38,78	12,24	10,20	22,45	10,20	34,69	... ...	0,00	0,00	0,00	2,04	97,96
2002 unemployed 1	75,00	25,00	100,00	0,00	0,00	0,00	0,00	0,00	100,00	... ...	25,00	50,00	0,00	0,00	25,00
2002 unemployed 1	33,33	66,67	0,00	100,00	0,00	0,00	0,00	0,00	0,00	... ...	0,00	100,00	0,00	0,00	0,00
2002 unemployed without marking	29,46	70,54	53,49	31,01	15,50	18,60	37,98	18,60	20,93	... ...	3,88	6,20	3,10	0,00	86,82
2002 not permanently 1	47,37	52,63	36,84	48,68	14,47	9,21	52,63	9,21	7,89	... ...	38,16	31,58	15,79	11,84	2,63
2002 not permanently 2	50,00	50,00	50,00	41,67	8,33	8,33	25,00	8,33	50,00	... ...	75,00	25,00	0,00	0,00	0,00
2002 not permanently 3	12,50	87,50	37,50	25,00	37,50	25,00	37,50	25,00	12,50	... ...	0,00	0,00	12,50	0,00	87,50
2002 not permanently 4	42,86	57,14	85,71	0,00	14,29	28,57	28,57	28,57	42,86	... ...	0,00	0,00	0,00	0,00	100,00
2002 not permanently without marking	30,51	69,49	43,22	33,90	22,88	15,25	40,68	15,25	16,10	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently 1	14,29	85,71	57,14	28,57	14,29	57,14	0,00	57,14	28,57	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently 1	0,00	100,00	42,86	57,14	0,00	14,29	28,57	14,29	28,57	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently 3	60,00	40,00	60,00	40,00	0,00	0,00	100,00	0,00	0,00	... ...	0,00	0,00	0,00	0,00	100,00
2002 permanently without marking	48,28	51,72	41,38	41,38	17,24	10,34	46,55	10,34	12,07	... ...	0,00	0,00	0,00	0,00	100,00

Table 7. Dissimilarity of the rules that have identical consequence

Unemployed 1997-2000	0,19
Unemployed 1997-2002	0,23
Unemployed 2000-2002	0,27
Not permanently employed 1997-2000	0,25
Not permanently employed 1997-2002	0,29
Not permanently employed 2000-2002	0,24
Employed 1997-2000	0,40
Employed 1997-2002	0,39
Employed 2000-2002	0,21

Table 8. Similarity among the rules at time *t*

Unemployed 1997 and all other rules 2000	0,59
Unemployed 1997 and all other rules 2002	0,62
Unemployed 2000 and all other rules 2002	0,63
Not permanently employed 1997 and all other rules 2000	0,66
Not permanently employed 1997 and all other rules 2002	0,65
Not permanently employed 2000 and all other rules 2002	0,71
Employed 1997 and all other rules 2000	0,74
Employed 1997 and all other rules 2002	0,74
Employed 2000 and all other rules 2002	0,68

Table 9. Dissimilarity among the rules system in the three surveys

1997-2000	*1997-2002*	*2000-2002*
0,47	0,49	0,26

By applying formula (8), we obtain the dissimilarity among the rules system in the three surveys shown in Table 9.

So, it is possible to see how the first survey, related to a period of unemployment for Neapolitan graduates, shows anomalous accessibility rules to the labour market, while “a getting back to normality” is expressed by the poor value of the dissimilarity index relative to the comparison between the years 2000 and 2002.

7. Conclusions and future developments

We proposed a strategy, the symbolic marking, as a tool for the analysis of the evolution of a phenomenon in a given time span. We have shown it is applicable to the examined dataset and may be applied in other cases.

The proposed comparison measure may be further enriched if, for its calculation, we introduce a weighting system related to the different strength of the applied logical rules. Other future developments may derive from the possibility of simultaneous treatment of different types of variables (multinomial, modal, continuous, interval variables) without operating any previous transformations.

References

AGRAWAL R., IMIELINSKI T., SWAMI A. (1993) Mining Associations between Sets of Items in Massive Databases, *Proceedings of the ACM-SIGMOD 1993 Int'l Conference on Management of Data*, Washington D.C.: 207-216.

BOCK H., DIDAY E. (2000) *Analysis of Symbolic Data*, Springer Verlag, Heidelberg.

BOCCI L., RIZZI A. (2000) Misure di prossimità nell'analisi dei dati simbolici. In: *Atti della XL Riunione Scientifica della Società Italiana di Statistica, Sessioni Plenarie e specializzate*, Firenze, 91-102.

BRUZZESE D., DAVINO C. (2002) Post Analysis of Association Rules in a Symbolic Framework. In: *Atti della XLI Riunione Scientifica della SIS,* Milano, 63-66.

GRASSIA M.G., MURATORE M.G. (2001) I controlli nei questionari CATI dell'indagine ISTAT sulla sicurezza dei cittadini. In: FABBRIS L. (ed) *CAPTOR 2000. Qualità della didattica e sistemi computer-assisted*, CLEUP, Padova: 309-326.

GETTLER-SUMMA M. (1998) *MGS in SODAS: Marking and Generalization by Symbolic Objects in the Symbolic Official Data Analysis Software*, Cahier9935, Université Dauphine LISE CEREMADE, Paris.

Human Capital Growth for University Education Evaluation

Pietro Giorgio Lovaglio, Giorgio Vittadini[1]

Statistics Department, University of Milan-Bicocca, Italy

Summary. In this paper, we propose a method for the evaluation of relative and impact external efficiency of university studies as effects of education on the long-term income of graduates. In order to evaluate the “ceteris paribus” levels and growth we applied a multilevel longitudinal model with random effects. The variables used in our analysis refer to Human Capital studies. The Bank of Italy has collected the data used for the analysis in the years 1998, 2000 and 2002.
Keywords: External efficiency, Human capital, Multilevel growth model.

1. The external efficiency of university studies

The organizations that distribute services of public utility, as managers of collective resources, must account for their performances to the funding authorities, the community, and the customers (Hanushek, 1997).

Guidelines are being delineated, at national and international levels, for the accreditation of university education and research programmes, and services for the students. The university accreditation processes should not only meet the *ex-ante* requirements, but also verify the results.

To evaluate an educational institution, we adopt the outline put forward by Lockheed & Hanushek’s (1994). The inputs of the system determine whether the terms "efficiency" or "effectiveness" should be used. The outputs of the system determine whether the descriptors "internal" or "external" are to be applied to efficiency and effectiveness (Table 1).

---

[1] This paper is the result of the joint effort of both the authors. G. Vittadini took care of the final editing of Section 1, whereas P.G. Lovaglio was responsible for all the other Sections.

Table 1. Criteria of evaluation for university education

	Internal	**External**
Effectiveness	INTERNAL EFFECTIVENESS: University/type of bachelor programme effect on the student achievement	EXTERNAL EFFECTIVENESS: University/type of bachelor programme effect on employment Short term Human Capital
Efficiency	INTERNAL EFFICIENCY: Costs-benefit analysis of the investment in higher education	EXTERNAL EFFICIENCY: University/type of bachelor programme effect on the returns of investment in higher education Long term Human Capital

The internal effectiveness and efficiency can be used to self-evaluation of universities, while the external ones may be the most proper way for evaluating the universities' performances (Elias, 2002). As of today, there are very few studies of the analysis of external efficiency of education based on the evaluation of the earned income of graduates in the vital cycle. In the following, we will define and apply the criteria for the evaluation of organizations that distribute public utility services.

The investment in higher education may be evaluated in terms of the graduates' earned income. In other words, we can evaluate the so-called Human Capital (HC) of an individual in terms of his or her expected earned income related to the competences achieved through education (Becker, 1964).

In the following, the evaluation of the university education does not coincide with the estimation of HC, but with the following aims: to evaluate "ceteris paribus" the universities' investment in HC with returns in terms of earned income (relative efficiency), and if and how the university education is a real advantage in monetary terms (impact of efficiency).

2. Gain score, added value, and multilevel growth model

In order to overcome the problem of "selection bias" (Garen, 1984) due to lack of randomization of the individuals between treatments, the effectiveness and efficiency must be evaluated *ceteris paribus*, i.e. adjusting the outcome for the different individual characteristics, university resources, and local job markets (Fitz-Gibbon, 1997; Scheerens & Bosker, 1997).

To this aim, even if a linear model may be applied, the nature of the dependent variable suggests non-linear approaches. To this purpose, we introduce the concepts of initial performance or pre-test (y_1), final status or post-test (y_2), as well as the concept of improvement or gain (y_2-y_1).

In the *gain score* approach the "adjusted outcome" is simply the expected value of gain, $E(y_2\text{-}y_1)$, whereas the added value or *residual gain score* approach[2] is concerned with the quantity $y_2\text{-}E(y_2/y_1)$, i.e. the difference between the post-test and its expected value $E(y_2/y_1)$, a result of the regression of y_2 onto y_1.

Both approaches present some drawbacks. Some researchers highlight the technical problems associated with the residual gain score approach (Cronbach & Furby, 1970; Scheerens & Bosker, 1997; Hanushek, 1997, 2002; Willett, 1988; Thum, 2002). First, it is reasonable to expect that the same set of factors affecting the pre-test would affect the post-test as well. Like the post-test, the pre-test is measured with error, so that a regression of post-test on a pre-test score violates the basic regression assumption that covariates are error-free. This implies that the residual gain score is an inefficient and inconsistent estimator of true change. Secondly, because the pre-test typically explains a large quota of the post-test outcome, this drastic reduction of residual variance limits the explicative power of other factors, for example the resources of the organization, its quality and so on.

Other researchers (Rogosa *et al*., 1982; Rogosa & Willet, 1985; Rogosa, 1995; Singer & Willet, 2003) dispel the notion that gain scores are inherently unreliable, so demonstrating that the correlation between gain and initial status (the major drawback of the gain score approach) is merely an artefact of the design. The nature of the correlation is the outcome of the parameterization we choose for the growth factors and, as a consequence, the meaning itself we attach to them.

Thum (2002) and Bryk *et al*. (1998) propose to measure the outcomes in time series in a longitudinal framework, an approach that allows insertion of covariates related to the institutions that can affect the performances of the micro units (students, classes). The repeated measurement approach we adopt in this paper places the pre-test and the post-test on an equal footing in that both serve as outcomes in our models. The change with respect to gain scores is evident in recent research (Collins, 1996; Maris, 1998; Mellenbergh, 1999; Williams & Zimmerman, 1996).

The second development, known as "hierarchical", "multilevel", "growth" or "mixed-effects" modelling, provides the methodological complement for a proper, flexible, treatment of growth data in stratified sampling designs that are common in educational research settings (repeated measures on students over time, students nested with schools in the district, etc.).

---

[2] Its origins can be traced to notions of goods, value and wealth in political economy (Saunders, 1999); value-added analyses address a model for comparison (Goldstein & Spiegelhalter, 1996) adjusted for the factors out of control for the researchers. Researchers have frequently applied the regression analysis of the post-test score on the pre-test score, motivating this choice with the notion of "fairness" in order to flatten the playing ground (Good *et al*. 1975).

Multilevel models facilitate the study of the effects of covariates of development at every level of nesting (Laird & Ware, 1982; Bryk & Raudenbush, 1992; Goldstein, 1995). The advent of multilevel models helped accountability agencies to avoid the misleading picture conveyed by aggregate scores, such as the school or district means, and focus on student level and/or change (Meyer, 1996). Many current accountability systems adopt multilevel models (Bryk *et al*., 1998).

A Multilevel Growth Model (MGM) analyses random effects in the sense that intercept and slopes describing individual trajectories over time vary between individuals (Bryk & Raudenbush, 1992). Such a model requires a specification of the covariance structure between observations at different times on the same subject, because it is plausible to assume that observations on the same unit are correlated (Willett, 1988).

There are two fixed effect factors, treatment and time. Random effects result from variation between subjects and from variation within subjects. Measures on the same subject at different times are usually correlated, with measures taken close in time being more highly correlated than measures taken far apart. Observations on different subjects are assumed independent, although the validity of this assumption depends on the study design. Multilevel models are used with repeated measures data to accommodate the fixed effects of treatment and time and the covariation between observations on the same subject at different times.

In a MGM, we will conceptualize the repeated measures as observations ‘nested’ within individuals. To start, we must again specify our model for the individual trajectory, which we will refer to as the Level 1 (or within-person) model for intra-individual change:

$$y_{tij} = \pi_{0ij} + \pi_{1ij}\, time_{tij} + r_{tij} \quad , \tag{1}$$

where: y_{tij} represents the *educational outcome* at time t (t=1, .., k) for the i-th subject (i = 1,…, n_j) nested in the j–th institution (university, school, type of programme) (j = 1,…, m); $time_{tij}$ is the variable time having cardinality equal to the number of instants; π_{0ij} and π_{1ij} are the intercept (mean of outcome at $time_{tij}$=0) and slopes (rate of increase over time) for subject ij. In equation (1), where for the sake of simplicity we did not insert covariates varying over time, it is assumed that the outcome follows a linear trend.

We also specify a Level 2 (or between-person) model for inter-individual differences. The Level 2 model decomposes the individual intercept and slopes (π_{0ij} and π_{1ij}) into the mean values of the parameter institutions across individuals (β_{00j} and β_{10j}), in the effect of a set of individual characteristics (x_{sij} with α_{sj} associated parameter) of subject ij and the individual effect (u_{0ij} and u_{1ij}).

Finally, the Level 3 model decomposes the mean effect of institutions into the mean values of the parameters across all institutions (γ_{000} and γ_{100}), in the effect of a set of characteristics (g_{rj} with associated parameter λ_r) of institution j and institution effect (m_{00j} and m_{10j}):

$$\pi_{0ij} = \beta_{00j} + \Sigma_s \alpha_{sj} x_{sij} + u_{0ij} \quad \pi_{1ij} = \beta_{10j} + \Sigma_s \alpha_{sj} x_{sij} + u_{1ij}, \tag{2}$$

$$\beta_{00j} = \gamma_{000} + \Sigma_r \lambda_r g_{rj} + m_{00j} \quad \beta_{10j} = \gamma_{100} + \Sigma_r \lambda_r g_{rj} + m_{10j}. \tag{3}$$

Each equation presents errors at all nested levels: r_{tij} is the random effect of time t on the trajectory of outcome evolution for subject ij; u_{0ij} and u_{1ij} are the random effects of subject ij with respect to the mean trajectory for the subjects nested in institution j; m_{00j} and m_{10j} are the random effects of institution j with respect to the mean trajectory for all institutions.

The distribution of the time specific disturbances, individual trajectory parameters and institutions' specific parameters are assumed normally distributed, with zero expectation, and mutually independent:

$$\begin{bmatrix} \mathbf{r_{tij}} \\ \mathbf{u_{\cdot ij}} \\ \mathbf{m_{\cdot\cdot j}} \end{bmatrix} \sim \mathbf{N}\left(\begin{matrix} \mathbf{0} \\ \mathbf{0}, \\ \mathbf{0} \end{matrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix} \right) \tag{3}$$

where $\mathbf{\Sigma}$ is the (k,k) covariance matrix of the time-specific disturbances (r_{tij}), $\mathbf{T}$ the (2,2) covariance matrix of the individual trajectory parameters ($\mathbf{u_{\cdot ij}}$) and $\mathbf{G}$ the (2,2) covariance matrix of the trajectory parameters between institutions residuals ($\mathbf{m_{\cdot\cdot j}}$). Even if MGMs neglect the specification of the covariance between observations at different times on the same subject, a special attention must be paid to $\mathbf{\Sigma}$ otherwise we risk incorrect conclusions on fixed effects (Rogosa, 1995; Laird & Ware, 1982; Cnaan *et al.*, 1997).

The model specified with formulas (1), (2) and (3) is applicable for estimating the specific effect of institutions on the level (in a temporal instant, m_{00j}) and on the rate of outcome increase over time (m_{10j}). The model can be extended (Lovaglio, 2005) to the case of the evaluation of the efficiency among institutions (relative evaluation) and to that of the distributed service (impact evaluation).

In the latter case, the model allows to check if an institution that distributes a service (higher education) obtains better performances in comparison with a (control) group of subjects with similar characteristics to whom the service was not distributed. To this aim, a dummy variable I_{ij} – which assumes value 0 for subject i of the control group and 1 otherwise – is inserted in equations (2) and (3) for the estimation of the impact level (κ_0) and slope (κ_1):

$$\boldsymbol{\pi}_{0ij} = \beta_{00j} + \Sigma_s \alpha_{sj} x_{sij} + \boldsymbol{\kappa}_0 \, \boldsymbol{I_{ij}} + u_{1ij} \qquad \boldsymbol{\pi}_{1ij} = \beta_{10j} + \Sigma_s \alpha_{sj} x_{sij} + \boldsymbol{\kappa}_1 \, \boldsymbol{I_{ij}} + u_{1ij} \tag{5}$$

$$\beta_{00j} = \boldsymbol{\gamma}_{000} + \Sigma_s \boldsymbol{\lambda}_r g_{rj} + I_{ij}\, m_{00j} \qquad \beta_{10j} = \boldsymbol{\gamma}_{100} + \Sigma_s \boldsymbol{\lambda}_r g_{rj} + I_{ij}\, m_{10j} \tag{6}$$

3. The external efficiency of Italian universities

The proposed methodology was applied to the evaluation of the external efficiency of Italian university education using the data collected by the Bank of Italy with the Survey on Household Income and Wealth (Banca d'Italia, 2002) for the years 1998, 2000 and 2002.

The dependent variable is the earned income for each individual in a time series and the covariates are individual factors (social-demographic-economic background), university covariates (input, resources, intangible aspects, personnel) and job market variables (region, province, etc. where the subject is addressed).

Several indicators concur to identify the individual human capital and household socio-economical status: Age, Gender, Region of residence, Marital status, Educational level, Employment status, Type of job, Economic activity sector, Years of schooling, Age of entrance in the labour market, Number of children; Years of full-time work, Years part-time work, Household total wealth, Household total debt, Household income, and Parents' educational level.

The outcome of the analysis is the partial disposable earned income (from here on *earned income*) composed of earnings of employed and self-employed people, pensions, transfers and economic assistance.

The sample was composed of subjects belonging to the labour force with positive earned income for all three years; we identified the graduates as individuals who had achieved a university bachelor or master degree before 1998.

The panel of income earners is composed of less than 2000 subjects, whereas the group of graduates of 559 individuals (Table 2). Because of the reduced number of panel graduates, it is not possible to deepen our analyses on the external efficiency by university, or type of degree. The analysis is therefore limited to the effect of the higher education *tout court* on the income dynamics (impact analysis that compares the level and the growth rates of graduates and non-graduates earnings in Italy)[3].

Table 2. Composition of the Bank of Italy panel sample

Period	*Families*	*Individuals*	*Income earners*	*Income earners Graduated after 1998*
1998	7147	20901	7635	764
2000	8001	22268	8140	825
2002	8011	21148	7313	793
1998 - 2000 – 2002	2591	6953	1978	559

[3] Such data moreover prevent the estimation of the short-term university education effect, because we must select the graduates who have recently acquired a degree, typically going back three to five years from the first survey.

Table 3. Means of the earned income (in Euro) by degree level and time

Period	*Graduate*	*Non Graduate*
1998	17,959	13,108
2000	20,700	14,269
2002	22,024	15,429

In order to estimate income temporal trend and the between-group differences, a MGM is specified for evaluating:

a) the temporal correlations for every subject (covariance structure);
b) the existence of temporal dynamics for earned income trajectories (time effect);
c) if such a trend differs between graduates and non-graduates (bachelor effect);
d) the individual characteristics that explain the levels and the growth rates over time.

The specified longitudinal model is therefore a two-level model composed of equation (1) and (5) describing, respectively, Level 1 (time effect) and Level 2 (subject effect). We highlight that both level and growth rate for graduates are larger than for non-graduates (Table 3).

We can state that the hypotheses on which the classical approaches of repeated measures (Anova and Contrast Analysis) are based (constancy of covariance between couples of measures when the temporal lag varies, and constancy of variances in time) are too rigid. A correct modelling of the covariance structure $\mathbf{\Sigma}$ - representing it as a function of a relatively small number of parameters - is fundamental for valid inference and estimation of fixed effects (impact of the bachelor study programme on the level and growth rates of earned income). For most of the structures, the covariance between two observations on the same subject depends only on the lag between measurements with constant variance.

Estimates of between-graduate variances (diagonal elements) and correlations within groups at each period are shown in Table 4, referring to the MGM specified with the effects of treatment, time, treatment-by-time interaction, and other significant covariates (Table 5). The estimated parameters show that variances between graduates within each group are approximately equal across times and correlations decrease with increasing interval between observational times.

Table 4. Estimated correlations and variances of incomes over time

Time	*1998*	*2000*	*2002*
1998	0.34	0.50	0.39
2000	0.50	0.32	0.49
2002	0.39	0.49	0.39

Table 5. Effect of covariates on earned income at 1998 and on growth 1998-2002

Level (Income at 1998)	*F Value*	*Pr > F*		*Growth rate*	*F Value*	*Pr > F*
Intercept	173.61	<.0001		*Intercept*	123.56	<.0001
Gender	161.90	<.0001		*t* Occupation*	99.29	<.0002
Occupation	80.29	<.0002		*t*Cl_age*	66.34	<.0003
Bachelor	58.43	<.0003		*t*Sector*	23.56	<.0006
Full_job	14.65	<.0005		*t*Bachelor*	3.83	0.0218
Sector	8.51	<.0006		*t*Gender*	1.68	0.0932
Marital	4.43	0.0042		*t*Full_job*	1.65	0.1021
Area	3.85	0.0039		*t*Area*	1.15	0.1362
Cl_age	3.09	0.0150		*t*Marital*	1.12	0.1423

Akaike's information criterion and Schwarz's Bayesian criterion are the indices of relative goodness-of-fit chosen for comparing models with the same fixed effects but different covariance structures. The covariance structure assumed for $\mathbf{\Sigma}$ is the first order autoregressive one (the estimated correlation parameter for unitary lag is 0.373), which specifies homogeneous variance and covariances between observations that decrease toward zero with increasing lag. The test of heterogeneity indicates a common covariance structure for graduates and non-graduates.

Once $\mathbf{\Sigma}$ is structured, the longitudinal analysis requires the construction of the Unconditional Means Model (UMM), a model that does not require covariates at each level (level 1, level 2). It is useful to both estimate whether a systematic variability exists in the outcome, and show how the total variance of the outcome is decomposed between and within subjects, depending on the time or on individual factors that changed over time.

The estimated intraclass correlation (the ratio of the between and total variances) shows that the 43% of the total variability of income is attributable to the differences between the individuals (i.e. individual characteristics) and for the remaining part to the differences within the subjects, confirming the hierarchical structure (multilevel) of the dataset.

Inserting the variable time (as a continuous quantitative covariate) in the UMM we obtain the Unconditional Growth Model (UGM) that explains how much of the within subject variance is explained by time. We estimated that 28% of the within-subject variability is determined by time and the remaining part by individual characteristics that change over time (occupation, sector, etc).

Relative to the fixed effects, the overall growth rate of earned income is positive and highly significant, so indicating a strong dynamic effect on the income trajectories, while, relatively to random parameters, the variance of income in 1998 (the reference time) and the variance of the growth rates - between the subjects- are significantly different from zero. The estimated covariance between the levels of incomes in 1998 and growth rates is not significant, and this implies unstable relations between slopes and initial income.

In order to explain the residual variability between the levels (earned incomes in 1998) and between the growth rates (yearly increment of income), we inserted individual covariates in the 2 level model (equation 5): some of which, for the sake of simplicity, are kept fixed in time and measured at the final year (2002).

The covariates and the significance levels of the associated parameters (fixed effects)[4] are shown in Table 5. The level of earned income at 1998 (left part of Table 5), gender, type of occupation at 2002 (Occupation), possession of a bachelor's degree (Bachelor), years of full-time job (Full_job), occupational sector at 2002 (Sector), area of residence at 2002 (Area), marital status at 2002 (Marital), and age at 2002 (Cl_age) are highly significant.

The factors that characterise the evolution of the growth rates are (right part of Table 5): type of occupation at 2002 (t*Occupation), age at 2002 (t*Cl_age), occupational sector at 2002 (t*Sector) and, to a lesser extent, the possession of a bachelor's degree (t*Bachelor), whereas the other covariates are not significant at 5% level. Relatively to the random parameters (not shown here), the residual variances in the 1998 level and growth rates remain significant.

We can conclude that income level at 1998 and income growth rates are greater and faster increasing (even though close to the limits of the customary significance levels) for graduates. The selected covariates explain a large quota (56%) of the UGM residual variance for the level and just a part (21%) of the between-subject growth rate variability, thus leaving open the search of other individual covariates that may affect the evolution of income over time. The trajectories estimated for each group of subjects (after adjustment for the selected covariates) give a mean growth annual rates of 1,051€ for a graduate and 426€ for a non-graduate.

Moreover, we estimated the means of growth rates for the two groups for the more relevant age class. By comparing the growth rates of a graduate (1507€) and a non-graduate (916€) in the age class less of 30 years it is possible to quantify how the labour market rewards university education. The impact of a university degree appears highly significant in absolute terms, while in comparison with the non-graduate rates this difference remains significant, but constant in the older age groups (606€ for 31-40 year group; 584€ for 41-50 group).

Finally, we estimated the differences between group income means at every time and for the income growth rates. The contrast analysis shows that for all

---

[4] The application of QQ plot test for Level 1 and Level 2 residuals shows that normality is supported by data.

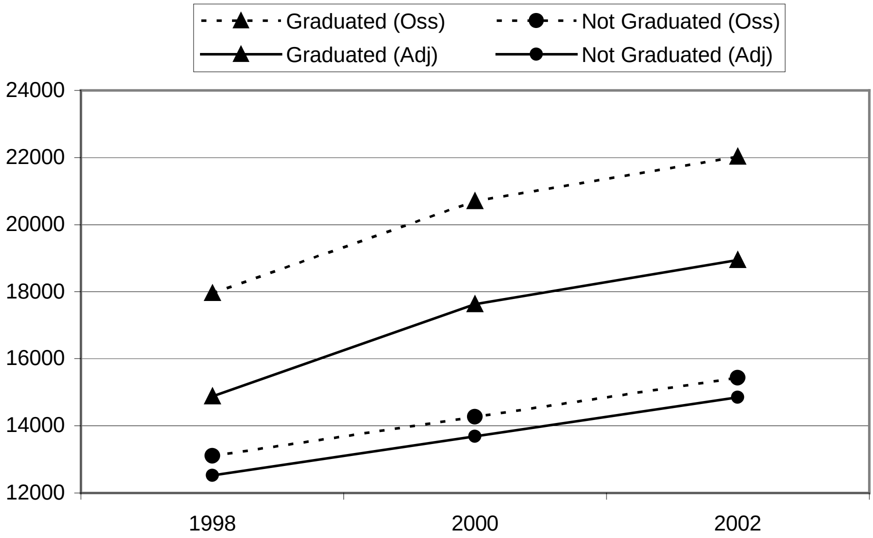


Figure 1. Observed and estimated income trajectories

three temporal moments the differences between income means are not only meaningful, but increasing in time (in particular, the difference is 1188€, 2015€ and 2118€, respectively for 1998, 2000 and 2002) and that the income growth rate of the graduates (estimate: 525€) is larger than that of non-graduates (*p*-value=0.0169).

Figure 1 shows the incomes trajectories for graduates and non-graduates, relatively to the average observed means (*Oss*) and to the average estimated means (*Adj*). The continuous lines, which show the income trends adjusted for the subjects' characteristics, confirm our previous results. The adjusted trajectories make it evident the "correction" effect (also called *risk-adjustment*).

4. Conclusions

The proposed methodology performs "ceteris paribus" comparisons, so allowing a better inference in the sense that the impact of the university degree is estimated net of the different group characteristics.

In fact, controlling for the possible factors of heterogeneity, the difference of the two trajectories is reduced, because graduates possess individual characteristics more coherent with the ability to produce high salaries (area of residence, household context, remunerative occupation, employment sector, etc.), as illustrated by the fact that the adjusted trajectory for graduates lowered drastically.

References

BANCA D'ITALIA (2002) I bilanci delle famiglie italiane, *Supplementi al Bollettino Statistico.*

BECKER G.S. (1964) *Human Capital*, Columbia University Press and NBER, New York.

BRYK A.S., THUM Y.M., EASTON J.Q., LUPPESCU S. (1998) Assessing school academic productivity: The case of Chicago School Reform, *Social Psychology of Education*, **2**: 103-142.

BRYK A.S., RAUDENBUSH S.W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publication, Newbury Park, CA.

COLLINS L.M. (1996) Is reliability obsolete? A commentary on "Are simple gain scores obsolete?", *Applied Psychological Measurement*, **20**: 289/292.

CNAAN A., LAIRD N.M., SLASOR P. (1997) Using the General Linear Mixed Model to analyse unbalanced repeated measures and longitudinal data, *Statistics in Medicine*, **16**: 2349–2380.

CRONBACH L.J., FURBY L. (1970) How should we measure "change" or should we?, *Psychological Bulletin*, **74**(1): 68-80.

ELIAS G. (2002) L'accreditamento del processo e la misura del risultato. In: Workshop "*Valutazione dell'Università, accreditamento del processo, misurazione del prodotto*", Franco Angeli, Università Milano Bicocca.

FITZ-GIBBON C.T. (1997) *The Value-Added National Project: Final Report. Feasibility Studies for a National System of Value-Added Indicators*, SCAA Publication, Hayes Middlesex.

GAREN J. (1984) The returns to schooling: a selectivity bias approach with a continuous choice variable, *Econometrica*, **52**, 23-54.

GOLDSTEIN H. (1995) *Multilevel Statistical Models, II ed.*, Edward Arnold, London.

GOLDSTEIN H., SPIEGELHALTER D.J. (1996) League table and their limitations: statistical issues in comparisons of institutional performances (with discussion), *Journal of Royal Statistical Society*, **159(5)**: 385-443.

GOOD T.L., BIDDLE B.J., BRODY J.E. (1975) *Teachers Make a Difference*, Holt, Rinehart and Winston, New York.

HANUSHEK E.A. (1997) Assessing the effects of school resources on economic performance, *Education Evaluation and Policy Analysis*, **19**(2): 141-164.

HANUSHEK E.A. (2002) Publicly provided education. In: AUERBACH A., FELDSTEIN M. (eds) *The Handbook of Public Economics*, IV.

LAIRD N.M. WARE J. H. (1982) Random effects models for longitudinal studies, *Biometrics*, **38**: 973-974.

LOCKHEED M.E., HANUSHEK E.R. (1994) Concepts of educational efficiency and effectiveness, *International Encyclopedia of Education, II Edition.*

LOVAGLIO P.G. (2005) Efficacia relativa e di impatto di iniziative nell'ambito dei servizi alla persona di pubblica utilità, *Statistica*, **1**: 1-18.

MARIS E. (1998) Covariance adjustment versus gain score-Revisited, *Psychological Methods*, **3**: 309-327.

MEYER D. (1996). *The Power of their Ideas: Lessons for America from a Small School in Harlem.* Boston: Beacon Press.

MELLENBERGH G.J. (1999) A note on simple gain score precision, *Applied Psychological Measurement*, **23(1)**: 87-89.

ROGOSA D.R. (1995) Myths about longitudinal research. In: GOTTMAN J.M. (ed), *The Analysis of Change*, LEA, Mahwah, NJ: 3-66.

ROGOSA D.R., BRAND D., ZIMOWSKI M. (1982) A growth curve approach to the measurement of change, *Psychological Bulletin*, **90**: 726-748.

ROGOSA D.R., WILLETT J.B. (1985) Understanding correlates of change by modeling individual differences in growth, *Psychometrika*, **50**: 203-228.

SAUNDERS L. (1999) A brief history of educational "value added": how did we get to where we are?, *School Effectiveness and School Improvement*, **10(2)**: 233-256.

SCHEERENS J., BOSKER R.J. (1997) *The Foundations of Educational Effectiveness*, Pergamon, Oxford.

SINGER J.D., WILLET J.B. (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, Oxford.

THUM Y.M. (2002) Measuring progress towards a goal: estimating teacher productivity using a multivariate multilevel model for value-added analysis (http://www.mff.org/pubs/tchprodv3.pdf).

WILLETT J.B. (1988) Questions and answers in the measurement of change, *Review of Research in Education*, **15**: 345-422.

WILLIAMS R.H., ZIMMERMAN D.W. (1996) Are gain scores obsolete?, *Applied Psychological Measurement*, **20**: 59-69.

Estimating University Human Capital through Growth Models

Marisa Civardi, Emma Zavarrone[1]

Department of Methods of Economics, University of Milano - Bicocca, Italy

Summary. Our paper focuses on the law of growth of the human capital deriving from the evaluation of undergraduates' human capital due to university education. For this purpose, we introduce the definition of University Human Capital (UHC), a kind of human capital that sums up to the other kinds of human capital and that acts, for the concerned companies, as a detector of competences owned by graduates. It follows that UHC can be interpreted also as a component of the "intellectual capital" that characterizes the different kinds of enterprises. UHC individual growth trajectories are to be established by means of two-level growth models. We attempt to synthesise the law of individual UHC growth through both a logistic and a Gompertz function.
Keywords: Latent growth curves; Human capital; Specific competences; Multilevel models; Gompertz function.

1. Introduction

All the theories on Human Capital (HC) share a macroeconomics approach that raises questions about the suitability of the analysis of HC. Indeed, at a macroeconomics level, the focus of such an analysis is on assessing the general relevance in establishing the contribution to the national wealth. At a microeconomics level, instead, the focus is on quantifying the internal companies' HC. A microeconomics approach brings about the need to determine the capital identifying the company market value, the so-called *Intellectual Capital* (IC) (Lev, 2000).

---

[1] In this joint work, M. Civardi was responsible for the final editing of Sections 1, 2, 3 and 4, whereas E. Zavarrone was responsible for Sections 4.1, 5 and 6.

The rapidly changing job market requires graduates to hold a good post-graduate degree and the ability of being flexible for an effortless and a rapid vocational rehabilitation. University education and flexibility are not only two aspects of company IC, but also the value added by universities to their graduates. Therefore, identifying the HC that can be offered by university is crucial if an analysis and a quantification of the level of competence required by companies are being carried out.

In this perspective, a microeconomics approach to determine and measure UHC – that is the HC increase in terms of competencies for work by university – is a three-phase process. In the first phase, HC is assessed according to a university perspective. For each graduate the HC increase is quantified assessing the competence acquired by graduation.

In the second, HC is analysed according to a company's perspective, which implies that the required HC typologies are identified.

In the last phase, a feedback is dealt with: procedures and strategies developed by university to adapt higher education to market demand are analysed.

2. From Human Capital to Intellectual Capital

In order to adopt a microeconomics perspective to quantify UHC, an agreement on terminology is required to identify the common terms shared by universities and companies. To start with, while companies consider IC an intangible asset, universities regard it as the increase in graduates' knowledge by way of the learning process, which essentially transforms a secondary school graduate into a university graduate.

From the university viewpoint, UHC can be defined as the difference between students' HC when entering university and the HC held at graduation. In other words, UHC represents the improvement of knowledge, skills and attitudes attained thanks to the educational activities, use of didactic structures, exams passed and social interaction with fellow students. Measuring this improvement can be an aid to assess university ability to produce competences, depending on the variety of students' HC at university enrolment.

A first simplified representation of the higher education process can involve only two actors: users of education (students) and providers of education (university). A more realistic approach needs to take into consideration the existing relations between students' HC (family background, social background, and educational background), the education provider and companies in the job market. Obviously, the choices about education made by family, as well as all interactions with the social environment, crucially contribute to identify students' HC typology at the enrolment time (human capital at time t_0: HC_0).

During the time spent at the university, the initial HC_0 changes: it modifies and specializes according to the education received, the experiences, and the opportunities taken. It is exactly during higher education that HC specifies itself into University Human Capital. This kind of human capital assumes specific relevance and distribution in each student depending on his/her speed of learning and time spent at the university. As a result, at the end of the educational process, this component, which plays a differential role in the recruitment selection process, characterizes the graduates' human capital.

Moreover, IC can be considered as the HC component that characterizes company typologies. In fact, even though it is never included into companies' balance, IC is a company's intellectual asset and is at least as basic as capital assets. Undoubtedly, in today's economy, knowledge and information increasingly influence national wealth and consequently reduce the input of manufacturing.

Developing and managing knowledge in conjunction with intellectual capital is the most effective way through which companies can develop a competitive advantage: company's value and company's performance depend on its ability to create, gather and share knowledge. Companies, therefore, measure IC as the difference between its own market value and the cost of assets replacement.

Human capital, organizational capital and relational capital are the three main IC components. Organizational capital consists of company culture, company philosophy and information systems while relational capital consists of all relations with customers, suppliers and the environment/area/region. The main contributors to the development of human capital are, on the one hand, the knowledge, the capacity and the ability of each individual to provide clients with appropriate solutions, and on the other, the company overall ability to choose the most appropriate solution among all those proposed.

According to Edvisson & Malone (1997), *know-how* concerns two distinct aspects: knowledge and experience. With reference to graduate students, knowledge is anything formal and informal offered to the individual throughout university education. The cumulative experiences, regardless of their nature (deliberate or not, documentary, verbal, tacit), are the main sources of influence on learning and individual behaviour. In turn, individual performance is the result of the causal link between the two competences mentioned above through a circle linking motivations, characteristics, behaviour and result. The result influences in its turn motivation and characteristics.

3. UHC quantification

Assuming that some features of the above-mentioned HC characterize IC a whole, as a result, UHC can be considered as a very important element connecting university, which trains professional figures, and the job market, which searches them. In other words, UHC becomes the link between students, universities and companies.

Briefly, while students aim at gaining the credits required to graduate within the given time of their studies, university aims at providing effective educational services for training professional figures qualified to fill in the vacancies offered by companies. From this viewpoint, specifying HC in UHC represents the characterizing and decisive key element when companies select new employees or partners.

Let us now define the measure of UHC adopted in this work and its characteristics. UHC is the increase of students' competence acquired through higher education, in a dynamic context. This process consists of:

a. an *individual component* related to:
   1. different initial HC,
   2. different UHC increase paths,
   3. the interaction between individual UHC levels and the time spent to obtain them,
b. a *group component* that expresses the influence of the system on the individual component (such as the chosen study programme).

For a graduate *j*, the value of accumulated UHC at the time *t* is determined by:

$$UHC_j(t) = \frac{100\sum_{i=0}^{t} uhc_j(i)}{\max\ UHC_j(T_j)} \qquad t=1,2,3,\ldots,T, \qquad [1]$$

where: $uhc_j(i) = \sum cfu_j(i) \times mark_j(i)$ represents the global credits gained by *j* at the time *i* (the time when the examination result was registered by Student services) weighted with the respective marks. Hence, $uhc_j(i)$ represents the graduate *j* UHC increase at the observation *i*.

The denominator of [1] expresses the theoretical maximum for the subject *j* at the time T_j, that is, the time within which *j* obtains his degree. If no marks are added for first-class honours, the maximum for a three years degree graduate is: max $UHC_j(T_j) = 180 \times 30$. Therefore: $0 \le UHC_j(T_j) \le 100$.

4. Methodology

The individual component, which depends on individual students' UHC accumulation paths (component a.2), can be evaluated by assessing the parameters of appropriate time functions. Clearly, as this involves both a constant growth rate *k* and absence of an upper limit $\lim_{t \to \infty} f(t) = \infty$, the linear function $f(t) = \beta + kt$ is not the most appropriate to consider (Rao, 1958).

More realistically, by assuming an upper value α of UHC, several options are possible among which three are chosen. In the first one, the growth rate at time t is directly proportional to total growth to be still achieved, where the proportionality factor k>0. Under these conditions, the following growth curve is identified as a modified exponential growth curve:

$$f(t) = \alpha\left(1 - \beta e^{-kt}\right) \qquad \beta > 0 \quad . \qquad [2]$$

In [2], α is not only a function scaling factor, but it is the horizontal asymptote. Therefore, it is also UHC maximum value. In our case, α = 100. Parameter k controls the scale along the time axis and, being UHC initial value UHC(0)=$\alpha(1-\beta)$, β determines the intersection between curve and vertical axis. If α is known, the transformed function $g(t)=f(t)-\alpha$ is an exponential function whose parameters can be estimated by the least squares method.

Provided the limit growth value equals α, the second option assumes that the growth rate at the time t is directly proportional to the total growth to be still achieved by the proportionality factor k>0. This leads to the following three-parameter *logistic function*:

$$f(t) = \frac{\alpha}{1 + \beta\, e^{-kt}} \qquad \beta > 0 \quad . \qquad [3]$$

This curve has two horizontal asymptotes: the lower is the time axis, the upper one is α, that is still UHC maximum value. The initial value is: $UHC(0)=\alpha/(1+\beta)$. In addition, the curve presents an inflection point.

As in [3] $f(t)$ reciprocal is a modified exponential function, introducing the function $g(t)=1/f(t)$ brings the identification of the logistic function back to identify the modified exponential function.

The third option, assumes that, given the upper limit α, time t growth rate is proportional to the difference between the maximum achievable level logarithm and the logarithm of the level achieved at the time t, with a proportional factor k>0. The corresponding function is *Gompertz curve*:

$$f(t) = \alpha\, exp\left(-\beta e^{-kt}\right) \qquad \beta > 0\,. \qquad [4]$$

In Gompertz curve both the time axis and α are horizontal asymptotes. Therefore, α is again UHC maximum value, while the initial value is: $UHC(0)=\alpha\, e^{-\beta}$. In the Gompertz function, $log\, f(t)$ is expressed as a modified exponential function, hence, introducing the function $g(t)=log(f(t))$ brings the identification of its parameters back to identifying modified exponential function.

The three growth curves described above represent respectively three behavioural models for gaining the university credits required to obtain a degree. The assumption underlying the Italian Ministry of Education, University and Research supply of higher education is that students' yearly accumulation rate of credits in each of the three studying years, should be constant. In fact, this is not the case: few students obtain their degree within the expected time and when they do, a constant accumulation rate is very rare.

As already proved, the modified exponential function involves growth rates, and therefore number of credits gained in a unit time slot, decreasing in any time. According to this behavioural rule, students should gain more credits in their first year than in the following years. Unsurprisingly, its numerical application leads to the rejection of the model.

The other two models seem to be definitely more appropriate. In the second model, the hypothesis of proportionality between the relative growth rate and that to be achieved is equivalent to the assumption that the ratio between the credits gained in a time unit slot and the total credits at the beginning of the slot is proportional to the credits still to be gained out of the 180 required to obtain a degree. In the initial period of their studies, that is during the first examination sessions, students gain credits slowly while they speed up during the central periods.

According to Gompertz model, the ratio between credits gained during the generic time unit slot and the total credits at the beginning of the slot, that is the relative growth rate, is a decreasing linear function of the gained credits logarithm. Like the previous model, this also shows a slow start, but with an earlier acceleration. As we will shown below, both models fit the observed data.

4.1 Two-level growth linear models

In order to consider the effects of the interaction between UHC total amount (or its appropriate transformation) and the time spent achieving it, a 2-levels linear growth model can be referred to, where level 1 expresses the individual growth model. Because of what has been said above, three different growth models have been set up. They are based on the three individual UHC transformations introduced above:

UHC=log(100-UHC)* →	growth model expressed by the modified exponential curve;
UHC=log(1/UHC-1/100)* →	growth model expressed by the logistic function;
UHC=log [log100-logUHC]* →	growth model expressed by the Gompertz curve.

If π indicates the model parameters for level 1 ("within" students) and β the parameters for level 2 ("amongst" students):

$$UHC^{*}_{ij} = \pi_{0j} + \pi_{1j} t_{ij} + r_{ij} \qquad r_{ij} \approx N\left(0, \sigma^{2}\right) , \qquad [5]$$

where

$$\pi_{0j} = \beta_{00} + u_{0j} \; ; \qquad \pi_{1j} = \beta_{10} + u_{1j} \quad , \qquad [6]$$

and

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \approx N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \; . \qquad [7]$$

Equation [5] expresses the variable $UHC_{ij}*$ for each student j at time i as a linear function of time. Parameters π_{0j} e π_{1j} correspond respectively to the mean value at initial time and to $UHC_{ij}*$ growth rate. Equations [6] indicate that for student j $UHC_{ij}*$ value at initial moment (π_{0j}) depends on the average UHC_i* of all students' (β_{00}) and on a specific casual effect of student j (u_{0j}). Likewise, it indicates that the growth rate (π_{1j}) depends on the mean growth rate of all students (β_{10}) as well as a specific casual effect of student j (u_{1j}).

The presence of random errors on both levels shows the random nature of the model. In particular, as shown in [5], r_{ij} expresses instant time i casual effect on $UHC_{ij}*$ growth curve while, as said above, u_{0j} e u_{1j} are the random effects on student j growth curve in relation to all students' average one, respectively by intercept and growth rate.

Merging the two levels in one equation generates the following:

$$UHC_{ij} = \left[\beta_{00} + \beta_{10}t_{ij}\right] + \left[u_{0j} + u_{1j}t_{ij} + r_{ij}\right] \; . \qquad [8]$$

Equation [8] indicates that the multilevel model can be expressed as the sum of two parts. The first includes two fixed effects, one for intercept and the other for time parameter - growth rate. The second consists of three random effects: on intercept, on t slope, and on the residuals within students, r_{ij}. In this formula, both intercept and slope are assumed to be random. Additionally, at level 2, there is no variable. These assumptions can be both modified (Singer, 1998).

5. Application

The application was conducted on a cohort of University of Milan-Bicocca Faculty of Economics students, registered in the academic year 2000-2001. The career of each student was observed until December 31, 2004, provided everyone graduated within the observation period. Hence, the cohort was observed for 48 months.

The following characteristics were collected for each student:

- first examination mark (VE_1)
- credits associated with first examination (CE_1),
- first examination date (mm/yy) (DTE_1)
- second examination mark (VE_2)

– credits associated with second examination (CE_2)
– second examination date (mm/yy) (DTE_2)
– ....
– ....
– last examination mark (VE_n)
– credits associated with last examination (CE_n)
– last examination date (mm/yy) (DTE_n)
– final oral examination (graduation) mark (VL)
– credits associated with final oral examination (CL)
– final oral examination date (mm/yy) (DTL)

According to Student secretariat, the initial database contained, for each student, n_j+1 records (one record for the student' n_j examination, plus the final oral examination record). Each of n_j record consisted of: registration number, examination name, mark, credits and examination date; final oral examination record consisted of: registration number, graduation mark, credits and final oral examination date.

Consequently, the same monthly inquiry frequency and observation period was established for each student. The 31st of December 2000 was chosen as initial observation time t_0 because that is when freshmen can first sit for examinations and accumulate own UHC, while the maximum time T_{48} is December 31st, 2004.

A record of 49 inquiries was then built for each graduate, the first of which consists of the registration number (the identifier) while the other 48 include the $uhc_j(t_i)$ values at monthly inquiries t_1, t_2,…, t_{48}.

The $uhc_j(t_i)$ values were obtained from the following formula:

$$uhc_j(t_i) = \sum CE_j(t_i) \times VE_j(t_i) \ ,$$

where the sum operator has to do with examinations passed within the time interval between t_{i-1} e t_i . Hence, if student j passed two examinations in January 2001, his/her $uhc_j(t_1)$ is obtained from the summation of the two examination products (credits $*$ mark); if he/she did not sit for any examination, his/her $uhc_j(t_1)$ is then equal to 0.

A new record based on 49 records and registration numbers was built including $UHC_j(t_i)$ values obtained as ratio between the cumulative of $uhc_j(t_i)$ and its theoretical maximum. The UHC growth curves for the five study programmes and for the whole faculty were calculated with these data.

The parameters of each of the three growth curves examined in Sections 4 and 4.1 were then evaluated with the procedure MIXED of SAS software (Littell *et al.*, 1996). With a focus of fixed effects (level 1), the first two columns on Table 1 show parameters of the functions (made linear in their parameters), on line one; and in brackets on line two, the related test t values. The other columns provide estimated UHC values at time 0 after 12, 24 and 36 months on line one, and on line two, the corresponding mean growth rates.

Table 1. Fixed effects for the three growth models: estimated averages.

Degree' Course	Linear Function parameters p_0	p_1	*UHC* estimated at time 0	*UHC* values (1° row) and Growth rate (2°row): after 12 months	after 24 months	after 36 months	Average *UHC* observed on 36 months
Exponential model							
1	4,968	-0,049	-43,728	20,585	56,120	75,755	69,758
t-value	*123,291*	*-9,682*		*3,926*	*2,169*	*1,199*	*14,800*
2	4,997	-0,056	-47,909	24,739	61,705	80,514	73,103
t-value	*161,576*	*-19,406*		*4,237*	*2,156*	*1,097*	*14,894*
3	5,016	-0,058	-50,830	24,689	62,397	81,224	77,960
t-value	*204,353*	*-17,846*		*4,359*	*2,176*	*1,087*	*12,142*
4	5,290	-0,084	-98,295	27,328	73,367	90,240	87,315
t-value	*66,989*	*-25,894*		*6,079*	*2,228*	*0,816*	*2,650*
5	4,993	-0,052	-47,375	21,246	57,916	77,511	73,434
t-value	*112,229*	*-11,055*		*4,113*	*2,198*	*1,174*	*14,740*
Faculty	**5,007**	**-0,056**	**-49,476**	**24,095**	**61,455**	**80,427**	**74,256**
t-value	***250,427***	***-27,981***		***4,286***	***2,177***	***1,105***	**14,277**
Logistic model							
1	-0,392	-0,149	1,459	8,090	34,359	75,686	69,758
t-value	*-1,239*	*-15,946*		*1,105*	*3,351*	*2,734*	*14,800*
2	-0,747	-0,130	2,066	9,156	32,504	69,705	73,103
t-value	*-3,385*	*-14,248*		*1,084*	*2,859*	*2,752*	*14,894*
3	-0,246	-0,155	1,263	7,635	34,811	77,527	77,960
t-value	*-0,609*	*-12,691*		*1,096*	*3,528*	*2,709*	*12,142*
4	-0,194	-0,162	1,200	7,848	37,397	80,733	87,315
t-value	*-0,172*	*-3,907*		*1,174*	*3,800*	*2,525*	*2,650*
5	0,010	-0,163	0,980	6,552	33,176	77,854	73,434
t-value	*0,030*	*-16,907*		*0,999*	*3,616*	*2,812*	*14,740*
Faculty	**-0,526**	**-0,141**	1,664	**8,435**	**33,397**	**73,185**	**74,256**
t-value	***-3,347***	***-22,928***		***1,091***	***3,141***	***2,771***	**14,277**
Gompertz							
1	1,583	-0,078	0,768	14,939	47,600	74,837	69,758
t-value	*23,331*	*-15,272*		2,226	2,769	1,700	*14,800*
2	1,793	-0,101	0,247	16,794	58,856	85,428	73,103
t-value	*19,012*	*-12,734*		*3,030*	*3,155*	*1,361*	*14,894*
3	1,772	-0,096	0,279	15,757	55,962	83,330	77,960
t-value	*17,262*	*-12,985*		*2,810*	*3,135*	*1,466*	*12,142*
4	2,388	-0,144	0,002	14,576	71,145	94,159	87,315
t-value	*9,424*	*-9,677*		*4,053*	*3,498*	*0,818*	*2,650*
5	1,668	-0,083	0,499	14,189	48,699	76,712	73,434
t-value	*23,072*	*-17,787*		*2,305*	*2,915*	*1,692*	*14,740*
Faculty	**1,775**	**-0,098**	0,273	**16,115**	**56,843**	**83,962**	**74,256**
t-value	***29,299***	***-19,229***		***2,875***	***3,139***	***1,435***	**14,277**

Since variables are absent from the model at level 2, the values above express mean values of UHC level and instant growth rate for single degrees and the Faculty. Test *t* values, which are all highly significant, particularly for the logistic and Gompertz model, allow us to reject the null hypothesis that parameters are equals to 0 in the population.

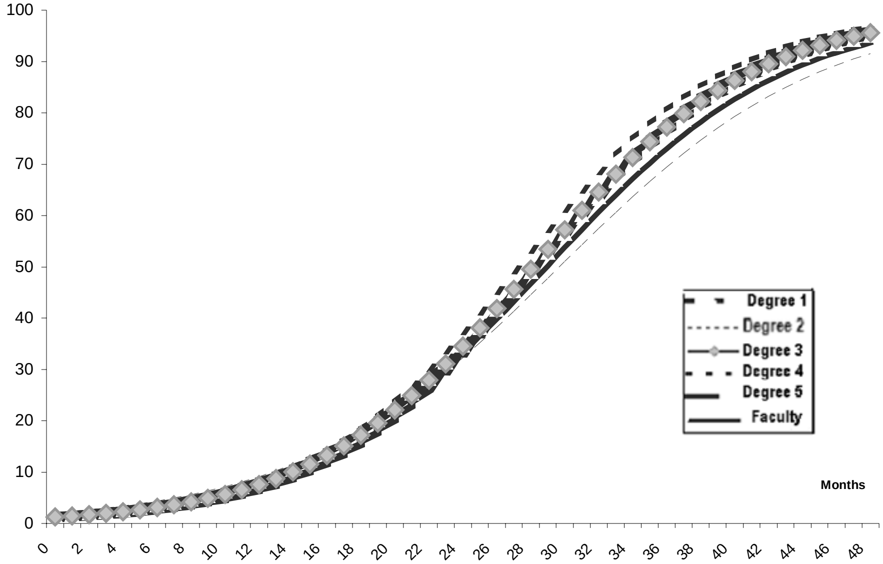


Figure 2. Mean Trajectory for each Degree' Course and Faculty - Logistic model.

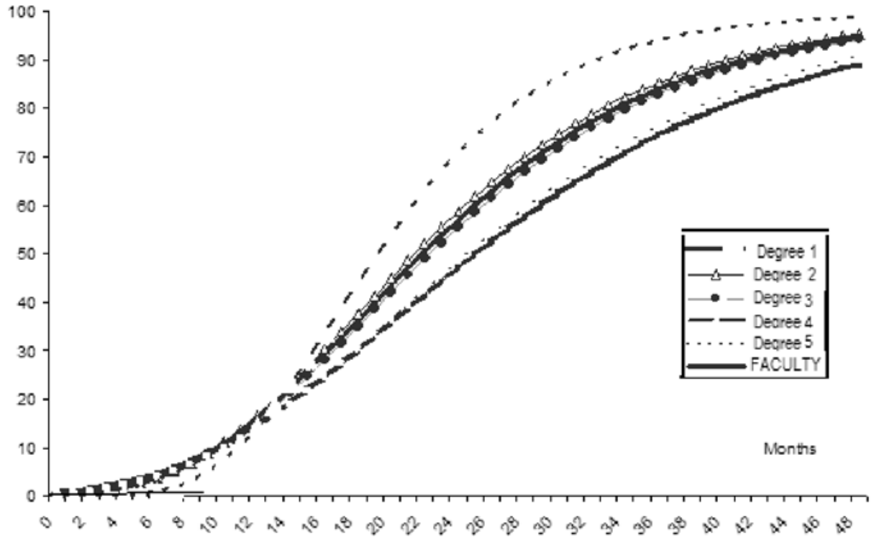


Figure 3. Mean Trajectory for each Degree' Course and Faculty for Gompertz model.

Figures 1 and 2 show the mean growth paths of each degree and the faculty, having assumed respectively a logistic model and a Gompertz model. Clearly, both models show slight differences in growth paths of study programmes. This may be due either to pre-selection, in other words, best student choose certain degrees and not others, or to a difference in the difficulty of each degree. The first hypothesis can be easily tested by introducing three individual characteristics (time independent variables x_{sij}, s =1,2,3) in equations [9], that is, type of secondary school diploma, secondary school final mark, university registration date.

Table 2 shows, instead, the evaluation of casual effects, particularly, of those related to intercept and slope and therefore to UHC mean levels and mean growth rate. The significance of these values shows the existence of intercept

Table 2. Comparison among fixed and random effects estimates

Degree' Course	Fixed Effects		Random Effects		
	Linear parameters		Covariance coefficients		
	β_{00}	β_{10}	τ_{00}	τ_{11}	τ_{01}
Logistic					
1	-0,392	-0,149	0.818	-0.018	0.001
t-value	*-1,239*	*-15,946*	*0.097*	*0.174*	*0.162*
2	-0,747	-0,130	2.298	-0.090	0.004
t-value	*-3,385*	*-14,248*	*0.000*	*0.000*	*0.000*
3	-0,246	-0,155	3.555	-0.099	0.003
t-value	*-0,609*	*-12,691*	*0.002*	*0.003*	*0.003*
4	-0,194	-0,162	3.080	-0.113	0.004
t-value	*-0,172*	*-3,907*	*0.418*	*0.412*	*0.410*
5	0,010	-0,163	1.556	-0.038	0.001
t-value	*0,030*	*-16,907*	*0.020*	*0.037*	*0.038*
Faculty	**-0,526**	**-0,141**	**2.003**	**-0.073**	**0.003**
t-value	*-3,347*	*-22,928*	*0.000*	*0.000*	*0.000*
Gompertz					
1	1,583	-0,078	0.043	-0.001	0.000
t-value	*23,331*	*-15,272*	*0.060*	*0.324*	*0.031*
2	1,793	-0,101	0.451	-0.036	0.003
t-value	*19,012*	*-12,734*	*0.000*	*0.000*	*0.000*
3	1,772	-0,096	0.224	-0.014	0.001
t-value	*17,262*	*-12,985*	*0.003*	*0.005*	*0.001*
4	2,388	-0,144	0.085	-0.007	0.001
t-value	*9,424*	*-9,677*	*no sig.*	*no sig.*	*no sig.*
5	1,668	-0,083	0.079	-0.003	0.000
t-value	*23,072*	*-17,787*	*0.012*	*0.112*	*0.006*
Faculty	**1,583**	**-0,078**	**0.317**	**-0.025**	**0.002**
t-value	*23,331*	*-15,272*	*0.000*	*0.000*	*0.000*

and slope variation that can be that explained with the individual characteristics, that is, with UHC level achieved by each student and through the time spent in achieving them. Covariance between intercept and slope, instead, is significant, although negative, only for two study programmes. These results indicate the existence of negative covariance between time and initial UHC so that longest times to achieve a given UHC level correspond to weaker students.

6. Conclusion

This work achieved the objective to obtain a measure of graduates specific competence growth during higher education based on results of each examination and related time dynamics of a cohort of students registered at five different degree courses of the University of Milan - Bicocca, Faculty of Economics who graduated within the 44 months of observation period. The objective was achieved with statistics methodologies recently presented in the area of variance components models.

The initial hypothesis is that the growth of competence corresponds to a particular type of human capital, namely, university human capital. This evolves in a dynamic context that obviously depends on students' characteristics (the individual component) and is affected by the type of degree course students' registered for (the group component).

On the one hand, the slightly different growth paths of different degree courses seem to confirm what the preliminary statistics analysis, concerned with type of degree and marks distribution, outlined: the existence of some sort of self-selection that brings clever students to choose certain courses rather than others. On the other hand, it seems to point at some disparity in the difficulty of different degrees.

In conclusion, it is this last hypothesis that needs to be further investigated if the effectiveness of education offered by a faculty is to be assessed.

References

EDVISSON L., MALONE M. (1997) *Intellectual Capital: Realizing your Company's True Value by Finding its Hidden Roots*, Harper Collins, New York.

LEV B. (2000) *Knowledge and Shareholder Value*, N.Y. University Press, New York.

LITTELL R., MILIKEN G., STROUP W., WOLFINGER R. (1996). *SAS Systems for Mixed Models*, North Carolina, SAS Institute.

RAO C.R. (1958) Some statistical methods for comparison of growth curves, *Biometrics*, **14**: 1-17.

SINGER J. (1998) Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models, *Journal of Educational and Behavioral Statistics*, **4**: 323-355.