Advances in Intelligent Systems and Computing 200

Editor-in-Chief

For further volumes:
http://www.springer.com/series/11156

Van-Nam Huynh, Vladik Kreinovich,
Songsak Sriboonchitta, and Komsan Suriya (Eds.)

Uncertainty Analysis in Econometrics with Applications

Springer

Editors
Prof. Van-Nam Huynh
Japan Advanced Institute of Science
and Technology
School of Knowledge Science
Ishikawa
Japan

Prof. Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
USA

Prof. Songsak Sriboonchitta
Faculty of Economics
Chiang Mai University
Chinag Mai
Thailand

Prof. Komsan Suriya
Faculty of Economics
Chiang Mai University
Chinag Mai
Thailand

ISSN 2194-5357
e-ISSN 2194-5365
ISBN 978-3-642-35442-7
e-ISBN 978-3-642-35443-4
DOI 10.1007/978-3-642-35443-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012953654

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains papers presented at **TES 2013** – The Sixth International Conference of the Thailand Econometric Society, which is held in Chiang Mai, Thailand, during January 10^{th}–11^{th}, 2013, and hosted by the School of Economics, Chiang Mai University, Thailand.

The aim of this conference is to bring together researchers and scientists working in econometrics and quantitative analysis in economics for an opportunity to present and discuss theoretical and applied research problems as well as to foster research collaborations.

The papers included in this volume were carefully evaluated and recommended for publication by the Scientific Committee. We appreciate the efforts of all the authors who submitted papers and regret that not all of them can be included. The volume begins with two papers from keynote speakers and is followed by the invited and contributed papers.

As a follow-up of TES 2013 conference, several special issues of journals such as *Journal of Econometrics*, *International Journal of Approximate Reasoning*, *International Journal of Intelligent Technology and Applied Statistics* will be organized to publish a small number of extended papers selected from the Conference as well as other relevant contributions received in response to subsequent open calls. These journal submissions will go through a fresh round of reviews in accordance with these journals' guidelines.

The TES 2013 conference is financially supported by the Chiang Mai School of Economics (CMSE), Thailand. We are very thankful to Dean Pisit Leeahtam and CMSE for providing crucial support throughout the organization of TES 2013. We are also especially grateful to Prof. Hung T. Nguyen for his valuable advice and constant support.

We sincerely wish to express our appreciation to all the members of Advisory Board, Administrative Committee, Scientific Committee and Local Organizing Committee for their great help and support. We would also like to thank Prof. Janusz Kacprzyk (Series Editor) and Dr. Thomas Ditzinger (Senior Editor, Engineering/Applied Sciences) for their support and cooperation in this publication.

Last, but not the least, we wish to thank all the authors and participants for their contributions and fruitful discussions that made this conference a success.

We hope that you the reader will find in reading this volume helpful and motivating.

Chiang Mai, Thailand,
January 2013

Van-Nam Huynh
Vladik Kreinovich
Songsak Sriboonchitta
Komsan Suriya

Organization

Organizers

Faculty of Economics, Chiang Mai University
Department of Agricultural Economics, Chiang Mai University
School of Economics & Public Policy, Srinakharinwirot University
Department of Economics, Faculty of Management Science, Khon Kaen University
Development Economics School, National Institue of Development Administration
Faculty of Economics, Kasetsart University

General Chair

Pisit Leeahtam – Faculty of Economics, Chiang Mai University, Thailand

Advisory Board

Pisit Leeahtam (*Chair*)
Pairut Kanjanakaroon
Prapatchon Jariyapan
Kanchana Chokethaworn
Sasipen Phuangsaichai
Kunchon Wattanakul
Manoj Pothaphon

Administrative Committee

Komsan Suriya (*Chair*)
Chaiwat Nimanussornkul
Rossarin Osathanunkul (*Co-Chair*)
Kunsuda Nimanussornku
Nalitra Thaiprasert (*Co-chair*)
Anaspree Chaiwan
Pathairat Paspipatkul
Chukiat Chaiboonsri
Anuphak Saosaovaphak
Supanika Leurcharusmee
Chatchai Khiewnagmdee

Scientific Committee

Hung T. Nguyen (*Chair*), USA & Thailand
Berlin Wu (*Co-chair*), Taiwan
Paul Embrechts (Switzerland)
Christian Gourieroux (France)
Cathy Chen, Taiwan
Hanfeng Chen, USA
Mei-Yuan Chen, Taiwan
Hsiao Cheng, USA
In Choi, Korea
Boris Choy, Australia
Thierry Denoeux, France
Richard Gerlach, Australia
Shih-Hsun Hsu, Taiwan
Wan-Tran Huang, Taiwan
Tsorng-Chyi Hwang, Taiwan
Peter H. Calkins (*Co-chair*), Canada & Thailand
Vladik Kreinovich (*Co-chair*), USA
Van-Nam Huynh, Japan
Chung-Ming Kuan, Taiwan
Kenneth Lin, Taiwan
Antonio Mele, UK
Kimio Morimune, Japan
Marc Paolella, Switzerland
Terry Rockafellar, USA
Daniel Schoch, Malaysia
Surach Tanboon, Thailand
Michael Wolf, Switzerland
Shihti Yu, Taiwan
Qingfeng Liu, Japan
Komsan Suriya, Thailand

Local Organizing Committee

Kulcharee Kritsadathikarnkul (*Organizer-in-chief*)
Tatcha Sudtasan (*Secretary*)
Rattana Charussaengsuriya
Thanikool Boonsriton

Contents

Part II Fundamental Theory

Part I
Keynote Addresses

On the State of the Art of Info-metrics

Amos Golan

Abstract. Info-metrics is the science and practice of inference and quantitatively processing information. In this paper I provide a brief discussion of the state of info-metrics. After defining and discussing the concept of information and types of information I relate these concepts to information processing and data analysis. The connection between info-metrics and the class of information-theoretic methods of inference is discussed here as well. The discussion concludes with a partial list of open questions in info-metrics.

1 Background, Objective and Motivation

Inference and estimation deal with ways to process all of the available information. To do so we first need to build models that connect the observable and unobservable information, and provide the framework for processing and evaluating the information we have. This basic issue of inference and processing of finite information is one of the most fascinating universal problems. Regardless of the quality and amount information we have, the main task is always how to process this information such that the inference - derivation of conclusions from given information or premises - is optimal and is based on minimal assumptions. In this article, I discuss some of the basic problems in information processing and inference (info-metrics) and concentrate on some of the basic solutions provided by the class of information-theoretic methods of inference.

The field of info-metrics is the science and practice of inference and quantitatively processing information. It crosses the boundaries of all sciences and provides a mathematical and philosophical foundation for inference with finite, noisy or incomplete information. Info-metrics is at the intersection of information theory,

Amos Golan
Info-Metric Institute and Department of Economics,
American University, Washington, DC
e-mail: agolan@american.edu

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 3–15.
DOI: 10.1007/978-3-642-35443-4_1

statistical and econometric methods of inference, applied mathematics, complexity theory, decision analysis, basic modeling and the philosophy of science. The study of info-metrics helps in resolving a major challenge for all scientists and all decision makers of how to reason under conditions of incomplete information. Though optimal inference and efficient information processing are at the heart of info-metrics, these issues cannot be developed and studied without understanding information, entropy, probability theory, statistical inference, information and complexity theory as well as the meaning and value of information, data analysis and other related concepts from across the sciences. Info-metrics is based on the notions of information, probabilities and relative entropy. It provides a unified framework for reasoning under conditions of incomplete information.

Though much progress has been made, there are still many deep philosophical, practical and conceptual open problems in info-metrics: What is information and what information is observed? What is a correct inference method? How should a new theory be developed? Is a unified approach to inference, learning and modeling necessary? If so, does info-metrics provide that unified framework? Or even simpler questions related to real data analyses and correct processing of different types and sizes of blurry data fall at the heart of info-metrics. Simply stated, modeling flaws and inference with imperfect information is a major challenge. Inconsistencies between theories and empirical predictions are observed across all scientific disciplines. Info-metrics deals with the study of that challenge. These issues - the fundamental problem, current state, thoughts on future directions and open questions - are the focus of that article.

Generally speaking, to answer the above questions demands a better understanding of both information and information processing. That includes understanding the types of information observed, connecting it to the fundamental - often unobserved - entities of interest and then meshing it all together and processing it in a consistent way that yields the best theories and the best predictions. Info-metrics provides that mathematical and philosophical framework. It generalizes the Maximum Entropy (ME) principle [12][13] which builds on the principle of insufficient reason. The ME principle states that in any inference problem, the probabilities should be assigned by maximizing Shannon's[22] information measure called entropy subject to all available information. Under this principle only the relevant information is used. All information enters as constraints in an optimization process: constraints on the probability distribution representing our state of uncertainty. Maximizing the entropy subject to no constraints yields the uniform distribution representing a state of complete uncertainty. Introducing meaningful information as constraints in the optimization shifts the distribution away from uniformity. The more information there is, the further away the resulting distribution is from uniformity or from a state of complete uncertainty.

The next section discusses, qualitatively and quantitatively, the concept of information and types of information. In Section 3, the classical ME principle (for solving under-determined problems) is formulated and extended. It follows, with a formulation of a whole class of Information-Theoretic (IT) methods of estimation and inference. In Section 4 a special member of the IT-estimators is discussed.

Unlike the other members of the IT Class of estimators, the IT, Generalized Maximum Entropy method can accommodate for all types of information and can be used even for very sparse or small data as well as for ill-behaved data. Some remaining open questions and problems in info-metrics are summarized briefly in Section 5.

2 Information

2.1 Information - A Qualitative Discussion

"Information" is a word packed with seemingly many meanings and somewhat vague in the more ordinary usage. See for example the discussion in [4] or the more practical discussion (as related to information processing) in [6]. One practical way of studying the concept and meaning of information is to study the way scientists across disciplines interpret and understand it. A common view is that when we do not know a certain fact with certainty, whatever reduces the bounds of possibility, or concentrates probability of possible outcomes, informs us. It is an addition to one's stock of knowledge; however measured and of whatever quality. For the applied researcher (who is interested in inference and learning) this means that *information* is anything, such as a fact or an opinion that affects one's estimates or decisions. It is "meaningful content." For example, the *information* in the random variable A about (the random variable) B is the extent to which A changes the uncertainty about B. When A is another random prospect, a particular outcome of it may increase or decrease uncertainty about B. On average, however, if A and B are related, knowledge of outcomes of A should decrease uncertainty about prediction of outcomes of B. More technically, the *information* content of an outcome (of a random variable) is an inverse function of its probability. Consequently, anything "surprising" (an outcome with a low probability) has much information. From a more human or behavioral perspective, like a force that induces a change in motion, *information* is whatever induces an agent (or any living being) to change her/his beliefs where these beliefs are constrained by the new *information*. Similarly, *information* is whatever induces a living entity to change its actions where these actions are constrained by the new *information*.

The above captures only a subset of the more comprehensive notion of "information." For practical purposes of inference and information processing, a main subset of the "input information" - the information used for inference, or the information being processed - may be studied by using a core definition of information and related concepts, such as entropy. That core concept has a precise mathematical definition and is the basic measure that allows us to understand information and information processing. That precise definition (discussed below) is the one provided by Shannon and is based on Hartley's formula [10]. But there is still something missing. The input information is a composite of three types of information: The hard data (quantitative information), the soft information and the prior information. The first and the last are well defined and can be quantitatively analyzed using Shannon, or other, information measures. The second one (soft information) deserves further discussion and represents the toughest part of the inferential information processing. This type of information, most often appearing as "input information" is composed of the assumptions, conjectures and axioms of the observer or modeler. It also captures the observer's beliefs and values as well as intuition. The following example demonstrates the issue and the fundamental problem of processing input information.

Consider studying the state of the economy as is conveyed by the observed information. Given a unique data set composed of the basic macro level indicators of the economy, we wish to infer the state of that economy. Stating it differently, what is the most probable "story" that is consisting with the observed information (hard data)? As we often see, different researchers end up with different outcomes. The problem here is that there is not enough information to ensure one solution. The hard data are not sufficient for recovering the complete story of the economy. There are numerous stories that are fully consistent with the available information and with the data in front of each one of the different researchers. Each researcher is facing a blurry and imperfect data. The only reason for the different conclusions ("output information") must be the researchers' use of soft information. This is the part that is hard to quantify. This is the information that enters every time we process information. This is the information that reflects one's intuition, values, subjective beliefs and other hard-to-define information.

To take into account the above issues demands a better understanding of both information and information processing. That includes understanding the types of information observed, connecting it to the fundamental - often unobserved - entities of interest and then meshing it all together and processing it in a consistent way that yields the best theories and the best predictions. Info-metrics provides that mathematical and philosophical framework. But, as discussed above, info-metrics and the practice of inference with limited and imprecise information cannot be understood without understanding information. Before returning to our discussion of info-metrics, we first quantify the concept of information.

2.2 Information - A Quantitative Discussion

Let $\eta = \{\eta_1, \eta_2, ..., \eta_K\}$ be a finite set and p be a proper probability mass function on η, where "proper probability distribution" means that all the elements are non negative and the sum over all elements equals exactly one. Hartley showed that the amount of information needed to fully characterize all of the K elements of this set is defined by $I(\eta_K) = log_2 K$. Developed within the context of communication theory, it is the logarithm of the number of possibilities (K), or simply a logarithm measure of information. Shannon built on Hartley's formula, within the context of communication process, to develop his information criterion. Shannon's information content of some outcome $h(\eta_i) = h(p_i) \equiv log_2 \frac{1}{p_i}$. If the base 2 is used for the logarithm, the resulting units are "bits". A "bit" is a binary digit - a one or a zero and it is a basic unit of information. All information (data) can be specified in terms of bits. A random variable, for example, with two possible outcomes stores one bit of information. N binary random variables store N bits of information because the total number of possible observed states is n^N and $log_2(2^N)$. The choice of base 2 seems to be a natural choice as it provides the most efficient (cheapest) way of coding and decoding the data, though one can use logarithms of any base.

Shannon's criterion, called *entropy*, reflects the expected informational content of an outcome and is defined as

$$H(\mathbf{p}) \equiv \sum_i p_i log_2 \frac{1}{p_i} = -\sum_i p_i log_2 p_i = E\left[log_2(1/p(X))\right]$$

for some random variable X and with $xlog_2(x)$ tending to zero as x tends to zero. This information criterion, expressed in bits, measures the uncertainty or informational content of X that is implied by p. See the original work of Shannon, or other texts such as [2] or [6] for a detailed discussion of information and entropy and their basic properties of this entropy measure.

Shannon's entropy measure can be extended for defining the informational distance between two proper distributions, say a prior and a posterior. This is called the relative entropy or more commonly known as the Kullback-Leibler divergence measure [17]. The *relative entropy* (or cross entropy) between the two probability mass functions p and q for the random variables X and Y is

$$D(\mathbf{p}||\mathbf{q}) \equiv \sum_k p_k log(p_k/q_k).$$

The relative entropy $D(p||q)$ reflects the gain in information resulting from the additional information in p relative to q. It is an information-theoretic distance of p from q that measures the"inefficiency" of assuming a priori that the distribution is q when the correct distribution is p (see [5]). In a more information theoretic language, if I knew the true distribution of the random variable, I could construct a code with an average description length of $H(\mathbf{p})$ to describe that random variable. But if I use my code for the incorrect distribution q, I need $H(\mathbf{p}) + D(\mathbf{p}||\mathbf{q})$ bits on the average to describe the same random variable. In more econometric terms, using the incorrect likelihood, or model, when analyzing data is costly not only in terms of efficiency and precision but also may lead to an inconsistent estimator. Note that $D(\mathbf{p}||\mathbf{q})$ is not a true distance and is not symmetric. For further discussion on this measure see [2].

Building on Shannon's information and entropy measures, a number of generalized information measures were developed. Though none of these measures exhibit the exact properties of the Shannon's entropy, these measures are used often in econometrics and statistics and provide a basis for defining the class of Information-Theoretic (IT) estimators. These generalized information measures are all indexed by a single parameter: α.

Starting with the idea of describing the gain of information, [21] developed the entropy of order α for incomplete random variables. The relevant generalized entropy measure of a *proper* probability distribution is

$$H_\alpha^R(\mathbf{p}) = \frac{1}{1-\alpha} log \sum_k p_k^\alpha$$

where Shannon measure is just a special case for $\alpha \to 1$. Other generalizations were developed. Among those, the more commonly used in econometrics and statistics are those that were developed during the 1980's by [3] and [24]. The Cressie-Read measure (accommodating for priors) is

$$D_\alpha^{CR}(\mathbf{x}||\mathbf{y}) = D_\alpha^{CR}(\mathbf{p}||\mathbf{q}) = \frac{1}{\alpha(1+\alpha)} \sum_k p_k \left[\left(\frac{p_k}{q_k} \right)^\alpha - 1 \right].$$

For a quantitative comparisons of these measures (including their properties) see, for example, [6]. It is important, however, to emphasize one point on the differences between these measures and Shannon's entropy. With Shannon's entropy, events with high or low probability do not contribute much to the measure's value. With the generalized entropy measures for $\alpha > 1$, higher probability events contribute more to the value than do lower probability events. The average logarithm is replaced by an average of probabilities raised to the power α. Thus, a change in α changes the relative contribution of event k to the total sum.

We now return to the basic problem of solving an under-determined problem where the number of unknown quantities is larger than the number of known quantities, a common problem in statistical and econometric inference, especially when one tries to impose minimal assumptions or structure on the model.

3 The Basic Information Theoretic Model

3.1 The Classical Maximum Entropy Framework

Given T (pure) moments $\mathbf{y}$, our objective is to estimate the K-dimensional, unknown distribution p for the case where $K > T+1$ (an under-determined problem). With under-determined problems we need to introduce more information (constraints) or to choose a certain criterion to pick one of the infinitely many solutions that are consistent with the observed information. In all of the IT family of estimators, the chosen criterion is an informational one.

$$ME = \begin{cases} \hat{\mathbf{p}} = argmax\{H(\mathbf{p}) \equiv -\sum_k p_k log p_k\} \\ s.t.\ \mathbf{y} - X\mathbf{p} = 0; \sum_k p_k = 1 \end{cases}$$

The ME solution is

$$\hat{p}_k = \frac{exp(-\sum_{t=1}^T \hat{\lambda}_t x_{tk})}{\sum_k exp(-\sum_{t=1}^T \hat{\lambda}_t x_{tk})} \equiv \frac{exp(-\sum_{t=1}^T \hat{\lambda}_t x_{tk})}{\Omega(\hat{\boldsymbol{\lambda}})}$$

where is the T-dimensional vector of estimated Lagrange multipliers.

The ME is formulated in terms of a constrained optimization where the optimization is carried out with respect to the p's. The concentrated model is

$$\begin{aligned}\ell(\lambda) &= -\sum_k p_k log(p_k) + \sum_t \lambda_t \left[y_t - \sum_k x_{tk} p_k \right] \\ &= -\sum_k p_k(\lambda) \left[-\sum_t \lambda_t x_{tk} - log(\Omega(\boldsymbol{\lambda})) \right] + \sum_t \lambda_t \left[y_t - \sum_k x_{tk} p_k \right] \\ &= \sum_t \lambda_t y_t + log(\Omega(\boldsymbol{\lambda}))\end{aligned} \tag{1}$$

Looking at the concentrated model it is clear that it has the same form of a likelihood function. In fact, it is equivalent to the maximum likelihood Logit for a discrete choice model.

The advantages of the concentrated model is that first, an unconstrained model is simpler (and computationally superior), second by moving from the probability space to the Lagrange multipliers' space the dimension of the model decreases significantly, and last, it allows a direct comparison with the more traditional likelihood methods.

Out of the infinitely many solutions that are consistent with the observed information (moments), the chosen ME distribution is the most uniform one. It is the most conservative estimates of p. For further motivations for using the ME, related test statistics and examples, see [6] and others.

So far the model handled only "hard" observed information - the first type of information defined earlier. We now extend the ME to include prior information (the third type of information defined). Substituting the Kullback-Leibler measure for Shannon's entropy, the Cross Entropy (CE) formalism (including priors) is

$$CE = \begin{cases} \tilde{\mathbf{p}} = argmin\,\{D(\mathbf{p}||\mathbf{q}) \equiv \sum_k p_k log(p_k/q_k)\} \\ s.t.\ \mathbf{y} - X\mathbf{p} = 0; \sum_k p_k = 1 \end{cases}$$

The estimated probabilities are:

$$\tilde{p}_k = \frac{q_k exp(\sum_{t=1}^T \tilde{\lambda}_t x_{tk})}{\sum_k exp(\sum_{t=1}^T \tilde{\lambda}_t x_{tk})} \equiv \frac{q_k exp(\sum_{t=1}^T \tilde{\lambda}_t x_{tk})}{\Omega(\tilde{\boldsymbol{\lambda}})}$$

and the concentrated model is

$$\ell(\boldsymbol{\lambda}) = \sum_k p_k log(p_k/q_k) + \sum_t \lambda_t \left[y_t - \sum_k x_{tk} p_k \right] = \sum_t \lambda_t y_t - log(\Omega(\boldsymbol{\lambda})).$$

Unlike the ME model where the probabilities are pushed toward uniformity, in this case the resulting distribution is the one that is closest to the priors and satisfies the observed moments. If no information is used, the estimates are just the priors. As more information is introduced (via the moment constraints), the estimates start to shift away from the priors (assuming the new information tells a different story than the priors). If the priors are uniform the CE and ME estimates are the same.

So far we concentrated on processing two types of information: hard data (quantities) and priors. We still need to accommodate these models for processing the other soft information (type 2 defined above). We do it in the next section. But first we build on the ME and CE framework to construct a generic IT estimator.

3.2 The Generic IT Framework

Within the same objective of solving an under-determined problem, or estimating with minimal assumptions, it is possible to generalize the ME and CE formalisms. The idea is to substitute the generalized entropy function for Shannon's entropy as the objective criterion within the ME/CE formulations. To define the problem in more general notations, let $Y = \{Y_1, Y_2, ..., Y_T\}$ be a sample of i.i.d. observations from an unknown distribution F_0. There is an N-dimensional parameter vector $\boldsymbol{\theta}_0$ that is related to F_0 in the sense that the information about $\boldsymbol{\theta}_0$ and F_0 is available in the form of some M moments (or functionally independent unbiased estimating functions). A possible IT likelihood of this parameter vector could be defined by considering the distributions supported on the sample, where each Y_i is assigned a probability P_i. For a specified value of the parameter vector, say $\boldsymbol{\theta}_S$ the IT (empirical) likelihood is defined as the maximal value of some function $f(.)$ over all such probability distributions satisfying the relationship among y, p, and $\boldsymbol{\theta}_S$ that are specified by the M-dimensional equations $\mathbf{g}(\mathbf{y}, \mathbf{p}, \boldsymbol{\theta}_S) = [\mathbf{0}]$. Under that approach, one starts by defining the feasible set of proper probability distributions supported on the sample observations. This feasible set is characterized by a set of M restrictions on the unknown probabilities p. Given the T observations and the M moment conditions, the objective is to estimate the probability distribution p. These estimates are the unobserved empirical weights of the T data points. Like the ME formulation, a simple way to solve such an under-determined problem is to transform it into a well-posed, constrained optimization problem. Using one of the generalized divergence measures forms a basis for optimal decision making and for statistical inference. To simplify presentation, for the rest of this paper we will only use the Cressie-Read measure, though the results hold for all IT measures. The IT estimator is

$$\text{Generic IT}=\begin{cases} \mathbf{p}^{*}=argmin\left\{ f(\mathbf{p}||\mathbf{q})=D_{\alpha}^{CR}(\mathbf{p}||\mathbf{q}) =\dfrac{1}{\alpha(1+\alpha)}\Sigma_k p_k\left[\left(\frac{p_k}{q_k}\right)^{\alpha}-1\right]\right\} \\ \text{s.t. } g_m(y,p,\theta_1)=[\mathbf{0}]; m=1,2,...,M \\ \Sigma_i p_i = 1; i=1,2,...,T \text{ and } M<T-1 \end{cases}$$

The class of IT estimators $\mathbf{p}^*$ depends on the pre specified α and on the exact specification of $g_m(.)$. If, for example, $\alpha \to 1$, $f(p||q) = D_{\alpha}^{R}(\mathbf{p}||\mathbf{q})$ is the Kullback-Liebler divergence measure $D(\mathbf{p}||\mathbf{q}) \equiv \sum_{t=1}^{T} p_i log(p_i/q_i)$, with the CE solution $\tilde{p}$. If in addition, the q's are uniform, then $\tilde{p} = \hat{p}$ which is equivalent to using the negative of the Shannon's entropy as the objective function, resulting in the ME solution $\hat{p}$.

Once the optimal solution $\mathbf{p}^*$ is found, the concentrated model can be formulated. The estimated Lagrange multipliers, which are directly related to the estimated parameters, also reflect the contribution of each constraint (moment) to the optimal value of the objective function. In the IT class, the objective is an informational criterion, meaning the estimated Lagrange multipliers reflect the marginal information of each constraint. It is the same Lagrange multipliers that enter as the parameters in the estimated probability distribution.

Depending on α, there are many members of this IT class of models. The case of $\alpha \to 1$ was already discussed. Another case is now briefly discussed. Letting $\alpha \to 0$, and assuming the priors are uniform, subject to the same set of $M+1$ restrictions yields the Empirical Likelihood (EL) method. The EL criterion is simply the probability of the observed sample or its natural logarithm. Following [18][19] and [20], the EL method is

$$EL = \begin{cases} \mathbf{p}^* = argmin \left\{ f(\mathbf{p}) = \frac{1}{T} \Sigma_{t=1}^{T} log(p_t) \right\} \\ s.t. \ \Sigma_t p_t g_m(y_t;\theta) = 0; \\ \Sigma_i p_i = 1; p_t \geq 0 \end{cases}$$

The resulting estimated weights (probabilities) are

$$\mathbf{p}^* = T^{-1} \left[\sum_m \lambda_m^* g_m(y_t;\theta) + 1 \right]^{-1}$$

Extending this framework for linear regression is trivial but is not discussed here. For more details, including examples and test-statistics, see [19] and [20]. For recent advancements and a survey of the EL (and Generalized EL) see [23] and [15]. For a further discussion of all members of the IT family of estimators see [6] and [14]. For Bayesian and IT see the seminal work of [25].

Two notes in conclusion. First, substituting the objective function within the EL framework for the entropy of order α takes us back to the generic IT estimator discussed above. This idea goes back to the work of Imbens et.al, [11] that discuss three special cases of that objective function, and to the work of [16] that connected the ME and the GMM in a very elegant way. Second, [23] considered a more general class of estimators which he called Generalized EL (GEL). Under the GEL, the constrained optimization model is transformed into a concentrated model. This idea is similar to the original work of [1] who were the first to construct the concentrated ME model, and then applied and extended to the Generalized ME by [8], and [9]. See [9] for detailed examples and derivations.

4 The Generalized Information Theoretic Model

We now consider the more realistic case where all of the observed information may be noisy or blurry, the information may be contradictory or complementary, and all the soft information can be incorporated directly into the processing framework.

The objective here is to establish a single framework that accommodates for possible noise in the observed information while allowing specification of all types of information (hard, priors and soft). A proper way to resolve it (without imposing additional assumptions/information) is via the information theoretic Generalized Cross Entropy (GCE) or Generalized Maximum Entropy (GME) framework of [9]. In this case, all the constraints are specified as stochastic constraints together with the proper probabilities requirements. Maximizing the joint entropies of the noise and the signal subject to these constraints yields a unique solution. To establish the notations and basic framework, consider extending the CE (or ME) framework for stochastic moments. Rather than specifying the observed moments as $y = Xp$, we specify it as $y = Xp + \varepsilon$ where ε is a zero mean noise vector (independent of X). In this case, regardless of the number of observed pieces of information (moments), the problem is always under-determined. But unlike the previous (pure) moments case, we cannot apply the CE directly. But we can do so after reformulating the noise (ε) as a set of proper probability distributions. To do so, we formulate each unobserved random error (ε_i) as an expected value over a J-dimensional, symmetric about zero, support space v. Then, the IT-GCE (with stochastic moments) is:

$$
\begin{aligned}
&Min_{\{p,W\}} D(\mathbf{P},\mathbf{W}||\mathbf{P}^0,\mathbf{W}^0) = \textstyle\sum_k p_k log(p_k/p_k^0) + \sum_{tj} w_{tj} log(w_{tj}/w_{tj}^0) \\
&s.t.\ y_t = \textstyle\sum_k p_k x_{tk} + \varepsilon_t = \sum_k p_k x_{tk} + \sum_j w_{tj} v_j \\
&\textstyle\sum_k p_k = 1, \text{ and } \sum_j w_{tj} = 1
\end{aligned}
$$

As was shown previously, the concentrated model with respect to the Lagrange multipliers λ can then be formulated as well.

Though this is a very trivial example, it does introduce the framework for processing all types of information: noisy and pure, hard and soft, contradictory and complimentary, linear and nonlinear. It provides a natural framework for empirically validating and testing theories, conjectures and assumptions with the observable information. It is one of the only approaches that allow us to incorporate soft and hard data together and to empirically test these different types of information for consistency. It is essential to emphasize that even though all of the information is entered as potentially noisy, this framework does not force the noise to be none zero, rather it allows it to be none zero. Further, and possibly most important, for most problems, that framework generalizes the classical ME/CE but without added complexity: the same number of basic parameters (the $\boldsymbol{\lambda}$'s that determine both $\mathbf{p}$ and $\boldsymbol{\varepsilon}$) capturing the story hidden in the information is not increased.

A basic property of the classical ME (or CE) approach, as well as most members of the IT family of estimators is that the moment conditions have to be exactly fulfilled (zero moment conditions). This property is satisfactory for (relatively) large samples or for well behaved samples. Unfortunately, in both the social and natural sciences we are often trying to understand small or ill-behaved (and often non-experimental) data where the zero-moments' restrictions may be too costly. Another basic concern is how to incorporate in the estimation procedure information resulting from economic-theoretic behavior such as agents' optimization. The IT-GME accommodates for these concerns for all regression models by introducing all information as noisy constraints. These stochastic moments can be introduced in two ways. The first is by allowing for some additive noise (with mean zero) for each one of the moment conditions. This was discussed above. The second is by viewing each observation as a noisy moment resulting from the same data generating process. Thus, each observed data point can be treated as a composite of two components: signal and noise. The IT-GME model that was developed in the early 1990's has the above view in mind and treats the moments (or each observation) as stochastic. For detailed background, development and applications of the IT-GME see [9] and [6]. The IT-GME for the traditional linear regression model, as well as for other statistical problems, can be formulated using the above framework. The exact details are beyond the scope of this short discussion on the basics of info-metrics, but we note that this framework allows us to incorporate directly all types of information. See above references as well as [7].

5 Open Questions

Keeping in mind that all types of information are finite and that the observed information is often very limited and blurry, all inferential problems are inherently underdetermined. That problem brings out some of the fundamental questions a researcher has to deal with when constructing a theory and processing information. For example, how can one connect the unobserved preferences with the observed actions? Or what inference method is the "correct" method to use? Or how can the theory be validated with the observed information? How should one handle the noise if the exact underlying distribution is unknown? How can one connect the observed noisy data with the basic entities of interest? All of these questions arise naturally when one deals with small and noisy data within the social sciences. Similar inferential problems exist also with big data. One trivial example is image reconstruction or a balancing of a very large matrix. The first problem is how to reduce the data (or the dimensionality of the problem) so the reconstruction will be computationally efficient. Within the IT framework one can solve that problem as well. For more derivations and examples (small and large data) see [6]. A partial list of the fundamental open questions in info-metrics is provided below. Though these questions are not necessarily independent of one another, it is helpful to include each one of these questions separately.

The questions are:

1. What is information?

2. What information do we observe?

3. How can we connect the observed information to the basic unit (or entity) of interest (usually unobserved)?

4. How should we process the information we observe while connecting it to the basic unit of interest?
5. Can we quantify all types of information?
6. How can we handle contradicting/complimentary evidence (or information)?
7. What is "useful" information?
8. Is there a way to assign relative/absolute value to information?
9. How is the macro level information connected to the basic micro level information?
10. How should we do inference and modeling with finite information?
11. How can we validate our theories and models?
12. What is a correct inference method? Is it universal to all inferential problems? (What are the mathematical foundations of that method?)
13. Under what conditions members of the IT family of estimators are equivalent to the more traditional estimation methods?

The above list is not complete but it provides a window toward some pressing issues within info-metrics that need more research. A more detailed discussion and potential answers to some of these questions is part of a current and future research agenda.

References

1. Agmon, N., Alhassid, Y., Levine, R.D.: An Algorithm for Finding the Distribution of Maximal Entropy. Journal of Computational Physics 30, 250–259 (1979)
2. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. John Wiley & Sons, New York (2006)
3. Cressie, N., Read, T.R.C.: Multinomial goodness-of-fit tests. J. Royal Stat. Soc. B 46, 440–464 (1984)
4. Floridi, L.: The Philosophy of Information. Oxford University Press, Oxford (2011)
5. Gokhale, D.V., Kullback, S.: The Information in Contingency Tables. Marcel Dekker, New York (1978)
6. Golan, A.: Information and Entropy Econometrics - A Review and Synthesis. Foundations and Trends in Econometric 2(1-2), 1–145 (2008)
7. Golan, A., Gzyl, H.: An Entropic Estimator for Linear Inverse Problems. Entropy 14(5), 892–923 (2012)
8. Golan, A., Judge, G.:: Recovering and Processing Information in the Case of Underdetermined Economic Inverse Models, UC Berkeley (1992)

9. Golan, A., Judge, G., Miller, D.: Maximum Entropy Econometrics: Robust Estimation with Limited Data. John Wiley & Sons, New York (1996)
10. Hartley, R.V.L.: Transmission of Information. Bell System Technical Journal, 535–563 (July 1928)
11. Imbens, G.W., Johnson, P., Spady, R.H.: Information-Theoretic Approaches to Inference in Moment Condition Models. Econometrica 66, 333–357 (1998)
12. Jaynes, E.T.: Information Theory and Statistical Mechanics. Physics Review 106, 620–630 (1957a)
13. Jaynes, E.T.: Information Theory and Statistical Mechanics II. Physics Review 108, 171–190 (1957b)
14. Judge, G., Mittelhammer, R.: An Information Theoretic Approach to Econometrics. Cambridge University Press (2011)
15. Kitamura, Y.: Empirical Likelihood Methods in econometrics: Theory and Practice, Cowles Foundation Discussion Paper No. 1569 (2006)
16. Kitamura, Y., Stutzer, M.: An Information Theoretic Alternative to Generalized Method of Moments Estimation. Econometrica 65(4), 861–874 (1997)
17. Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Math. Stat. 22, 79–86 (1951)
18. Owen, A.: Empirical Likelihood for Linear Models. The Annals of Statistics 19(4), 1725–1747 (1991)
19. Owen, A.: Empirical Likelihood. Chapman & Hall/CRC (2001)
20. Qin, J., Lawless, J.: Empirical Likelihood and General Estimating Equations. The Annals of Statistics 22, 300–325 (1994)
21. Renyi, A.: On measures of information and entropy. In: Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960, vol. I, 547 (1961)
22. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 379–423 (1948)
23. Smith, R.J.: Empirical Likelihood Estimation and Inference. In: Salmon, M., Marriott, P. (eds.) Applications of Differential Geometry to Econometrics, pp. 119–150. Cambridge University Press (2000)
24. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys. 52, 479–487 (1988)
25. Zellner, A.: The Bayesian Method of Moments (BMOM): Theory and Applications. In: Fomby, T., Hill, R.C. (eds.) Advances in Econometrics (1997)

A Test for Strict Stationarity

Luiz Renato Lima and Breno Neri

Abstract. We introduce a test for strict stationarity based on the fluctuations of the quantiles of the data, and we show that this test has power against the alternative hypothesis of unconditional heteroskedasticity while other tests for first order (level) stationarity as the KPSS test (Kwiatkowski et al., 1992) and, its robust version, the IKPSS test (de Jong et al., 2007) have low power against this alternative of time-varying variance. Moreover, our test has power against the alternative hypothesis of time-varying kurtosis, while the test for second order (covariance) stationarity introduced by Xiao and Lima (2007) has power close to size against this alternative.

Keywords: strict stationarity testing, time-varying volatility, time-verying kurtosis.

JEL Classification: C12, C22.

Disclaimer: The opinions of the authors do not necessarily represent the opinions of Analysis Group.

1 Introduction

Several techniques employed in time-series econometrics rely on stationarity. So, the development of tests for stationarity is an active field of research.

In 1992, Kwiatkowski, Phillips, Schmidt and Shim (KPSS) proposed a test for for first order (level) stationarity[1] based on the following standardized empirical process:

Luiz Renato Lima
Department of Economics, The University of Tennessee,
514 Stokely Management Center, Knoxville, TN 37996-0550,
and Federal University of Paraiba
e-mail: llima@utk.edu

Breno Neri
Analysis Group, 10 Rockefeller Center, New York, NY 10020
e-mail: bneri@analysisgroup.com

[1] A stochastic process $\{y_t\}_{t=-\infty}^{\infty}$ is said to be first order (or level) stationary if the unconditional mean is finite and constant over time, $\mathrm{e}y_t = \mu < \infty, \forall t$.

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 17–30.
DOI: 10.1007/978-3-642-35443-4_2

$$S_T(r) := \frac{1}{\widehat{\omega}\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} (y_t - \overline{y}_T),$$

where $r \in [0,1]$, $\overline{y}_T$ is the sample mean of $\{y_t\}_{t=1}^{T}$ and $\widehat{\omega}^2$ is a nonparametric consistent estimator of the long-run variance

$$\omega^2 = \lim_{T\to\infty} \mathrm{e}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_t - \overline{y}_T)\right)^2.$$

In order to measure the fluctuation of $S_T(r)$, they consider the Cramér-von Mises metric,

$$h(S_T(r)) := \frac{1}{T} \sum_{k=1}^{T} S_T\left(\frac{k}{T}\right)^2,$$

an alternative to the Kolmogorov-Smirnov metric. The KPSS test statistic is then given by

$$KPSS = \frac{1}{(\widehat{\omega}T)^2} \sum_{k=1}^{T} \left(\sum_{t=1}^{k} (y_t - \overline{y}_T)\right)^2,$$

and, under the null hypothesis of level stationarity,

$$KPSS \xrightarrow{d} \int_0^1 \kappa(\alpha)^2 d\alpha,$$

where $\xrightarrow{d}$ indicates convergence in distribution and $\kappa(\alpha)$ is the standard Brownian bridge. That is, $\kappa(\alpha) := W(\alpha) - \alpha W(1)$, where $W(\alpha)$ is the Wiener process. The critical values can be found in KPSS (1992).

In a recent paper, de Jong et al. (2007) proposed a robust version of the KPSS test based on the following empirical process:

$$I_T(r) := \frac{1}{\widehat{\sigma}\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \mathrm{sign}\, y_t - m_T,$$

where m_T is the sample median of $\{y_t\}_{t=1}^{T}$, $\widehat{\sigma}^2$ is a nonparametric consistent estimator of the long-run variance,

$$\sigma^2 = \lim_{T\to\infty} \mathrm{e}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathrm{sign}\, y_t - m_T\right)^2$$

and sign$x$ is the sign of x: 1 if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$.

And they applied the Cramér-von Mises metric to measure the fluctuation of the empirical process $I_T(r)$. This gives rise to the IKPSS test statistic

$$IKPSS = \frac{1}{(\widehat{\sigma}T)^2} \sum_{k=1}^{T} \left(\sum_{t=1}^{k} \operatorname{sign} y_t - m_T \right)^2 .$$

Under the null hypothesis of level stationarity, de Jong et al. (2007) show that

$$IKPSS \overset{d}{\Longrightarrow} \int_0^1 \kappa(\alpha)^2 d\alpha,$$

the same limiting distribution as the KPSS test statistic. Unlike the KPSS test, the IKPSS has correct size[2] under the presence of fat-tailed errors. When the alternative hypothesis is unit root,[3] the indicator test has lower power than the KPSS when tails are thin, but higher power when tails are fat.

However, when the aforementioned traditional stationarity tests are applied to test stationarity, it is difficult to detect alternatives with unconditional volatility (distribution scale) that changes over time.

In the same year, 2007, Xiao and Lima proposed a test for second order (covariance) stationarity[4] based on the following standardized bivariate empirical process:

$$Z_T(r) := \frac{1}{\sqrt{T}} \widehat{\Omega}^{-\frac{1}{2}} \sum_{t=1}^{\lfloor Tr \rfloor} \begin{pmatrix} \widetilde{y}_t \\ v_t \end{pmatrix},$$

where

$$\widetilde{y}_t := y_t - \frac{1}{T} \sum_{j-1}^{T} y_j$$

is the demeaned data, $v_t := \widetilde{y}_t^2 - \sigma_y^2$,

$$\sigma_y^2 := \frac{1}{T} \sum_{t=1}^{T} \widetilde{y}_t^2,$$

and $\widehat{\Omega}^{-\frac{1}{2}}$ is the inverse of the Choleski decomposition[5] of $\widehat{\Omega}$, a nonparametric consistent estimator of the long-run covariance matrix

---

[2] In hypothesis testing, given a confidence level, the size of a test is the probability of rejecting the null hypothesis when the null hypothesis is indeed true. Analogously, the power of a test is the probability of rejecting the null hypothesis when the null hypothesis is indeed false.

[3] A stochastic process is said to be a unit root process when at least one of the roots of its characteristic equation lies on the unit circle. A unit root process is said to be integrated of order n, $I(n)$, when it has n unit roots. An integrated process is non-stationary. For more on unit root process or stationarity, see Hamilton, 1994.

[4] A stochastic process $\{y_t\}_{t=-\infty}^{\infty}$ is said to be second order (covariance, weak or wide-sense) stationary if it is first order stationary and e$y_t y_{t-j} = \gamma_j < \infty, \forall t$.

[5] Unless in a degenerate case, the covariance matrix $\widehat{\Omega}$ is positive definite, and hence hence it is invertible, as it is its Choleski decomposition.

$$\Omega = \lim_{T\to\infty} \mathrm{e}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\widetilde{y}_t\\ v_t\end{pmatrix}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\widetilde{y}_t\\ v_t\end{pmatrix}\right)'.$$

Then, they applied the Kolmogorov-Smirnov metric to measure the fluctuation of the empirical process $Z_T(r)$. Their test statistic is then

$$XL = \max_{1\leq k\leq T}\left\|\frac{1}{\sqrt{T}}\widehat{\Omega}^{-\frac{1}{2}}\sum_{t=1}^{k}\begin{pmatrix}\widetilde{y}_t\\ v_t\end{pmatrix}\right\|_1.$$

Under the null hypothesis of covariance stationarity,

$$XL \overset{d}{\Longrightarrow} \sup_{0\leq r\leq 1}\left\|\begin{pmatrix}W_1(r)-rW_1(1)\\ W_2(r)-rW_2(1)\end{pmatrix}\right\|_1,$$

where $\left(W_1(r)-rW_1(1)\ W_2(r)-rW_2(1)\right)'$ is the 2-dimensional standardized Brownian bridge. The critical values can be found in Xiao and Lima (2007).

Unlike the KPSS or the IKPSS, the XL test has power not only against the alternative hypothesis of distribution location varying on time but also against the alternative hypothesis of distribution scale (unconditional volatility) varying on time. However, all of the aforementioned tests have power close to size against the alternative hypothesis of time-varying kurtosis.

As Busetti and Harvey (2007) discuss, the distribution of a random variable may presents changes over time that does not impact the level or the variance. For instance, maybe the asymmetry or fatness of the tail is time-varying. This is particularly important in analyzing financial time-series. To exemplify this point, consider how changes in lower tail quantiles may impact decisions of a risk manager or a regulatory agency.[6]

In this paper, we propose a new test for the null hypothesis of strict stationarity[7] as a useful complement to the previous procedures. This new test uses the sign of the data minus the sample quantiles. In this way, this new test can be seen as a generalization of the IKPSS test, since the latter uses the sign of the data minus the sample median only. Comparing to the KPSS, IKPSS and XL tests, the proposed test has power not only against unit root alternative, alternatives with structural changes in the mean and alternatives with unconditional heteroskedasticity, but also has good power in detecting changes in higher moments of the unconditional distribution.

This paper is organized as follows: Section 2 describes our testing procedure; Section 3 brings the Monte Carlo; an empirical exercise is done in Section 4; and Section 5 concludes.

---

[6] See for instance Value-at-Risk (VaR), a measure of risk based on a lower tail quantile, that is of considerable importance in financial regulation (Lima and Neri, 2007).

[7] A stochastic process $\{y_t\}_{t=-\infty}^{\infty}$ is said to be strict (or strict sense) stationary if the unconditional distribution of y_t is constant over time.

2 A Test for Strict Stationarity

Let $\{y_t\}_{t=1}^T$ be the data and define $b(\tau) := \arg\min_{b\in\mathbb{R}} \sum_{t=1}^T \rho_\tau(y_t - b)$ for $\tau \in [0,1]$, where $\rho_\tau(u) = (\tau - 1_{u<0})u$. That is, $\rho_\tau(u)$ is equal to $(\tau-1)u$ if $u<0$ and τu otherwise. Therefore, $b(\tau)$ is simply the τ^{th} sample unconditional quantile of $\{y_t\}_{t=1}^T$.

Notice that ρ_τ is not everywhere differentiable but, since it is convex, we can still compute the subgradient. The subgradient plays the same role in quantile estimation as the score function in maximum likelihood estimation. The subgradient of ρ_τ is given by[8]

$$\psi_\tau(u) = \tau - 1_{u<0}.$$

We now define the empirical process

$$S_T(r,\tau) := \frac{1}{\widehat{\pi}(\tau)\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \psi_\tau\left(y_t - b(\tau)\right),$$

where $r \in [0,1]$, and $\widehat{\pi}(\tau)^2$ is a nonparametric consistent estimator of the long-run variance

$$\pi(\tau)^2 := \lim_{T\to\infty} \mathrm{e}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_\tau\left(y_t - b_0(\tau)\right)\right)^2,$$

where $b_0(\tau)$ is the population τ^{th} unconditional quantile of the $\{y_t\}_{t=1}^T$. $\pi(\tau)^2$ can be computed as the HAC estimator[9]

$$\widehat{\pi}(\tau)^2 := \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T K\left(\frac{i-j}{q_T}\right) \psi_\tau\left(y_i - b(\tau)\right) \psi_\tau\left(y_j - b(\tau)\right),$$

where K is a kernel function that satisfies, as in de Jong et al. (2007),

1. $\int_{-\infty}^{\infty} |\omega(\xi)|\, d\xi < \infty$, where $\omega(\xi) := \frac{1}{2\pi}\int_{-\infty}^{\infty} K(x)e^{-ix\xi}dx$;
2. K is continuous at all but a finite number of points, $K(x) = K(-x)$, $|K(x)| \le l(x)$, where $l(x)$ is non-increasing and $\int_0^\infty |l(x)|\, dx < \infty$, and $K(0) = 1$;
3. $\lim_{T\uparrow\infty} q_T = \infty$ and $\lim_{T\uparrow\infty} \frac{q_T}{T} = 0$.

This paper proposes to test for strict stationarity by using the Kolmogorov-Smirnov metric to measure the fluctuation of $S_T(r,\tau)$ across various quantiles $\tau \in \Gamma_w = [w, 1-w]$, for some $w \in \left(0, \frac{1}{2}\right)$, which gives rise to the following test statistic:

$$SS = \max_{\tau\in\Gamma_w} \max_{1\le k\le T} \frac{1}{\widehat{\pi}(\tau)\sqrt{T}} \left| \sum_{t=1}^k \psi_\tau\left(y_t - b(\tau)\right) - \frac{k}{T}\sum_{t=1}^T \psi_\tau\left(y_t - b(\tau)\right) \right|.$$

[8] In fact, the subgradient of ρ_τ at zero is not unique; it can be any element of the closed interval $[\tau-1, \tau]$.

[9] See Newey and West (1987) for more on Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimation.

Assumption 1. *(Null Hypothesis H_0)*

1. $\{y_t\}_{t=1}^{\infty}$ *is a strictly stationary stochastic process and* $b_0(\tau)$ *is the unique population* τ^{th} *unconditional quantile of* y_t*;*
2. $\{y_t\}_{t=1}^{\infty}$ *is strong* $(\alpha-)$ *mixing*[10] *and, for some finite* $\kappa > 2$*,* $C > 0$ *and* $\eta > 0$*,* $\alpha(m) \leq Cm^{-\frac{\kappa}{\kappa-2}-\eta}$*;*
3. $y_t - b_0(\tau)$ *have a continuous density* f *in a neighborhood* $[-\eta,\eta]$ *of 0 for some* $\eta > 0$*, and* $\inf_{y\in[-\eta,\eta]} f(y) > 0$*;*
4. $\sigma^2 \in (0,\infty)$.

Theorem 1. *Under Assumption 1,* $SS \xrightarrow{d} \sup_{\tau\in\Gamma_w} \sup_{0\leq r\leq 1} |B(r,\tau)|$*, where* $B(r,\tau)$ *is the Brownian Pillow (or the tucked Brownian Sheet).*

A proof for Theorem 1 for the case in which the innovations are i.i.d. was done by Qu (2005). The critical values of our test are computed through the simulation of 10^5 time series with 1000 observations $e_t \sim i.i.d. U[0,1]$.[11] $B(r,\tau)$ is then approximated by

$$\frac{1}{\widehat{\gamma}(\tau)\sqrt{T}}\left(\sum_{t=1}^{k} 1_{e_t\leq\tau} - \frac{k}{T}\sum_{t=1}^{T} 1_{e_t\leq\tau}\right),$$

where $k = \lfloor Tr \rfloor$, and $\widehat{\gamma}(\tau)^2$ is the sample variance (over k) of $\sum_{t=1}^{k} 1_{e_t\leq\tau} - \frac{k}{T}\sum_{t=1}^{T} 1_{e_t\leq\tau}$. The supremum of the absolute value of this approximating process is obtained by maximizing over k and τ. We considered $\tau \in [0.10, 0.90]$ with increments of 0.01. The critical values for the significance levels of 10%, 5% and 1% are 1.65, 1.77 and 2.01, respectively.

3 Monte Carlo Experiment

In this section we report the results of our Monte Carlo experiment that investigate the size and power of the KPSS, IKPSS, XL and our test for strict stationarity (SS). Our experiment is coded in R and it is run in one of the Linux HPCCs (High Performance Computation Clusters) at New York University (NYU). We follow de Jong et al. (2007) and vary tail thickness by considering t distributions with different degrees of freedom. In particular, we consider t_∞ (normal), t_5, t_3, t_2, and t_1 (Cauchy). We consider sample sizes $T = 100$, $T = 500$ and $T = 1000$. The significance level of the tests is 5%. For the SS test, we set $\tau \in [0.10, 0.90]$ with increments of 0.01. Our results are based on $N = 10^5$ replications.

---

[10] For the definition of strong mixing stochastic process, also knows as $\alpha-$mixing stochastic process, see White, 2001, pp. 46-8.

[11] All numerical procedures used in this paper are implemented in R, and can be downloaded from `https://files.nyu.edu/bpn207/public/`. R is a free computer programming language very suited to statisticians and econometricians, and can be downloaded from `http://www.r-project.org`.

We report our experiment in which the errors ε_t are i.i.d., but we also investigate the effects of short memory via a bootstrap experiment. The results are similar to the case with serially independent innovations, but they are not shown in this paper due to length restrictions. All results are available upon request.

Since the errors are i.i.d., we use $q_T = 0$ lags to compute the long-run variance for all the four tests.

3.1 Size

We first consider the size of the tests, so our Data Generating Process (DGP) is $y_t = \varepsilon_t$, with ε_t i.i.d. $t_\infty, \ldots$ or t_1. Our results are displayed in Table 1 and are easy to summarize. For the KPSS and IKPSS, our results are, as expected, very close to the ones obtained by de Jong et al. (2007).

The XL test requires the existence of the first two moments, so it is very undersized under t_2 and Cauchy distributions. Analogously, the KPSS test presents size distortion under Cauchy distribution.

The IKPSS and SS tests are robust to distributions without finite mean and/or variance. However, for very small samples ($T = 100$), the SS test is more conservative than the IKPSS. This happens because we estimate 81 unconditional quantiles in order to compute the SS test. Since the precision of such estimates depends on the density of observations around the quantiles, the performance of the SS test tends to deteriorate in very small samples.

Table 1 Size of the tests at 5% significance level

	t_∞	t_5	t_3	t_2	t_1
$T = 100$					
KPSS	0.049	0.050	0.048	0.044	0.029
IKPSS	0.049	0.050	0.050	0.049	0.050
XL	0.030	0.023	0.019	0.015	0.006
SS	0.039	0.040	0.040	0.040	0.039
$T = 500$					
KPSS	0.050	0.049	0.049	0.045	0.028
IKPSS	0.051	0.050	0.050	0.050	0.049
XL	0.043	0.035	0.027	0.020	0.007
SS	0.050	0.049	0.049	0.049	0.049
$T = 1000$					
KPSS	0.050	0.049	0.049	0.046	0.028
IKPSS	0.050	0.050	0.051	0.049	0.049
XL	0.047	0.039	0.032	0.022	0.007
SS	0.050	0.051	0.051	0.051	0.051

3.2 Power against Alternatives with Unit Root

We parameterize the alternative hypothesis of unit root in a fashion similar to de Jong et al. (2007), $y_t = \lambda r_t + \varepsilon_t$, where $r_t = \sum_{j=1}^{t} \mu_j$ is a random walk, and μ_t and ε_t are i.i.d. and independent from each other, and follow the same distribution (normal, … or Cauchy). The scale factor λ measures the relative importance of the random walk component. We considered $\lambda = 0.01$ and $\lambda = 0.1$.

Table 2 Power of the tests, at 5% significance level, against the alternative hypothesis of unit root

	$\lambda = 0.01$					$\lambda = 0.1$				
	t_∞	t_5	t_3	t_2	t_1	t_∞	t_5	t_3	t_2	t_1
$T = 100$										
KPSS	0.061	0.060	0.064	0.068	0.141	0.588	0.590	0.590	0.587	0.564
IKPSS	0.057	0.060	0.069	0.100	0.477	0.488	0.561	0.633	0.735	0.921
XL	0.035	0.028	0.026	0.029	0.146	0.442	0.468	0.502	0.563	0.679
SS	0.047	0.047	0.055	0.079	0.453	0.500	0.558	0.627	0.739	0.951
$T = 500$										
KPSS	0.307	0.308	0.315	0.337	0.413	0.988	0.987	0.986	0.974	0.873
IKPSS	0.230	0.299	0.394	0.593	0.980	0.972	0.983	0.991	0.997	1.000
XL	0.213	0.211	0.229	0.293	0.513	0.980	0.980	0.982	0.984	0.963
SS	0.241	0.290	0.375	0.571	0.983	0.982	0.989	0.995	0.999	1.000
$T = 1000$										
KPSS	0.606	0.607	0.608	0.608	0.582	1.000	0.999	0.999	0.996	0.934
IKPSS	0.507	0.595	0.697	0.858	0.999	0.998	0.999	1.000	1.000	1.000
XL	0.509	0.512	0.533	0.596	0.713	0.999	0.999	0.999	0.999	0.990
SS	0.539	0.605	0.694	0.853	1.000	0.999	1.000	1.000	1.000	1.000

The results summarized in Table 2 indicate that the power of all the four tests is increasing on both λ and T, as one would expect.

The IKPSS test has more power than the KPSS test for fat tail distributions, but it has less power for normal and t_5 distributions. Actually, the power of the IKPSS test is increasing on the fatness of the tail, which also happens with the SS test. Under normality, the KPSS has more power than the other three tests. Both the SS and the IKPSS tests have more power than the XL test in all cases.

The SS test has performance very similar to the IKPSS test. In all the cases, the SS test has power very close to the winner, when it is not the winner itself. For the infinite mean cases (Cauchy distribution), the SS test is the most powerful test among all the four tests analyzed, except for one case ($T = 100$ and $\lambda = 0.01$).

3.3 Power against Alternatives with Unconditional Heteroskedasticity

Recall that the driving force of the KPSS (IKPSS) test is the fluctuation of the data around the sample mean (median). So they should have low power to detect processes with a constant distribution location, but with a distribution scale that changes over time.

To investigate this possibility, we consider the DGP $y_t = \sqrt{1+st}\varepsilon_t$, so the scale factor is varying over time. We considered $s = 0.01$ and $s = 0.05$. In other words, that the variance, when it exists, is changing linearly over time at rate s.

Table 3 Power of the tests, at 5% significance level, against the alternative hypothesis of time-varying volatility

	$s = 0.01$					$s = 0.05$				
	t_∞	t_5	t_3	t_2	t_1	t_∞	t_5	t_3	t_2	t_1
$T = 100$										
KPSS	0.049	0.049	0.049	0.043	0.028	0.052	0.051	0.050	0.044	0.026
IKPSS	0.051	0.051	0.050	0.051	0.051	0.055	0.056	0.056	0.056	0.056
XL	0.098	0.050	0.032	0.019	0.006	0.388	0.171	0.086	0.039	0.008
SS	0.072	0.064	0.061	0.058	0.051	0.239	0.190	0.167	0.146	0.102
$T = 500$										
KPSS	0.051	0.053	0.050	0.046	0.027	0.053	0.053	0.052	0.046	0.025
IKPSS	0.057	0.057	0.057	0.056	0.057	0.066	0.065	0.065	0.066	0.066
XL	1.000	0.823	0.399	0.110	0.010	1.000	0.943	0.600	0.188	0.012
SS	0.974	0.913	0.851	0.758	0.505	1.000	0.999	0.996	0.984	0.860
$T = 1000$										
KPSS	0.054	0.053	0.051	0.047	0.026	0.056	0.052	0.051	0.046	0.026
IKPSS	0.060	0.060	0.060	0.061	0.060	0.072	0.068	0.070	0.070	0.070
XL	1.000	0.982	0.705	0.210	0.010	1.000	0.989	0.786	0.280	0.013
SS	1.000	1.000	1.000	1.000	0.973	1.000	1.000	1.000	1.000	0.999

Table 3 exhibits our results. Basically, the KPSS test has power equal to size even for large sample sizes ($T = 1000$). In fact, it is a biased test (power less than size) in several instances.

The IKPSS test has power close to size. Even for large samples ($T = 1000$), the maximum power offered by the IKPSS is never more than 0.072.

The XL test has power against this alternative of time-varying scale for thin tail distributions. For the t_2 distribution, its power is low. For the Cauchy distribution, its power is never greater than 0.013, even when $T = 1000$.

The SS test has more power than all the other tests in almost all cases. Even for moderate sample sizes ($T = 500$), it offers power 1, or very close to 1, for almost all distributions.

3.4 Power against Alternative with Time-Varying Kurtosis

Consider a family of real-valued discrete random variables $X(\nu)$ parametrized by $\nu \in [\sqrt{2}, \infty)$ and defined by the following probability mass distribution:

$$P(X(\nu) = x) = \begin{cases} \frac{1}{\nu^2} & \text{if } x = -\frac{\nu}{\sqrt{2}}, \\ 1 - \frac{2}{\nu^2} & \text{if } x = 0, \\ \frac{1}{\nu^2} & \text{if } x = \frac{\nu}{\sqrt{2}}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathrm{e}X(\nu) = 0$, $\mathrm{e}X(\nu)^2 = 1$, $\mathrm{e}X(\nu)^3 = 0$ and $\mathrm{e}X(\nu)^4 = \frac{\nu^2}{2}$, so the expectation, variance and skewness do not vary with ν, but the kurtosis depends on ν. Now, define $\eta_t := X(\sqrt{2} + 8\frac{t}{T})$ and consider the DGP $y_t = \eta_t + \varepsilon_t$. That is, the process is now the error ε_t, that can be distributed as normal, ..., or Cauchy, plus a discrete random variable η_t that has zero mean (and median) and skewness, and unit variance, but has time-varying kurtosis.[12]

It is worthwhile to notice that Kapetanios (2007) says that stationarity tests applied to such processes with changes only in higher unconditional moments have not been analyzed in the literature, and Xiao and Lima (2007) say that many widely used stationarity tests cannot even capture changes in the unconditional variance.

Since the KPSS and the IKPSS tests are not even able to detect time-varying variance when the mean (when it exists) and median are constant over time, they are not able to detect time-varying kurtosis when both the distribution location and the distribution scale are constant over time, as we see in Table 4. Their power and size are about the same.

The XL test presents very low power (never greater than 0.06). Except for the normal distribution case, its power is less than the significance level, 5%.

Our new SS test has good power when the sample size is moderate ($T = 500$ and, specially, $T = 1000$). Our test performs well when the kurtosis exists (normal and t_5 distributions), as one would expect; its power decreases with the fatness of the tail.

These results show that the SS test can reveal lack of stationarity in the data even when they have constant mean (or median), variance and skewness (if they exist). The new test is actually testing the null hypothesis of strict stationarity.

[12] We choose the equation $\nu(t) := \sqrt{2} + 8\frac{t}{T}$ because $X(\nu)$ is not defined for $\nu < \sqrt{2}$ and $P(X(\nu) = 0) > 0.98$ if $\nu > 10$. The results of our simulation are sensitive to the choice of this parametrization. More precisely, the SS test loses some power if $\eta_t \neq 0$ too seldom or too often, as one could expect. However, the other tests have never power against the alternative of time-varying kurtosis, no matter the parametrization we choose.

Table 4 Power of the tests, at 5% significance level, against the alternative hypothesis of time-varying kurtosis

	t_∞	t_5	t_3	t_2	t_1
			$T = 100$		
KPSS	0.049	0.048	0.049	0.045	0.027
IKPSS	0.052	0.051	0.051	0.051	0.052
XL	0.050	0.032	0.024	0.017	0.006
SS	0.085	0.070	0.065	0.060	0.051
			$T = 500$		
KPSS	0.050	0.050	0.048	0.046	0.027
IKPSS	0.051	0.050	0.050	0.052	0.052
XL	0.060	0.044	0.032	0.022	0.007
SS	0.386	0.273	0.221	0.178	0.112
			$T = 1000$		
KPSS	0.049	0.048	0.050	0.046	0.028
IKPSS	0.051	0.050	0.051	0.050	0.051
XL	0.058	0.046	0.035	0.023	0.007
SS	0.723	0.536	0.438	0.342	0.199

4 An Empirical Illustration

In this section we present an empirical analysis in which the use of the SS test can lead to a significant different finding.

We use $T = 4438$ observations of log returns on the S&P 500 index, from 01/03/1991 to 08/11/2008. The first panel of Figure 1 leads us to the belief that the returns r_t exhibit mean reversion, which suggests that the returns r_t do not have a unit root. However note, yet in the first panel, that the variance seems to change over time. Therefore, both the KPSS and the IKPSS tests cannot reject the null hypothesis of stationarity, but that both the XL and the SS tests can, as we can see in the first column of Table 5.

To visualize how the variance is varying on time, we compute variances using a rolling window of length $2h+1$.[13] More specifically, given $h \in \mathbb{N}$, define

$$\left\{V_t^{(h)}\right\}_{t=h+1}^{T-h}, \text{ where } V_t^{(h)} := \frac{1}{2h+1} \sum_{j=t-h}^{t+h} r_j^2.$$

We show the case for $h = 66$, or 3 months, so the total window length is one semester, but the results are similar when we use other windows lengths, available upon request. We plot $V_t^{(66)}$ in the third panel of Figure 1 with a continuous curve.

[13] We use a rolling window instead of an ARCH type variance (Engle, 1982) because we are interested in the unconditional variance rather than the conditional one.

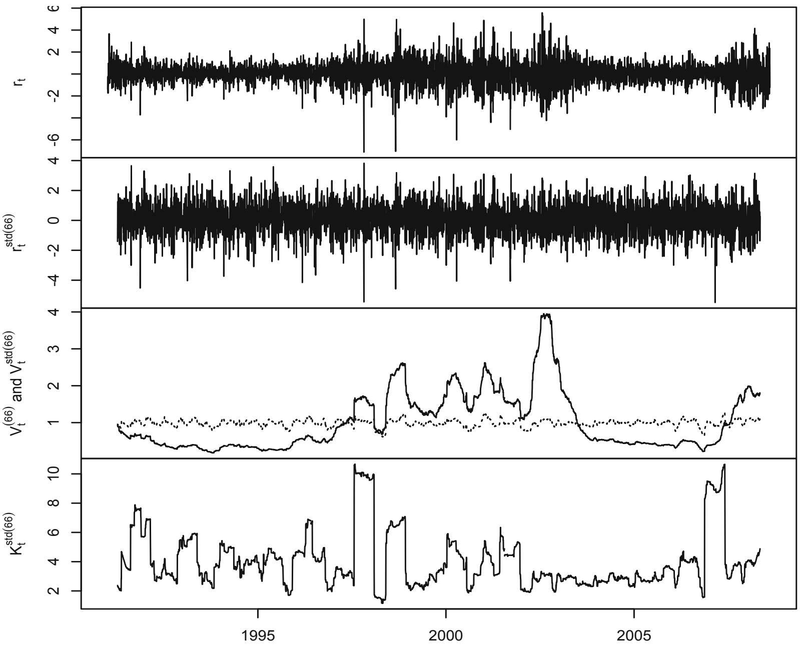


Fig. 1 (1) Plot of the log returns on S&P 500 from 01/03/1991 to 08/11/2008; (2) plot of the standardized returns; (3) plots of the variances of both the returns and the standardized returns; (4) plot of the kurtosis of the standardized returns

Table 5 Tests statistics of the four analyzed tests applied to 4438 observations of returns on the S%P 500 index. To indicate statistical significance at 10%, 5% and 1%, we use *, ** and ***, respectively.

	r_t	$r_t^{std(121)}$	$r_t^{std(66)}$	$r_t^{std(33)}$
Tests Statistics				
KPSS	0.282	0.184	0.178	0.214
IKPSS	0.264	0.262	0.256	0.281
XL	6.298***	1.389	1.083	1.138
SS	4.933***	2.129***	2.153***	2.042***

Now, let us define a standardized return,

$$r_t^{std(h)} := \frac{r_t}{\sqrt{V_t^{(h)}}},$$

plotted in the second panel of Figure 1, and compute its variance using the rolling window, $V_t^{std(h)}$, plotted in the third panel, as a dotted curve.

The variance of the standardized return (the dotted curve) is always around one, so the standardized returns are probably covariance stationary. Indeed, when applying the four stationarity tests to the standardized return, we find the the KPSS, the IKPSS and the XL tests fail to reject the null hypothesis of stationarity. However, the SS test rejects it at any usual significance level.

Perhaps the standardized returns have higher moments that are time-varying. To investigate this, we compute the kurtosis of the standardized returns with a rolling window, $K_t^{std(66)}$, plotted in the fourth and last panel of Figure 1. Indeed, the kurtosis is not constant over time.

In summary, the SS test can capture these fluctuations in higher moments of the returns, and even in higher moments of the standardized returns, so it can strongly reject the null hypothesis of strict stationarity. This empirical exercise casts doubts on results in the literature that are obtained from models that assume stationarity of returns.

5 Conclusion

In this paper we introduce a new test for strict stationarity. We show, through comprehensive Monte Carlo experiments and an empirical exercise that this test is comparable to both classical and new tests for stationarity in terms of power against alternative hypothesis with unit root or unconditional heteroskedasticity.

More importantly, we show that this test has good power against alternative hypotheses with higher moments varying on time, like a time-varying kurtosis, while the other tests fail to reject these hypotheses.

Moreover, the new test is particularly more powerful than the other analyzed tests for fat tail distributions, which makes it very suitable when analyzing financial time series.

References

1. Busetti, F., Harvey, A.: Tests of Time-Invariance, Mimeo, University of Cambridge (2007)
2. de Jong, R.M., Amsler, C., Schmidt, P.: A Robust Version of the KPSS Test Based on Indicators. Journal of Econometrics 137, 311–333 (2007)
3. Engle, R.F.: Autoregressive Conditional Heteroscedasticity with Estimates of The Variance of United Kingdom Inflation. Econometrica 50(4), 987–1007 (1982)
4. Hamilton, J.D.: Time Series Analysis. Princeton University Press (1994)
5. Kapetanios, G.: Testing for Strict Stationarity, Working Paper 602, Queen Mary University of London (2007)
6. Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.: Testing the Null Hypothesis of Stationarity Against the Alternative of Unit Root. Journal of Econometrics 54, 159–178 (1992)
7. Lima, L.R., Neri, B.: Comparing Value-at-Risk Methodologies. Brazilian Review of Econometrics 27(1), 1–25 (2007)

8. Newey, W.K., West, K.D.: A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55(3), 703–708 (1987)
9. Qu, Z.: Test for Structural Change in Regression Quantiles, Working paper, University of Illinois (2005)
10. R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2008)
11. White, H.: Asymptotic Theory for Econometricians, Revised Edition. Academic Press (2001)
12. Xiao, Z., Lima, L.R.: Testing Covariance Stationarity. Econometric Reviews 26(6), 643–667 (2007)

Part II
Fundamental Theory

Statistical Inference from Ill-known Data Using Belief Functions

Thierry Denœux

Abstract. As a general formalism for uncertain reasoning, the theory of belief functions extends the logical and probabilistic approaches to uncertainty: a belief function (or a completely monotone Choquet capacity) can be seen both as a non additive measure and as a generalized set. In this paper, the theory of belief functions is argued to be a suitable framework for statistical analysis of low quality, i.e., imprecise and/or partially reliable data. After a reminder of general concepts of the theory, we show how this approach can be applied to statistical inference by viewing the normalized likelihood function as defining a consonant belief function. The links with likelihood-based and Bayesian inference are discussed. We then show how this method can be extended to the analysis of uncertain data. The approach is illustrated using a running example.

1 Introduction

Whereas current research in statistics and econometrics mainly focuses on the development of more complex models and inference procedures, data quality is recognized by applied statisticians as a key factor influencing the validity of the conclusions drawn from a statistical analysis. As noted by Cox [5], "issues of data quality and relevance, while underemphasized in the theoretical statistical and econometric literature, are certainly of great concern in much statistical work". Arguing for better consideration of empirical practice in econometric theory, Heckman [22] also remarked that "Data quality, data collection and economic interpretation of statistical evidence are perceived as topics off limits to econometricians, but central to the field of empirical economics".

Thierry Denœux
UMR CNRS 7253 Heudiasyc,
Université de Technologie de Compiègne,
BP 20529 - F-60205 Compiègne cedex - France
e-mail: Thierry.Denoeux@hds.utc.fr

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 33–48.
DOI: 10.1007/978-3-642-35443-4_3

One of the reasons why data quality, in spite of its importance, has received relatively little attention in the statistical literature, may be that its evaluation often requires subjective judgements that do not easily fit with the standard likelihood-based or Bayesian frameworks. While the latter approach allows for the introduction of personalistic prior information, it does so in a very specific and questionable manner (by treating all unknown quantities as random variables), which raises a number of theoretical and practical issues [36, 16].

In the past thirty years, alternatives to the Bayesian framework for reasoning from weak information have emerged, including Possibility Theory [39], Imprecise Probabilities [36] and the theory of Belief Functions [7, 25]. In particular, the latter approach, also referred to as Dempster-Shafer or Evidence theory, was introduced by Dempster [6, 8] with the objective to reconcile Bayesian and fiducial inference. Shafer [25] later formalized this approach as a general method for representing and combining evidence, not necessarily statistical. Smets [29, 33] emphasized the singularity of the theory of belief functions as opposed to related but distinct frameworks such as imprecise probabilities [36] and random sets [24].

The main feature of the theory of belief functions is that is subsumes both the logical and probabilistic approaches to uncertainty: a belief function may be seen as a non-additive probability measure [25] and as a generalized set [18]. Also, basic mechanisms for reasoning with belief functions extend both probabilistic operations (such as marginalization and conditioning) and set-theoretic operations (such as intersection and union). In particular, the belief function approach coincides with the Bayesian approach when all variables are described by probability distributions, while allowing for considerably more flexibility when the available knowledge does not allow for the specification of a reasonable probability distributions without introducing unsupported assumptions.

In this paper, the theory of belief functions is advocated as a suitable framework for statistical analysis of low quality, i.e., imprecise and/or partially reliable data. The main concepts of the theory will first be recalled in Section 2 and its application to the representation of statistical evidence will be discussed in Section 3. The use of belief functions for representing data uncertainty and corresponding inferential procedures will be introduced in Section 4. Finally, Section 5 will conclude the paper with a summary of the main results and the presentation of some research challenges.

2 Belief Functions

This section recalls the necessary background notions related to Dempster-Shafer theory. Belief functions on finite domains and Dempster's rule of combination are first presented in Subsections 2.1 and 2.2, respectively. Some notions regarding the definition and manipulation of belief functions on continuous domains are then recalled in Subsection 2.3.

2.1 Belief Functions on Finite Domains

Let $\boldsymbol{\theta}$ be a variable taking values in a finite domain Θ, called the *frame of discernment*. Uncertain evidence about $\boldsymbol{\theta}$ may be represented by a *mass function* m on Θ, defined as a function from the powerset of Θ, denoted as 2^Θ, to the interval $[0,1]$, such that $m(\emptyset) = 0$ and

$$\sum_{A \subseteq \Theta} m(A) = 1. \tag{1}$$

Any subset A of Θ such that $m(A) > 0$ is called a *focal set* of m. A categorical mass function has only one focal set (it is thus equivalent to a set), while a Bayesian mass function has only focal sets of cardinality one and is thus equivalent to a probability distribution. The mass function m such that $m(\Theta) = 1$ is said to be vacuous.

Each number $m(A)$ is interpreted as a *degree of belief* attached to the proposition $\boldsymbol{\theta} \in A$ and to *no more specific proposition*, based on some evidence. As argued by Shafer [27], the meaning of such degrees of belief can be better understood by assuming that we have compared our evidence to a canonical chance set-up. The set-up proposed by Shafer consists of an encoded message and a set of codes $\Omega = \{\omega_1, \ldots, \omega_n\}$, exactly one of which is selected at random. We know the list of codes as well as the chance p_i of each code ω_i being selected. Decoding the encoded message using code ω_i produces a message of the form "$\boldsymbol{\theta} \in A_i$" for some $A_i \subseteq \Theta$. Then

$$m(A) = \sum_{\{1 \le i \le n : A_i = A\}} p_i \tag{2}$$

is the chance that the original message was "$\boldsymbol{\theta} \in A$". Stated differently, it is the probability of knowing that $\boldsymbol{\theta} \in A$. In particular, $m(\Theta)$ is, in this setting, the probability that the original message was vacuous, i.e., the probability of knowing nothing.

The above setting thus consists of a set Ω, a probability measure P on Ω and a multi-valued mapping $\Gamma : \Omega \to 2^\Theta \setminus \{\emptyset\}$ such that $A_i = \Gamma(\omega_i)$ for each $\omega_i \in \Omega$. This is the framework initially considered by Dempster in [7]. The triple (Ω, P, Γ) formally defines a finite *random set* [24]: mass functions are thus exactly equivalent to random sets from a mathematical point of view. However, the meaning of mass functions differs from the usual interpretation of a random set as the outcome of a random experiment: here, $m(A)$ is *not* the chance that A was selected, but it can be viewed as the chance of the evidence meaning that $\boldsymbol{\theta}$ is in A [27].

To each normalized mass function m, we may associate belief and plausibility functions from 2^Θ to $[0,1]$ defined as follows:

$$Bel(A) = P(\{\omega \in \Omega | \Gamma(\omega) \subseteq A\}) = \sum_{B \subseteq A} m(B) \tag{3a}$$

$$Pl(A) = P(\{\omega \in \Omega | \Gamma(\omega) \cap A \neq \emptyset\}) = \sum_{B \cap A \neq \emptyset} m(B), \tag{3b}$$

for all $A \subseteq \Theta$. These two functions are linked by the relation $Pl(A) = 1 - Bel(\overline{A})$, for all $A \subseteq \Theta$. Each quantity $Bel(A)$ may be interpreted as the degree to which the evidence *supports* A, while $Pl(A)$ can be interpreted as the degree to which the evidence *is not contradictory* with A. The following inequalities always hold: $Bel(A) \leq Pl(A)$, for all $A \subseteq \Theta$. The function $pl : \Theta \to [0,1]$ such that $pl(\theta) = Pl(\{\theta\})$ is called the *contour function* associated to m.

If m is Bayesian, then function Bel is identical to Pl and it is a probability measure, and pl is the corresponding probability mass function. Another special case of interest is that where m is *consonant*, i.e., its focal elements are nested. The plausibility function is then a *possibility measure* [39, 19] with possibility distribution pl, i.e., the plausibility function can be recovered from the contour function as follows: [25]:

$$Pl(A) = \max_{\theta \in A} pl(\theta). \tag{4}$$

for all $A \subseteq \Theta$.

Given two mass functions m_1 and m_2, m_1 is said to be *less specific* than m_2 if it can be obtained from m_2 by transferring belief masses $m_2(A)$ to supersets $B \supseteq A$ [38, 18]. In this case, m_1 can be considered as less informative, or less committed[1] than m_2. The *Least Commitment Principle* (LCP) [31] states that, given some constraints on an unknown mass function, the least committed should be selected. This principle provides a justification of consonant mass functions: given a function $\pi : \Theta \to [0,1]$ such that $\max \pi = 1$, the least specific mass function m with contour function pl such that $pl = \pi$ is consonant; its plausibility function, given by (4), will be denoted as pl^*.

2.2 Dempster's Rule

A key idea in Dempster-Shafer theory is that beliefs are elaborated by aggregating different items of evidence. The basic mechanism for evidence combination is Dempster's rule of combination, which can be naturally derived using the random code metaphor as follows.

Let m_1 and m_2 be two mass functions induced by triples $(\Omega_1, P_1, \Gamma_1)$ and $(\Omega_2, P_2, \Gamma_2)$ interpreted under the random code framework as before. Let us further assume that the codes are selected independently. For any two codes $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$, the probability that they both are selected is then $P_1(\{\omega_1\})P_2(\{\omega_2\})$, in which case we can conclude that $\boldsymbol{\theta} \in \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$. If $\Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) = \emptyset$, we know that the pair of codes (ω_1, ω_2) could not have been selected: consequently, the joint probability distribution on $\Omega_1 \times \Omega_2$ must be conditioned, eliminating such pairs [27]. This line of reasoning yields the following combination rule, referred to as Dempster's rule [25]:

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{5}$$

for all $A \subseteq \Theta$, $A \neq \emptyset$ and $(m_1 \oplus m_2)(\emptyset) = 0$, where

[1] Alternative comparative orderings between belief functions have been proposed, see, e.g., [18].

$$\kappa = \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{6}$$

is the *degree of conflict* between m_1 and m_2. If $\kappa = 1$, there is a logical contradiction between the two pieces of evidence and they cannot be combined. Dempster's rule is commutative, associative, and it admits as neutral element the *vacuous* mass function defined as $m(\Omega) = 1$.

Dempster's rule can be easily expressed in terms of contour functions: if pl_1 and pl_2 are the contour functions of two mass functions m_1 and m_2, then the contour function of $m_1 \oplus m_2$ is, using the same symbol $\oplus$ as used for mass functions and contour functions

$$(pl_1 \oplus pl_2)(\theta) = \frac{pl_1(\theta) pl_2(\theta)}{1 - \kappa} \tag{7}$$

for all $\theta \in \Theta$, where κ is the degree of conflict. If m_1 or m_2 is Bayesian, then so is m_1 and m_2 and the degree of conflict is then

$$\kappa = 1 - \sum_{\theta \in \Theta} pl_1(\theta) pl_2(\theta). \tag{8}$$

2.3 Random Real Intervals

The definition of belief functions and random sets in infinite spaces implies greater mathematical sophistication than it does in finite spaces [26, 24]. Here, we will restrict our discussion to random closed intervals on the real line (see, e.g., [9, 32, 11]), which constitute a simple yet sufficiently general framework for expressing beliefs on a real variable.

Let $(\Omega, \mathscr{A}, P)$ be a probability space and $(U,V) : \Omega \to \mathbb{R}^2$ a two-dimensional real random vector such that $P(\{\omega \in \Omega | U(\omega) \leq V(\omega)\}) = 1$. Let Γ be the multi-valued mapping that maps each $\omega \in \Omega$ to the closed interval $[U(\omega), V(\omega)]$. This setting defines a random interval, as well as belief and plausibility functions on $\mathbb{R}$ defined, respectively, by

$$Bel(A) = P(\{\omega \in \Omega | [U(\omega), V(\omega)] \subseteq A\}) \tag{9}$$
$$Pl(A) = P(\{\omega \in \Omega | [U(\omega), V(\omega)] \cap A \neq \emptyset\}) \tag{10}$$

for all elements A of the Borel sigma-algebra $\mathscr{B}(\mathbb{R})$ on the real line [9]. The intervals $[U(\omega), V(\omega)]$ are referred to as the focal intervals of $[U,V]$. We note that, when U and V are continuous, the notion of mass function should be replaced by that of mass density function defined by $m([u,v]) = p(u,v)$, where $p(u,v)$ denotes the joint probability density function (pdf) of (U,V). To simplify the terminology, we will continue to use the term "mass function" in this case.

If $U = V$, then we have a random point, which is equivalent to a real random variable. Another special case of interest is that of consonant random closed intervals defined as follows. Let $\Omega = [0,1]$ and$\pi : \mathbb{R} \to [0,1]$ a function such that, for each $\omega \in \Omega$,

$$\Gamma(\omega) = \{x \in \mathbb{R} | \pi(x) \geq \omega\}$$

is a closed interval $[U(\omega), V(\omega)]$. Finally, let P denote the Lebesgue measure on Ω. Then, $[U,V]$ is a random interval and π is its contour function, i.e., $pl(x) = Pl(\{x\}) = \pi(x)$ for all $x \in \mathbb{R}$. Such a random interval is said to be consonant because its focal intervals $\Gamma(\omega)$ are nested.

Dempster's rule can be defined for random intervals as follows. Let us assume that we have two random intervals $(\Omega_i, \mathscr{A}_i, P_i, \Gamma_i)$ with $i = 1,2$ and $[U_i(\omega), V_i(\omega)] = \Gamma_i(\omega)$. Let Γ_{12} be the mapping from $\Omega_1 \times \Omega_2$ to the set of closed real intervals defined by

$$\Gamma_{12}(\omega_1, \omega_2) = \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2), \quad \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$$

and let P_{12} be the product measure $P_1 \times P_2$ conditioned on the set $\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 | \Gamma_{12}(\omega_1, \omega_2) \neq \emptyset\}$. Then, $(\Omega_1 \times \Omega_2, \mathscr{A}_1 \times \mathscr{A}_2, P_{12}, \Gamma_{12})$ define a random interval $[U_{12}, V_{12}] = [U_1, V_1] \oplus [U_2, V_2]$. Its contour function is

$$(pl_1 \oplus pl_2)(x) = \frac{pl_1(x) pl_2(x)}{1 - \kappa}$$

for all $x \in \mathbb{R}$, where κ is the degree of conflict between the two random intervals defined as:

$$\kappa = P(\{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 | \Gamma_{12}(\omega_1, \omega_2) \neq \emptyset\}).$$

In general, the combination of two random intervals by Dempster's rule is not easy to compute analytically. However, a special case in which the computations are very simple is that were a random point with pdf p_1 is combined with a random interval with contour function pl_2. The results is a random point with pdf

$$(p_1 \oplus pl_2)(x) = \frac{p_1(x) pl_2(x)}{1 - \kappa}, \tag{11}$$

where the degree of conflict κ is

$$\kappa = 1 - \int_{-\infty}^{+\infty} p_1(x) pl_2(x) dx. \tag{12}$$

3 Modeling Statistical Evidence

Let us now turn our attention to the representation of statistical evidence. Assume that we have observed a realization $\mathbf{x}$ of a random vector $\mathbf{X}$ with pdf $p(\mathbf{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$ is an unknown parameter. What does this item of evidence tell us about $\boldsymbol{\theta}$? Shafer's solution [25] derived from the Likelihood and Least Commitment principles will first be recalled in Subsection 3.1. Arguments for and against this solution will then be discussed in Subsection 3.2 and an illustrative example will be presented in Subsection 3.3.

3.1 Least Committed Solution Based on Likelihoods

In the standard statistical framework, information about $\boldsymbol{\theta}$ is typically assumed to be represented by the likelihood function defined by $L(\theta;\mathbf{x}) = p(\mathbf{x};\theta)$ for all $\theta \in \Theta$. More precisely, the likelihood principle [2] [3] [20, chapter 3] states that "Within the framework of a statistical model, *all* the information which the data provide concerning the relative merits of two hypotheses is contained in the likelihood ratio of these hypotheses on the data". In statistical parlance, the likelihood ratio is often referred to as the "relative plausibility", which suggests translating the likelihood ratio in the belief function framework as follows:

$$\frac{pl(\theta_1;\mathbf{x})}{pl(\theta_2;\mathbf{x})} = \frac{L(\theta_1;\mathbf{x})}{L(\theta_2;\mathbf{x})},$$

for all $(\theta_1, \theta_2) \in \Theta^2$ or, equivalently,

$$pl(\theta;\mathbf{x}) = cL(\theta;\mathbf{x})$$

for all $\theta \in \Theta$ and some positive constant c. The LCP then leads us to giving the highest possible value to constant c, i.e., defining pl as the relative likelihood :

$$pl(\theta;\mathbf{x}) = \frac{L(\theta;\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta;\mathbf{x})} \tag{13}$$

and representing evidence about $\boldsymbol{\theta}$ by the least committed plausibility function induced by pl, i.e.,

$$Pl(A;\mathbf{x}) = \sup_{\theta\in A} pl(\theta;\mathbf{x}) = \frac{\sup_{\theta\in A} L(\theta;\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta;\mathbf{x})}, \tag{14}$$

for all $A \subseteq \Theta$. The corresponding belief function is called a likelihood-based belief function by Wasserman [37].

3.2 Discussion

Equation (14) was first proposed by Shafer in [25, chapter 11] who, however, did not justify it by the LCP, but by the more questionable requirement that the belief function on Θ be consonant. In the special case where $\Theta = \{\theta_1, \theta_2\}$ has only two points, Wasserman [37] showed that the plausibility function (14) corresponds to the unique belief function $Bel(\cdot;\mathbf{x})$ verifying the following requirements:

1. If $L(\theta_1;\mathbf{x}) = L(\theta_2;\mathbf{x})$, then $Bel(\cdot;\mathbf{x})$ should be vacuous;
2. $Bel(\{\theta\};\mathbf{x})$ should be nondecreasing in $L(\theta;\mathbf{x})$;
3. If $Bel = Bel(\cdot;\mathbf{x}) \oplus P_0$ and P_0 is a probability measure, then Bel should be equal to the Bayesian posterior.

This argument can be extended to the case where Θ is a complete, separable metric space [37].

One of the main criticisms against the use of the likelihood-based plausibility function (14) for represented statistical evidence is its incompatibility with Dempster's rule in the case of independent observations [28]. More precisely, assume that $\mathbf{X}$ is an independent sample $(X_1,\ldots,X_n)$ and each observation X_i has a marginal pdf $p(x_i;\boldsymbol{\theta})$ depending on $\boldsymbol{\theta}$. We could combine the n observations at the "aleatory level" by computing $Pl(\cdot;\mathbf{x})$ using (14), or we could combine them at the "epistemic level" by first computing the consonant plausibility functions $Pl(\cdot;x_i)$ induced by each of the independent observations and applying Dempster's rule. Obviously, these two procedures yield different results in general, as consonance is not preserved by Dempster's rule.

Shafer [28] seems to have regarded the above argument as strong enough to reject (14) as a reasonable method to represent statistical evidence. However, Aickin [1] proposed to keep (14) but questioned Dempster's rule as a mechanism for combining statistical evidence. Additional arguments against the use of Dempster's rule for combining evidence from independent observations can be found in [35].

Based on the above discussion, we propose to adopt (13) and (14) as models of statistical evidence. Further arguments in favor of this approach are summarized below:

1. This method of inference is considerably simpler than other methods such as Dempster's initial proposal [8] and other methods discussed in [28], while being more widely applicable than Smets' Generalized Bayesian Theorem [30, 17].
2. Combining $Pl(\cdot;\mathbf{x})$ given by (14) with a Bayesian prior P_0 on Θ using Dempster's rule yields a Bayesian plausibility function $Pl(\cdot;\mathbf{x}) \oplus P_0$ which is identical to the posterior probability obtained using Bayes' rule: consequently, the proposed method of inference boils down to Bayesian inference when a Bayesian prior is available.
3. Finally, viewing the relative likelihood function as a possibility distribution seems to be consistent with statistical practice, although this point of view has not been adopted explicitly in the statistical literature. For instance, likelihood intervals [23, 34] are focal intervals of the relative likelihood viewed as a possibility distribution. In the case where $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ and θ_2 is considered as a nuisance parameter, the relative profile likelihood function can be written

$$pl(\theta_1;\mathbf{x}) = \sup_{\theta_2 \in \Theta_2} pl(\theta_1, \theta_2;\mathbf{x}),$$

which is the marginal possibility distribution on Θ_1. Eventually, we can remark that the usual likelihood ratio statistics $\Lambda(\mathbf{x})$ for a composite hypothesis $H_0 \subset \Theta$ is nothing but the plausibility of H_0, as

$$\Lambda(\mathbf{x}) = \frac{\sup_{\theta \in H_0} L(\theta;\mathbf{x})}{\sup_{\theta \in \theta} L(\theta;\mathbf{x})} = \sup_{\theta \in H_0} pl(\theta;\mathbf{x}) = Pl(H_0;\mathbf{x}).$$

3.3 *Illustrative Example*

As a concrete example, let us consider the following problem using a real dataset. Average public teacher pay and spending on public schools per pupil in 1985 for 49 states and the District of Columbia were reported by the Albuquerque Tribune[2]. The data are plotted in Figure 1 for each of the three areas: Northeast and North Central, South and West. We can see that public teacher pay is approximately linearly related to spending on public schools. Is there any statistical evidence of different relations holding in the three regions?

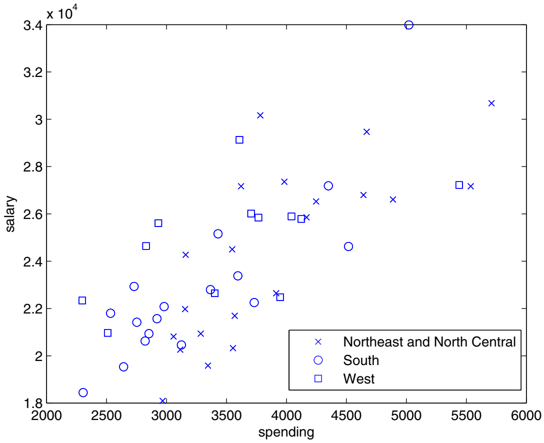


Fig. 1 Average public school teacher annual salary ($) as a function of spending on public schools per pupil ($) for 49 states and the District of Columbia

Let y_{ki} and x_{ki} denote, respectively, the teacher pay and spending on public schools in state i of region k ($k = 1, 2, 3$). We assume that $\mathbf{y}_k = \{y_{ki}\}_{i=1}^{n_k}$ is a realization of a Gaussian random vector $\mathbf{Y}_k \sim \mathcal{N}(\mathbf{X}_k \mathbf{b}_k, \sigma_k^2 I_n)$, where $\mathbf{X}_k$ is the fixed design matrix with line i equal to $(1, x_{ki})$, I_n is the identity matrix of size n, and $\boldsymbol{\theta}_k = (\mathbf{b}_k, \sigma_k)'$ is the parameter vector.

Figure 2 shows the contour functions $pl(\mathbf{b}_k; \mathbf{y}_k)$. We recall that this function is obtained as the relative profile likelihood function considering variance as a nuisance parameter, i.e.,

$$pl(\mathbf{b}_k; \mathbf{y}_k) = \sup_{\sigma_k > 0} pl(\mathbf{b}_i, \sigma_k; \mathbf{y}_k) = \frac{\sup_{\sigma_k > 0} L(\mathbf{b}_k, \sigma_k; \mathbf{y}_k)}{\sup_{\mathbf{b}_k \in \mathbb{R}^2, \sigma_i > 0} L(\mathbf{b}_k, \sigma_k; \mathbf{y}_k)},$$

[2] The dataset can be downloaded from the Data and Story Library at `http://lib.stat.cmu.edu/DASL`. The data for Alaska is an outlier and was not considered in this analysis

with

$$L(\mathbf{b}_k, \sigma_k; \mathbf{y}_k) = \phi(\mathbf{y}_k; \mathbf{X}_k' \mathbf{b}_k, \sigma_k^2 I_n) = \prod_{i=1}^{n} \phi(y_{ki}; (1, x_{ki})' \mathbf{b}_k, \sigma_k^2),$$

We can see from Figure 2(d) that the contour at level 0.1 for region 3 does not intersect the corresponding contour for region 2, which suggests that $\mathbf{b}_3$ is different from $\mathbf{b}_2$ with a high plausibility. To carry the analysis further, we can compute the plausibilities $Pl(\mathbf{b}_i = \mathbf{b}_j)$ for each pair of regions, as well as the plausibility $Pl(\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3)$ that the three parameters are equal. It is easy to see [15] that these plausibilities are equal to one minus the degree of conflict between the belief functions related to each parameter. These degrees of conflict are not easy to compute analytically, but they can be estimated by Monte-Carlo simulation. This is achieved by picking a focal set at random independently for each of the belief function, and estimating the probability for the focal sets to be disjoint. We obtain the following values:

$$Pl(\mathbf{b}_1 = \mathbf{b}_2) = 0.70, \quad Pl(\mathbf{b}_1 = \mathbf{b}_3) = 0.12, \quad Pl(\mathbf{b}_2 = \mathbf{b}_3) = 0.02$$

$$Pl(\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3) = 0.01.$$

which confirms that the hypotheses $\mathbf{b}_2 = \mathbf{b}_3$ and $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3$ can be discarded as having very small plausibility.

4 Inference from Uncertain Data

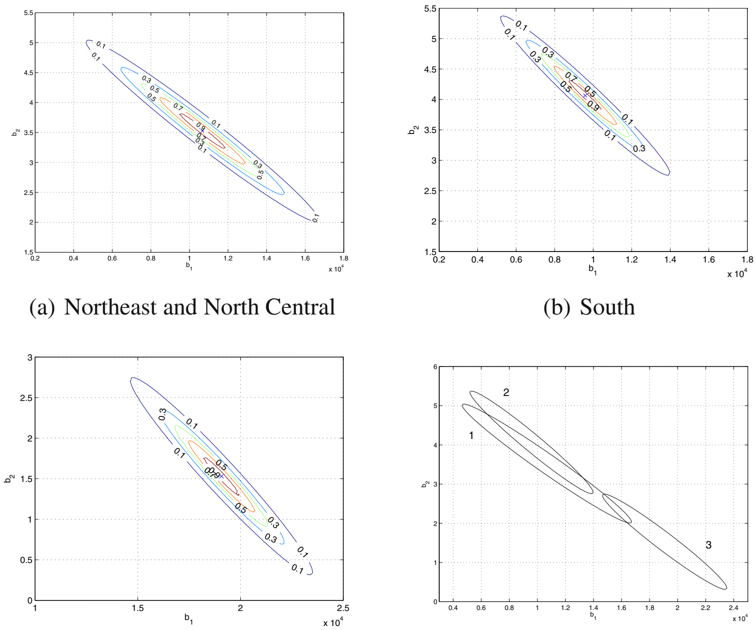
We consider in this section the situation where the data $\mathbf{x}$ have been generated by a random process but have been imperfectly observed [12, 13, 14]. Our partial knowledge of $\mathbf{x}$ will then be described by a mass function m on the data space $\Omega_X \subseteq \mathbb{R}^d$. Our objective will be to find a suitable representation of the information about the parameter provided by such data, in the belief function framework. Our approach will be to generalize the likelihood function and, as before, to consider the relative likelihood as the contour function of a consonant plausibility measure.

Before we describe our approach, it must be emphasized that, in this model, the pdf or probability mass function $p(\mathbf{x}, \boldsymbol{\theta})$ and the Dempster-Shafer mass function m represent two different pieces of knowledge:

- $p(\mathbf{x}, \boldsymbol{\theta})$ represents *generic* knowledge about the data generating process or, equivalently, about the underlying population; it corresponds to *random uncertainty*;
- m represents *specific* knowledge about a given realization $\mathbf{x}$ of $\mathbf{X}$; this knowledge is only partial because the observation process is imperfect; function m captures *epistemic uncertainty*, i.e., uncertainty due to lack of knowledge.

The uncertain data m is thus not assumed to be produced by a random experiment, which is in sharp contrast with other approaches based on random sets [24] or fuzzy random variables [21].

Our approach will first be described in Subsection 4.1. The impact of stochastic and cognitive independence assumptions will then be examined in Subsection 4.2.



(a) Northeast and North Central (b) South

(c) West (d) 0.1-level contours

Fig. 2 Contour functions $pl(\mathbf{b}_k;\mathbf{y}_k)$ for each of the three regions (a-c) and 0.1-level contours (d). Please note that the x and y axes have different ranges in the four plots.

4.1 Representation of Uncertain Statistical Evidence

Let us assume that the mass function m is induced by a random set $(\Omega,\mathscr{A},P,\Gamma)$. We will further assume that one of the following two conditions holds:

- $\mathbf{X}$ is discrete, or
- $\mathbf{X}$ is continuous an if $\Gamma(\omega)$ is not reduced to a point (which would correspond an infinite precision).

Under these assumptions, the probability of observing the result $\Gamma(\omega)$ given that the interpretation $\omega\in\Omega$ holds is

$$P(\Gamma(\omega);\theta)=\int_{\Gamma(\omega)}p(\mathbf{x};\theta)d\mathbf{x},$$

assuming that the integral in the right-hand side is well defined. The probability of the uncertain observation m may then defined as the average of $P(\Gamma(\omega);\theta)$ over $\omega\in\Omega$, which can be written as

$$P(m;\theta)=\sum_{\omega\in\Omega}p(\omega)P(\Gamma(\omega);\theta)$$

if Ω is finite and

$$P(m;\theta) = \int_\Omega p(\omega) P(\Gamma(\omega);\theta) d\omega$$

otherwise, assuming this integral to be defined. The likelihood function given the uncertain observation m can then be defined as $L(\theta;m) = P(m;\theta)$ for all $\theta \in \Theta$. It is easy to show that $L(\theta;m)$ only depends on the contour function. To see this, we may write:

$$L(\theta;m) = \int_\Omega p(\omega) \left(\int_{\Gamma(\omega)} p(\mathbf{x};\theta) d\mathbf{x} \right) d\omega, \tag{15}$$

$$= \int_{\Omega_X} p(\mathbf{x};\theta) \left(\int_{\{\omega | \Gamma(\omega) \ni \mathbf{x}\}} p(\omega) d\omega \right) d\mathbf{x}, \tag{16}$$

$$= \int_{\Omega_X} p(\mathbf{x};\theta) pl(\mathbf{x}) d\mathbf{x} \tag{17}$$

$$= \mathbb{E}_\theta \left[pl(\mathbf{X}) \right]. \tag{18}$$

As a natural extension of (13), we propose to represent the information on $\boldsymbol{\theta}$ provided by the uncertain data by the consonant plausibility function with the following contour function:

$$pl(\theta;m) = \frac{L(\theta;m)}{\sup_{\theta \in \Theta} L(\theta;m)}. \tag{19}$$

An iterative procedure for finding a value $\widehat{\theta}$ of θ that maximizes $pl(\theta;m)$ has been introduced in [4] and generalized in [12, 14]. This procedure, called the Evidential Expectation Maximization (E^{2}M) algorithm, is an extension of the EM algorithm [10].

4.2 *Independence Assumptions*

Let us assume that the random vector $\mathbf{X}$ can be written as $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, where each $\mathbf{X}_i$ is a p-dimensional random vector taking values in $\Omega_{\mathbf{X}_i}$. Similarly, its realization can be written as $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Omega_{\mathbf{X}}$. Two different independence assumptions can then be made:

1. Under the *stochastic independence* of the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the pdf or probability mass function $p(\mathbf{x};\boldsymbol{\theta})$ can be decomposed as:

$$p(\mathbf{x};\boldsymbol{\theta}) = \prod_{i=1}^n p(\mathbf{x}_i;\boldsymbol{\theta}), \quad \forall \mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Omega_{\mathbf{X}} \tag{20}$$

2. Under the *cognitive independence* of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with respect to m (see [25, page 149]), we can write:

$$pl(\mathbf{x}) = \prod_{i=1}^n pl_i(\mathbf{x}_i), \quad \forall \mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Omega_{\mathbf{X}}, \tag{21}$$

where pl_i is the contour function corresponding to the mass function m_i obtained by marginalizing m on $\Omega_{\mathbf{X}_i}$.

We can remark here that the two assumptions above are totally unrelated as they are of different natures: stochastic independence of the random variables $\mathbf{X}_i$ is an objective property of the random data generating process, whereas cognitive independence pertains to our state of knowledge about the unknown realization $\mathbf{x}$ of $\mathbf{X}$.

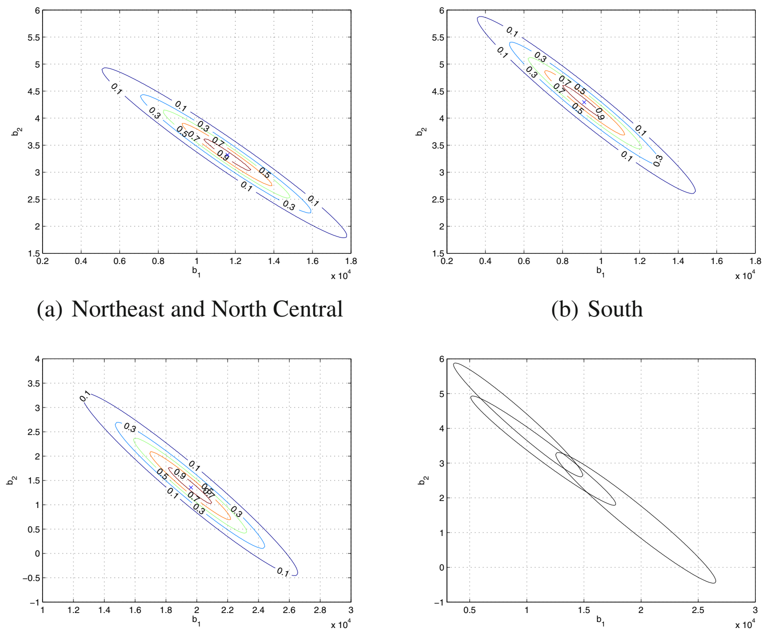
If both assumptions hold, the likelihood criterion (18) can be written as a product of n terms:

$$L(\boldsymbol{\theta};m)=\prod_{i=1}^{n}\mathbb{E}_{\boldsymbol{\theta}}\left[pl_i(\mathbf{X}_i)\right] \tag{22}$$

and $pl(\theta;m)$ can be written as:

$$pl(\theta;m)=\frac{\prod_{i=1}^{n}pl(\theta;m_i)}{\sup_{\theta\in\Theta}\prod_{i=1}^{n}pl(\theta;m_i)}. \tag{23}$$

Example 1. *Let us come back to the analysis of Subsection 3.3, this time assuming that the observations of the dependent variable are uncertain. This is reasonable if we assume that teacher pay data for each state are not known exactly but are estimated by surveys carried out with samples of different sizes and under different*



(a) Northeast and North Central (b) South

(c) West (d) 0.1-level contours

Fig. 3 Contour functions $pl(\mathbf{b}_k;\mathbf{y}_k)$ for each of the three regions (a-c) and 0.1-level contours (d), with simulated data uncertainty. Please note that the x and y axes have different ranges in the four plots.

conditions. As we do not know in which conditions the data were collected, we simulated data uncertainty by assuming the contour functions $pl_{ki}(y_{ki})$ *to be normalized Gaussians centered at each data point and with standard deviation* s_{ki} *selected at random from a uniform distribution in* $[0, 5000]$.

The results are shown in Figure 3. We can see that the consideration of data uncertainty actually leads to less committed plausibility functions in the parameter space. The plausibility values for the same hypotheses as considered in Subsection 3.3 are now:

$$Pl(\mathbf{b}_1 = \mathbf{b}_2) = 0.61, \quad Pl(\mathbf{b}_1 = \mathbf{b}_3) = 0.39, \quad Pl(\mathbf{b}_2 = \mathbf{b}_3) = 0.13,$$

$$Pl(\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3) = 0.08,$$

which shows that the hypotheses $\mathbf{b}_2 = \mathbf{b}_3$ *and* $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3$ *can no longer be rejected based on the uncertain statistical evidence.*

5 Conclusion

The Dempster-Shafer theory of belief functions places emphasis on the representation of evidence for evaluating degrees of belief. This generality and flexibility of this framework makes it suitable for representing and combining expert judgments and statistical evidence.

In this paper, we have focused on the representation of statistical evidence, seeing the relative likelihood function as the contour function of a consonant belief function in the parameter space, as originally proposed by Shafer. Likelihood-based and Bayesian inference schemes can both be seen as special cases of this approach.

We have shown that this method can be extended in a simple way to the representation of uncertain statistical evidence or ill-known data, where lack of knowledge comes from imperfectness of the observation process. Maximum plausibility estimation can still be performed in this case using a computationally simple iterative procedure that extends the EM algorithm.

A interesting perspective of this approach concerns situations in which statistical evidence needs to be combined with expert judgements. Such problems typically arise in climate change studies, in which statistical data cannot be considered as a unique source of information but have to be pooled with expert opinions summarizing findings from physical modeling. Results concerning the application of the belief approach to such problems will be reported in future publications.

References

1. Aickin, M.: Connecting Dempster-Shafer belief functions with likelihood-based inference. Synthese 123, 347–364 (2000)
2. Barnard, G.A., Jenkins, G.M., Winsten, C.B.: Likelihood inference and time series. Journal of the Royal Statistical Society 125(3), 321–372 (1962)

3. Birnbaum, A.: On the foundations of statistical inference. Journal of the American Statistical Association 57(298), 269–306 (1962)
4. Côme, E., Oukhellou, L., Denœux, T., Aknin, P.: Learning from partially supervised data using mixture models and belief functions. Pattern Recognition 42(3), 334–348 (2009)
5. Cox, D.R.: Some remarks on statistical aspects of econometrics. In: Panaretos, J. (ed.) Stochastics Musings. Lawrence Erlbaum, Mahwah (2003)
6. Dempster, A.P.: New methods for reasoning towards posterior distributions based on sample data. Annals of Mathematical Statistics 37, 355–374 (1966)
7. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
8. Dempster, A.P.: A generalization of Bayesian inference (with discussion). J. R. Statistical Society B 30, 205–247 (1968)
9. Dempster, A.P.: Upper and lower probabilities generated by a random closed interval. Annals of Mathematical Statistics 39(3), 957–966 (1968)
10. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, B 39, 1–38 (1977)
11. Denœux, T.: Extending stochastic ordering to belief functions on the real line. Information Sciences 179, 1362–1376 (2009)
12. Denœux, T.: Maximum likelihood from evidential data: an extension of the EM algorithm. In: Borgelt, C., et al. (eds.) Combining Soft Computing and Statistical Methods in Data Analysis (Proceedings of SMPS 2010), Oviedeo, Spain. AISC, pp. 181–188. Springer (2010)
13. Denœux, T.: Maximum likelihood estimation from fuzzy data using the fuzzy EM algorithm. Fuzzy Sets and Systems 18(1), 72–91 (2011)
14. Denœux, T.: Maximum likelihood estimation from uncertain data in the belief function framework. IEEE Transactions on Knowledge and Data Engineering (to appear, 2012), doi:10.1109/TKDE.2011.201
15. Denœux, T., Masson, M.-H.: EVCLUS: Evidential clustering of proximity data. IEEE Trans. on Systems, Man and Cybernetics B 34(1), 95–109 (2004)
16. Dubois, D.: Possibility theory and statistical reasoning. Computational Statistics and Data Analysis 51(1), 47–69 (2006)
17. Dubois, D., Denœux, T.: Statistical inference with belief functions and possibility measures: a discussion of basic assumptions. In: Borgelt, C., et al. (eds.) Combining Soft Computing and Statistical Methods in Data Analysis (Proceedings of SMPS 2010), Oviedo, Spain. AISC, pp. 217–225. Springer (2010)
18. Dubois, D., Prade, H.: A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. International Journal of General Systems 12(3), 193–226 (1986)
19. Dubois, D., Prade, H.: Possibility Theory: An approach to computerized processing of uncertainty. Plenum Press, New York (1988)
20. Edwards, A.W.F.: Likelihood (expanded edition). The John Hopkins University Press, Baltimore (1992)
21. Ferraro, M.B., Coppi, R., González Rodríguez, G., Colubi, A.: A linear regression model for imprecise response. International Journal of Approximate Reasoning 51(7), 759–770 (2010)
22. Heckman, J.J.: Econometrics and empirical economics. Journal of Econometrics 100, 3–5 (2001)
23. Hudson, D.J.: Interval estimation from the likelihood function. J. R. Statistical Society B 33(2), 256–262 (1973)

24. Nguyen, H.T.: An Introduction to Random Sets. Chapman and Hall/CRC Press, Boca Raton, Florida (2006)
25. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
26. Shafer, G.: Allocations of probability. Annals of Probability 7(5), 827–839 (1979)
27. Shafer, G.: Constructive probability. Synthese 48(1), 1–60 (1981)
28. Shafer, G.: Belief functions and parametric models (with discussion). J. Roy. Statist. Soc. Ser. B 44, 322–352 (1982)
29. Smets, P.: Resolving misunderstandings about belief functions. International Journal of Approximate Reasoning 6, 321–344 (1990)
30. Smets, P.: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning 9, 1–35 (1993)
31. Smets, P.: The Transferable Belief Model for quantified belief representation. In: Gabbay, D.M., Smets, P. (eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, pp. 267–301. Kluwer Academic Publishers, Dordrecht (1998)
32. Smets, P.: Belief functions on real numbers. International Journal of Approximate Reasoning 40(3), 181–223 (2005)
33. Smets, P., Kennes, R.: The Transferable Belief Model. Artificial Intelligence 66, 191–243 (1994)
34. Sprott, D.A.: Statistical Inference in Science. Springer, Berlin (2000)
35. Walley, P.: Belief function representations of statistical evidence. The Annals of Statistics 15(4), 1439–1465 (1987)
36. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
37. Wasserman, L.A.: Belief functions and statistical evidence. The Canadian Journal of Statistics 18(3), 183–196 (1990)
38. Yager, R.R.: The entailment principle for Dempster-Shafer granules. Int. J. of Intelligent Systems 1, 247–262 (1986)
39. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)

Brief Introduction to Probabilistic Compositional Models

Radim Jiroušek

Abstract. Any field of social sciences is based on uncertain knowledge, uncertain information and uncertain data. The economics is not an exception. This is why probability theory and probabilistic modeling play an important role in econometrics. In practical applications one has to cope with the fact that even relatively small models have to take into account rather hundreds than tens of factors. This is why the methods for multidimensional probability distribution representation, like Bayesian networks, have become so popular in this field. The goal of this paper is to promote an alternative approach, so called compositional models.

1 Introduction

There are more and more fields of human activities which are giving rise to databases of enormous size. In some of them, the research data bases are a side product of other business activities, like, for example, in banking where even small banks store hundreds or rather thousands of records describing their clients' activities every day. As another example we can consider the research in the field of customer relationship management, which is based on the analysis of records describing the customer spending. On the other hand, creating large data bases has become a business of its own, as the different media research companies attest to. These companies collect data on all possible marketing activities, like data from TV-meters, or data monitoring advertising investments, such as data on advertising in journals and on the Internet. An existence of such institutions proves the fact that data have become a business product and that their analysis and processing is an important part of business life.

Radim Jiroušek
Faculty of management, University of Economics,
Prague, Jarošovská 1117/II,
377 01 Jindřichův Hradec, Czech Republic
e-mail: radim@utia.cas.cz

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 49–60.
DOI: 10.1007/978-3-642-35443-4_4

So, it is not surprising that there is abundant literature on application techniques such as Bayesian networks [3, 14], which is perhaps the most popular tool to describe and process multidimensional probability distributions. Here we are expected to present some examples of research papers describing typical applications of Bayesian networks, but we do not dare to do it; in a few seconds Google has found more than two million incidences of 'application of Bayesian network to ...'. In this paper we do not intend to present the *two-millionth-first* paper on the Bayesian networks. On the contrary, we want to present a survey paper (summarizing some of the results published in [5, 9, 8]) on an alternative approach to multidimensional probability distribution representation and processing, an approach based on the so-called *operator of composition*.

In contrast to Bayesian networks, an advantage of the models described in the current paper, which we call *compositional models*, is that we can make do with probability theory. Though they are as powerful as Bayesian networks (they can model the same class of distributions), they do not use graphs to represent the distribution structure. For other advantages of compositional models see Conclusions.

2 Notation and Basic Concepts

We consider variables $u \in N$, each having a finite (non-empty) set of values that will be denoted by $\mathbb{X}_u$. The set of all combinations of the considered values will be denoted $\mathbb{X}_N = \times_{u \in N} \mathbb{X}_u$. Analogously, for a subset of variables $K \subset N$, $\mathbb{X}_K = \times_{u \in K} \mathbb{X}_u$.

Distributions of the considered variables will be denoted by Greek letters $\kappa, \lambda, \ldots$ with possible indices; thus for $K \subseteq N$, we can consider a distribution $\kappa(K)$, which is a $|K|$-dimensional distribution and $\kappa(x)$ denotes the value of probability distribution κ for point $x \in \mathbb{X}_K$.

For a probability distribution $\kappa(K)$ and $J \subset K$, we will often consider a *marginal distribution* $\kappa^{\downarrow J}$ of κ, which can be computed for all $x \in \mathbb{X}_J$ by

$$\kappa^{\downarrow J}(x) = \sum_{y \in \mathbb{X}_K : y^{\downarrow J} = x} \kappa(y),$$

where $y^{\downarrow J}$ denotes the *projection* of $y \in \mathbb{X}_K$ into $\mathbb{X}_J$. Note that we do not exclude situations when $J = \emptyset$. By definition, we get $\kappa^{\downarrow \emptyset} = 1$.

Having two distributions $\pi(K)$ and $\kappa(K)$, we say that κ dominates π (in symbol $\pi \ll \kappa$) if for all $x \in \mathbb{X}_K$, for which $\kappa(x) = 0$ also $\pi(x) = 0$. As a measure of similarity of these two distributions we will consider their *Kullback-Leibler divergence* [13] (or crossentropy) defined[1]

$$Div(\pi; \kappa) = \sum_{x \in \mathbb{X}_K} \pi(x) \log \frac{\pi(x)}{\kappa(x)},$$

which is known to be zero if and only if $\pi = \kappa$.

[1] In this paper we take $\frac{0.0}{0} = 0$ by definition.

The most important notion of this paper is the operator of composition, which realizes an operation in a way inverse to marginalization. For a probability distribution $\kappa(K)$ and $J \subset K$, the respective marginal distribution $\kappa^{\downarrow J}$ is unique. For a distribution $\pi(J)$ there are (infinitely) many distributions $\nu(K)$ such that $\nu^{\downarrow J} = \pi$. All these distributions ν are *extensions of π for variables K*. But if we want to find that $\nu(K)$, which is as similar as possible to a given distribution $\mu(K)$, we can take the distribution

$$\nu = \arg \min_{\lambda(K):\lambda^{\downarrow J}=\pi} Div(\lambda;\mu),$$

which is unique if the divergence is defined. In this case we say that ν is a *projection* of μ into the set (space) of all the extensions of π for variables K.

The operator of composition is designed in the way that the projection of μ into the set of all the extension of π is got as a composition of π and μ - see Property 3 of the following Proposition.

Definition 1. *For two arbitrary distributions $\kappa(K)$ and $\lambda(L)$, for which $\kappa^{\downarrow K\cap L} \ll \lambda^{\downarrow K\cap L}$, their* composition *is, for each $x \in \mathbb{X}_{L\cup K}$, given by the following formula*

$$(\kappa \triangleright \lambda)(x) = \frac{\kappa(x^{\downarrow K})\lambda(x^{\downarrow L})}{\lambda^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.$$

In case $\kappa^{\downarrow K\cap L} \not\ll \lambda^{\downarrow K\cap L}$, the composition remains undefined.

Let us summarize the most important properties of the composition operator that were proved in [5, 9]

Proposition 1. *Suppose $\kappa(K)$ and $\lambda(L)$ are probability distributions for which $\lambda^{\downarrow K\cap L} \gg \kappa^{\downarrow K\cap L}$. Then the following statements hold:*

1. Domain: $\kappa \triangleright \lambda$ *is a distribution for $K \cup L$.*
2. Composition preserves first marginal: $(\kappa \triangleright \lambda)^{\downarrow K} = \kappa$.
3. Projection: $\kappa \triangleright \lambda = \arg \min_{\nu(K\cup L):\nu^{\downarrow K}=\kappa} Div(\nu^{\downarrow L};\lambda)$.
4. Non-commutativity: *In general, $\kappa \triangleright \lambda \neq \lambda \triangleright \kappa$.*
5. Commutativity under consistency: *If $\kappa^{\downarrow K\cap L} = \lambda^{\downarrow K\cap L}$, then $\kappa \triangleright \lambda = \lambda \triangleright \kappa$.*
6. Non-associativity: *Suppose $\mu(M)$ is a probability distribution, then, in general, $(\kappa \triangleright \lambda) \triangleright \mu \neq \kappa \triangleright (\lambda \triangleright \mu)$.*
7. Associativity under a special condition: *Suppose $\mu(M)$ is a probability distribution, and suppose $L \supset (K \cap M)$. Then, $(\kappa \triangleright \lambda) \triangleright \mu = \kappa \triangleright (\lambda \triangleright \mu)$, if the right hand side formula is defined.*
8. Stepwise composition: *Suppose M is such that $(K \cap L) \subseteq M \subseteq L$. Then $(\kappa \triangleright \lambda^{\downarrow M}) \triangleright \lambda = \kappa \triangleright \lambda$.*
9. Simple marginalization: *Suppose M is such that $(K \cap L) \subseteq M \subseteq K \cup L$. Then $(\kappa \triangleright \lambda)^{\downarrow M} = \kappa^{\downarrow K\cap M} \triangleright \lambda^{\downarrow K\cap M}$.*
10. Maximum entropy extension: *If $\kappa^{\downarrow K\cap L} = \lambda^{\downarrow K\cap L}$, then $\kappa \triangleright \lambda = \arg \max_{\nu \in \Pi(\kappa,\lambda)} \mathbf{H}(\nu)$, where $\Pi(\kappa,\lambda)$ is the set of all common extensions of κ and λ, and $\mathbf{H}(\nu)$ is a Shannon entropy of ν.*

3 Compositional Models

To avoid some technical problems and the necessity of repeating some assumptions to excess, let us make three conventions.

In this and the next section we will consider a system of n distributions $\kappa_1(K_1)$, $\kappa_2(K_2), \ldots, \kappa_n(K_n)$. Therefore, whenever we speak about a distribution κ_k, if not explicitly specified otherwise, the distribution κ_k will always be assumed to be a distribution of variables K_k. Thus, for example, $\kappa_2 \triangleright \kappa_1 \triangleright \kappa_4$, if it is defined, will determine the distribution of variables $K_1 \cup K_2 \cup K_4$.

Our second convention pertains to the fact that the operator of composition is neither commutative nor associative. To avoid having to write too many parentheses in the formulas, in the rest of the paper we will apply the operators from left to right. Thus

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 \triangleright \ldots \triangleright \kappa_n = (\ldots((\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3) \triangleright \ldots \triangleright \kappa_n),$$

and the parentheses will be used only when we want to change this default ordering. Therefore, to construct a multidimensional distribution it is sufficient to determine a sequence – we call it a *generating sequence* – of oligodimensional distributions.

The third convention is of a rather technical nature. Since in the remaining part of the paper we are interested in a construction of multidimensional models, it is quite natural that we will always assume that all the models (compositions) we speak about are defined.

3.1 Perfect Sequences

Definition 2. *A generating sequence of probability distributions $\kappa_1, \kappa_2, \ldots, \kappa_n$ is called* perfect *if all the distributions from this sequence are marginals of the distribution $(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)$, i.e., if for all $i = 1, 2, \ldots, n$*

$$(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{\downarrow K_i} = \kappa_i.$$

Notice that when defining a perfect sequence, let alone a generating sequence, we have not imposed any conditions on sets of variables for which the distributions were defined. For example, considering a generating sequence where one distribution is defined for a subset of variables of another distribution (i.e., $K_j \subset K_k$) is fully sensible and may provide some information about the resulting multidimensional distribution. If, e.g., $\kappa(u), \lambda(v), \mu(u,v,w)$ is a perfect sequence, it is quite obvious that

$$\kappa(u) \triangleright \lambda(v) \triangleright \mu(u,v,w) = \mu(u,v,w)$$

(because all the elements of a perfect sequence are marginals of the resulting distribution and therefore μ must be marginal to $\kappa \triangleright \lambda \triangleright \mu$). Nevertheless, it can happen that for some reason or another, it may be more advantageous to work with the model defined by the perfect sequence than just with the distribution μ. From this model one can immediately see that variables u and v are independent, which, not knowing the numbers defining the distribution, one cannot say about distribution μ.

Let us present two important properties on perfect sequences (Theorem 10.14 and Theorem 10.15 in [9]).

Proposition 2. *If a sequence of distributions $\kappa_1, \kappa_2, \ldots, \kappa_n$ is perfect, then*

$$\mathbf{H}(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n) \geq \mathbf{H}(\nu)$$

for any $\nu \in \{\pi(K_1 \cup K_2 \cup \ldots \cup K_n) : \pi^{\downarrow K_i} = \kappa_i \;\; \forall i = 1, 2, \ldots, n\}$.

Proposition 3. *If a sequence of distributions $\kappa_1, \ldots, \kappa_n$ and its permutation $\kappa_{i_1}, \ldots,$ κ_{i_n} are both perfect, then $\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n = \kappa_{i_1} \triangleright \kappa_{i_2} \triangleright \ldots \triangleright \kappa_{i_n}$.*

From the point of view of practical applications it is important to know that each generating sequence can be transformed into a perfect sequence. The process of transformation is described in the following assertion proved in [9] (Theorem 10.9).

Proposition 4. *For any generating sequence $\kappa_1, \kappa_2, \ldots, \kappa_n$, the sequence $\pi_1, \pi_2, \ldots,$ π_n computed by the following process*

$$\begin{aligned}
\pi_1 &= \kappa_1, \\
\pi_2 &= \pi_1^{\downarrow K_2 \cap K_1} \triangleright \kappa_2, \\
\pi_3 &= (\pi_1 \triangleright \pi_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \triangleright \kappa_3, \\
&\vdots \\
\pi_n &= (\pi_1 \triangleright \ldots \triangleright \pi_{n-1})^{\downarrow K_n \cap (K_1 \cup \ldots K_{n-1})} \triangleright \kappa_n
\end{aligned}$$

is perfect and $\kappa_1 \triangleright \ldots \triangleright \kappa_n = \pi_1 \triangleright \ldots \triangleright \pi_n$.

From the theoretical point of view, this process is simple. Unfortunately, it need not be valid from the point of view of computational complexity. The process requires marginalization of models, which are distributions represented by generating sequences, and this may be computationally very expensive [6]. To avoid these computational problems we will use decomposable generating sequences introduced in the following paragraph.

3.2 Decomposable Sequences

We call a generating sequence $\kappa_1, \kappa_2, \ldots, \kappa_n$ *decomposable* if the corresponding sequence of variable sets $K_1, K_2, \ldots, K_n$ meets the *running intersection property* (RIP), i.e., if

$$\forall i = 2, \ldots, n \;\; \exists j (1 \leq j < i) \quad \left(K_i \cap (\bigcup_{k=1}^{i-1} K_k) \subseteq K_j \right).$$

The importance of these sequences follows, among others, from the following assertion [9].

Proposition 5. *If $\kappa_1, \kappa_2, \ldots, \kappa_n$ is a sequence of pairwise consistent probability distributions such that $K_1, \ldots, K_n$ meets RIP, then this sequence is perfect.*

The reader can notice, that if the sequence $K_1, K_2, \ldots, K_n$ in Proposition 4 meets RIP, then $K_3 \cap (K_1 \cup K_2)$ equals either $K_3 \cap K_1$ or $K_3 \cap K_2$. Similarly, $K_4 \cap (K_1 \cup K_2 \cup K_3)$ equals $K_4 \cap K_j$ for some $j \leq 3$. It means that, thanks to RIP, for all $i = 3, 4, \ldots, n$ the necessary marginal distributions

$$(\pi_1 \triangleright \ldots \triangleright \pi_{i-1})^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}$$

can be computed from some π_j as $\pi_j^{\downarrow K_i \cap K_j}$, because $\pi_1, \ldots, \pi_{i-1}$ is a perfect sequence and therefore π_j is marginal to $\pi_1 \triangleright \ldots \triangleright \pi_{i-1}$. All this means that for this type of distributions the process of perfectization can be performed locally.

4 Conditioning

In this short section we will show that the operator of composition can also serve as a tool for computation of conditional distributions. Define a *degenerated* one-dimensional probability distribution $\pi_{|u;\alpha}$ as a distribution of variable u achieving probability 1 for value $u = \alpha$, i.e.,

$$\pi_{|u;\alpha}(x) = \begin{cases} 1 & \text{if } x = \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

Now, consider a probability distribution $\kappa(K)$ for which $\{u, v\} \subset K$ and compute $(\pi_{|u;\alpha} \triangleright \kappa)^{\downarrow \{v\}}$. For any $y \in \mathbb{X}_v$

$$\begin{aligned} (\pi_{|u;\alpha} \triangleright \kappa)^{\downarrow \{v\}}(y) &= ((\pi_{|u;\alpha} \triangleright \kappa)^{\downarrow \{u,v\}})^{\downarrow \{v\}}(y) \\ &= (\pi_{|u;\alpha} \triangleright \kappa^{\downarrow \{u,v\}})^{\downarrow \{v\}}(y) \\ &= \sum_{x \in \mathbf{X}_u} \frac{\pi_{|u;\alpha}(x) \cdot \kappa^{\downarrow \{u,v\}}(x, y)}{\kappa^{\downarrow \{u\}}(x)} \\ &= \frac{\kappa^{\downarrow \{u,v\}}(\alpha, y)}{\kappa^{\downarrow \{u\}}(\alpha)} = \kappa(v = y | u = \alpha). \end{aligned}$$

Thus we have got that $\kappa(v|u = \alpha) = (\pi_{|u;\alpha} \triangleright \kappa)^{\downarrow \{v\}}$.

In the same way it can be shown for any $L \subseteq K \setminus \{u\}$ that $\left(\pi_{|u;\alpha} \triangleright \kappa\right)^{\downarrow L}$ is an $|L|$-dimensional conditional distribution κ under the condition that variable u attains value α, i.e., $\kappa(L|u = \alpha)$. Proceeding analogously even further we can get that for any $v \in K \setminus (L \cup \{u\})$ and $\beta \in \mathbb{X}_v$

$$\kappa(L|u = \alpha, v = \beta) = \left(\pi_{|v;\beta} \triangleright \left(\pi_{|u;\alpha} \triangleright \kappa\right)\right)^{\downarrow L}$$

is a conditional distribution for variables from L given that variables u and v attain values α and β, respectively.

5 Local Computations

By local computations we understand a process based on the ideas published in the famous paper by Lauritzen and Spiegelhalter [15]. Here we have especially in mind the idea that when computing the required conditional probability, one performs computations only on the system of marginal distributions defining the decomposable model. It means that during the computational process one does not need to store more data than what is necessary to store for the decomposable model.

In the preceding paragraph we showed that the conditional distribution can be expressed as a composition of a degenerated distribution with the distribution for which we want to compute the conditional distribution. So, let us assume that a distribution κ is decomposable, i.e.,

$$\kappa = \kappa^{\downarrow K_1} \triangleright \kappa^{\downarrow K_2} \triangleright \ldots \triangleright \kappa^{\downarrow K_n}$$

for a sequence $K_1, K_2, \ldots, K_n$ meeting RIP, and we want to compute, say, $\kappa(L|u = \alpha, v = \beta) = \left(\pi_{|v;\beta} \triangleright \left(\pi_{|u;\alpha} \triangleright \kappa\right)\right)^{\downarrow L}$.

For this, we will have to take advantage of the famous fact (an immediate consequence of the existence of a join tree, see [1]) that if $K_1, K_2, \ldots, K_n$ can be ordered to meet RIP, then there are many of such orderings, and for each $k \in \{1, 2, \ldots, n\}$, at least one of them starts with K_k. Therefore, thanks to Proposition 3, we can consider any of these orderings. So, consider any K_k for which $u \in K_k$, and find the ordering meeting RIP which starts with this K_k. Without loss of generality let it be $K_1, K_2, \ldots, K_n$ (so, $u \in K_1$).

Thus, our goal is to compute in the first step $\left(\pi_{|u;\alpha} \triangleright \kappa\right)$

$$\pi_{|u;\alpha} \triangleright \kappa = \pi_{|u;\alpha} \triangleright \left(\kappa^{\downarrow K_1} \triangleright \kappa^{\downarrow K_2} \triangleright \ldots \triangleright \kappa^{\downarrow K_n}\right).$$

Now applying $(n-1)$ times *Associativity under a special condition* (Property 7 of Proposition 1) we get (recall that we selected the RIP ordering, for which $u \in K_1$)

$$\begin{aligned} \pi_{|u;\alpha} \triangleright \left(\kappa^{\downarrow K_1} \triangleright \kappa^{\downarrow K_2} \triangleright \ldots \triangleright \kappa^{\downarrow K_n}\right) &= \\ \pi_{|u;\alpha} \triangleright \left(\kappa^{\downarrow K_1} \triangleright \kappa^{\downarrow K_2} \triangleright \ldots \triangleright \kappa^{\downarrow K_{n-1}}\right) \triangleright \kappa^{\downarrow K_n} &= \\ \ldots = \pi_{|u;\alpha} \triangleright \kappa^{\downarrow K_1} \triangleright \kappa^{\downarrow K_2} \triangleright \ldots \triangleright \kappa^{\downarrow K_n}&, \end{aligned}$$

from which the following computationally local process (see Proposition 4 and the comment in Section 3.2)

$$\begin{aligned} \nu_1 &= \pi_{|u;\alpha} \triangleright \kappa^{\downarrow K_1}, \\ \nu_2 &= \nu_1^{\downarrow K_2 \cap K_1} \triangleright \kappa^{\downarrow K_2}, \\ \nu_3 &= \left(\nu_1 \triangleright \nu_2\right)^{\downarrow K_3 \cap (K_1 \cup K_2)} \triangleright \kappa^{\downarrow K_3}, \\ &\vdots \\ \nu_n &= \left(\nu_1 \triangleright \ldots \triangleright \nu_{n-1}\right)^{\downarrow K_r \cap (K_1 \cup \ldots K_{n-1})} \triangleright \kappa^{\downarrow K_n}, \end{aligned}$$

yields a perfect decomposable sequence $\nu_1, \ldots, \nu_n$, such that $\pi_{|u;\alpha} \triangleright \kappa = \nu_1 \triangleright \ldots \triangleright \nu_n$.

Now, it has remained to compute in the second step the required

$$\kappa(L|u=\alpha,v=\beta)=\left(\pi_{|v;\beta}\triangleright\left(\pi_{|u;\alpha}\triangleright\kappa\right)\right)^{\downarrow L}=\left(\pi_{|v;\beta}\triangleright(\nu_1\triangleright\ldots\triangleright\nu_n)\right)^{\downarrow L}.$$

Thanks to decomposability of the sequence $\nu_1,\ldots,\nu_n$ the computations will proceed in the same way as in the first step. First, distributions ν_i will be reordered in the way that $\nu_{j_1},\ldots,\nu_{j_n}$ meet RIP and variable v is among the variables for which ν_{j_1} is defined. Then we can, as in the first step, due to *Associativity under a special condition* deduce that

$$\pi_{|v;\beta}\triangleright(\nu_{j_1}\triangleright\nu_{j_2}\triangleright\ldots\triangleright\nu_{j_n})=(\pi_{|v;\beta}\triangleright\nu_{j_1})\triangleright\nu_{j_2}\triangleright\ldots\triangleright\nu_{j_n},$$

which can be, again, converted into a perfect sequence by the computationally local process of perfectization.

6 Heuristics for Model Construction

The reader interested in other theoretical issues concerning the operator of composition and perfect sequence models is referred to [9] and the papers cited there. Here, we want to briefly introduce a possible heuristic way to create a perfect sequence model from a data file – see Figure 1. For a more detailed description of this process, as well as for an example of its application to a small data file, the reader is referred to [8]. Notice that the described process is fully driven by an expert, and thus the following decisions must be made by a human expert:

1. Selection of oligodimensional distributions at the beginning of the whole process.
2. Decision which type of "refinement" procedure should be chosen (detailed explanation is given below).
3. Stopping rule.

As it can be seen from the diagram in Figure 1, the process is initiated with definition of a system of oligodimensional distributions. Regarding the fact that the process cyclically employs steps of *verification* and *refinement*, during which this initial system is gradually changed, the result is fairly independent of the initial selection. For example, starting with all two-dimensional distributions may be quite reasonable (for application to small data files with a limited number of variables one can consider a possibility to start with three-dimensional marginal distributions). In other situations, an expert can select the initial marginal distributions from which the model should be constructed. Generally, we propose to select distributions carrying a greater amount of information. This idea is supported by the following assertion, proved in [7] (Corollary 1.). It claims that the higher information content of a perfect sequence, the better approximation of the unknown distribution.

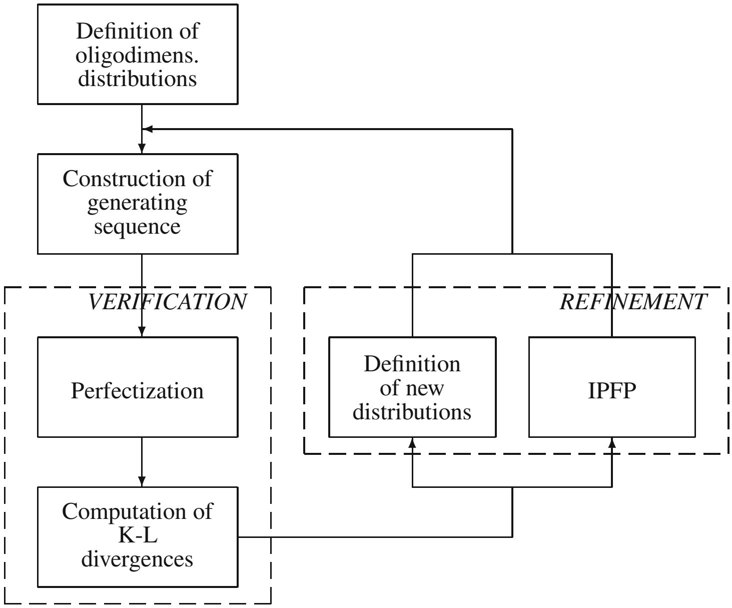


Fig. 1 Process of model construction

Proposition 6. *Consider an arbitrary distribution* κ, *and a generating sequence consisting of its marginals* $\kappa^{\downarrow K_1}, \kappa^{\downarrow K_2}, \ldots, \kappa^{\downarrow K_n}$. *If this generating sequence is perfect, then*

$$Div(\kappa \| \kappa^{\downarrow K_1} \triangleright \ldots \triangleright \kappa^{\downarrow K_n}) = I(\kappa) - I(\kappa^{\downarrow K_1} \triangleright \ldots \triangleright \kappa^{\downarrow K_n}),$$

where the Information content $I(\pi)$ *of a distribution* $\pi(J)$ *is the Kullback-Leibler divergence of* π *and a product distribution of its one-dimensional marginal distributions:*

$$I(\pi) = Div(\pi \| \prod_{u \in J} \pi^{\downarrow \{u\}}) = \sum_{x \in \mathbb{X}_J} \pi(x) \log \frac{\pi(x)}{\prod_{u \in J} \pi^{\downarrow \{u\}}(x^{\downarrow \{u\}})}.$$

Let us stress that the information content is a generalization of a *Shannon mutual information*, which will be used in the algorithm further in this text, and which is for two disjoint (nonempty) $L, M \subset J$ defined by the formula

$$MI_\pi(K;L) = \sum_{x \in \mathbb{X}_K} \sum_{y \in \mathbb{X}_L} \pi^{\downarrow K \cup L}(x,y) \log \frac{\pi^{\downarrow K \cup L}(x,y)}{\pi^{\downarrow K}(x) \cdot \pi^{\downarrow L}(y)}.$$

If we want to construct a perfect sequence model approximating an unknown distribution κ, we have to aim at getting the model with the highest possible information content (under the assumption that the oligodimensional distributions, which the perfect sequence consists of, are marginals of the approximated distribution). In [8] we have published the following heuristic algorithm producing a sub-optimal generating sequence from a system of oligodimensional distributions.

Algorithm

Input: System of low-dimensional distributions $\kappa_1(K_1),\ldots \kappa_n(K_n)$.

Initialization: Select a variable u and a distribution κ_j such that $u \in K_j$. Set $\pi_1 := \kappa_j^{\downarrow\{u\}}$, $L := \{u\}$ and $k := 1$.

Computational Cycle: While $K_1 \cup \ldots \cup K_n \setminus L \neq \emptyset$ perform the following 3 steps:

1. for all $j = 1,\ldots,n$ and all $w \in K_j \setminus L$ compute the mutual information
$$MI_{\kappa_j}(w; K_j \cap L).$$
2. Fix j and w for which $MI_{\kappa_j}(w; K_j \cap L)$ achieved its maximal value.
3. Increase k by 1. Set $\pi_k := \kappa_j^{\downarrow(K_j \cap L)\cup\{w\}}$ and $L := L \cup \{m\}$.

Output: Generating sequence $\pi_1, \pi_2, \ldots, \pi_k$.

What can be said about the resulting generating sequence $\pi_1, \pi_2, \ldots, \pi_k$? Distribution $\nu = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_k$ is a probability distribution of variables $K_1 \cup K_2 \cup \ldots \cup K_n$. The algorithm realizes a greedy (therefore very efficient) process, which seeks to find a sequence utilizing the information content of individual oligodimensional distributions in a maximal possible way. The result is a generating sequence which, unfortunately, need not be perfect. It means that some of the input distributions are not marginals of the resulting multidimensional model. As a rule, the expert (the model constructor) has to accept some deviations of the model marginals from the input oligodimensional distributions. To decide whether the obtained deviations are acceptable, i.e., whether the whole model construction process depicted in Figure 1 should be terminated, the expert must be provided with some additional information. To get it, the process employs the perfectization procedure described in Proposition 4. Then it is possible to compare original oligodimensional distributions with the corresponding marginals defined by the model. The comparison may be done with the help of Kullback-Leibler divergence; as already said above, its value equals 0 *iff* $\pi = \kappa$, otherwise it is always positive. Therefore, the lower this value, the closer κ to π. The goal of this step is to find all the marginal distributions which are unacceptably distorted by the model. If there is no such a marginal distribution, the process is terminated. In the opposite case, the expert proposes to perform another cycle of the whole process with a modified system of oligodimensional distributions. The described process then proceeds so that several original distributions are substituted with one *a-little-bit-more*-dimensional one in the *refinement* step.

As the reader can see from Figure 1, there are two possibilities to get these new distributions. If it is possible (i.e., the data file is large enough) the expert can decide to get them as estimations from the given data file (going along the left branch of the *refinement* box in Figure 1). However, if the data file is too small to get reliable estimations (which may happen easily if one needs to substitute several distributions with a distribution whose dimensionality is high – let us say, 6 or more), then one can take advantage of the well-known Iterative Proportional Fitting Procedure (IPFP) (see [2]; for its effective implementation, which makes it possible to compute distributions of pretty high dimensions, see [4]). In this way, when all the desired substitutions are realized, a new system of oligodimensional distributions is set up, to which the heuristic algorithm for generating sequence construction is again applied. The described cycle is repeated until the expert decides that a suitable multidimensional model representing (approximating) all the required oligodimensional distributions has been achieved.

Let us stress once more that the process shown in Figure 1 is fully controlled by the expert. The more cycles of the process are performed, the higher dimensions of the input distributions are considered. If the expert had continued ad absurdum, the process would have, in fact, finished with an application of IPFP to all of the initial oligodimensional distributions (which is, as a rule, computationally intractable in practical situations).

7 Conclusions

In this paper we summarized most of the practically oriented properties of compositional models and showed that they can be applied to multidimensional distribution representation. We also showed that conditional distributions can be computed as a composition of one or several degenerated distributions with the respective model, and that these computations can be, for decomposable models, performed locally.

Let us, now, mention another advantage of perfect compositional models that is important for another computational process that was not discussed in this paper. We have in mind the process of marginalization. Since the perfect model is composed of a system of its marginal distributions, it is not difficult to show on examples that there are number of situations when marginalization in a compositional model is simple but the same process in the corresponding Bayesian network is either computationally very expensive or even intractable. This advantageous property of compositional models is employed in algorithms described in [6].

As the last remark, let us mention that compositional models where introduced not only within the framework of probability theory, but also in possibility theory [16], the theory of belief functions [12], and recently also for the Shenoy's valuation-Based Systems [11]. Thus, most of the results presented in this paper can easily be extended into the above mentioned theoretical frameworks. For example, the content of Sections 4 and 5 have originally been published for belief functions [10], not for probability theory.

Acknowledgement. This work was supported in part by the National Science Foundation of the Czech Republic by grant no. GACR 403/12/2175.

References

1. Beeri, C., Fagin, R., Maier, D., Yannakakis, M.: On the Desirability of Acyclic Database Schemes. J. ACM 30(3), 479–513 (1983)
2. Deming, W.E., Stephan, F.F.: On a least square adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Stat. 11, 427–444 (1940)
3. Jensen, F.V.: Bayesian Networks and Decision Graphs. IEEE Computer Society Press, New York (2001)
4. Jiroušek, R.: Solution of the Marginal Problem and Decomposable Distributions. Kybernetika 27(5), 403–412 (1991)
5. Jiroušek, R.: Composition of probability measures on finite spaces. In: Geiger, D., Shenoy, P.P. (eds.) Uncertainty in Artificial Intelligence: Proceedings of the 13th Conference (UAI 1997), pp. 274–281. Morgan Kaufmann (1997)
6. Jiroušek, R.: Marginalization in Composed Probabilistic Models. In: Boutilier, C., Goldszmidt, M. (eds.) Uncertainty in Artificial Intelligence, pp. 301–308. Morgan Kaufmann Publishers, San Francisco (2000)
7. Jiroušek, R.: On approximating multidimensional probability distributions by compositional models. In: Bernard, J.-M., Seidenfeld, T., Zaffalon, M. (eds.) Proceedings of the 3rd International Symposium on Imprecise Probabilities and Their Application, pp. 305–320. Carleton Scientific (2003)
8. Jiroušek, R.: Data-Based Construction of Multidimensional Probabilistic Models with MUDIM. Logic Journal of the IGPL 14(3), 501–520 (2006)
9. Jiroušek, R.: Foundations of compositional model theory. International Journal of General Systems 40(6), 623–678 (2011)
10. Jiroušek, R.: Local Computations in Dempster-Shafer Theory of Evidence. Int. J. Approx. Reason. 53(8), 1155–1167 (2012)
11. Jiroušek, R., Shenoy, P.P.: Compositional models in valuation-based systems. Working Paper No. 325, School of Business, University of Kansas, Lawrence, KS (2011)
12. Jiroušek, R., Vejnarová, J., Daniel, M.: Compositional models of belief functions. In: de Cooman, G., Vejnarová, J., Zaffalon, M. (eds.) Proc. of the Fifth Int. Symposium on Imprecise Probability: Theories and Applications, Praha, pp. 243–252 (2007)
13. Kullback, S., Leibler, R.A.: On information and sufficiency. Annals of Mathematical Statistics 22, 76–86 (1951)
14. Lauritzen, S.L.: Graphical models. Oxford University Press (1996)
15. Lauritzen, S.L., Spiegelhalter, D.: Local computation with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society Series B 50, 157–224 (1988)
16. Vejnarová, J.: Composition of possibility measures on finite spaces: preliminary results. In: Bouchon-Meunier, B., Yager, R.R. (eds.) Proceedings of 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU 1998, pp. 25–30. Editions E.D.K., Paris (1998)

Some Aspects of Information Theory in Gambling and Economics

Hai Q. Dinh

Abstract. We discuss applications of information theory to the fields of gambling and economics, such as the problem of gambling on horse races with causal side information, and process of portfolio selection in the stock market. One of the center points is the gambling strategy proposed by Kelly, that, on the one hand, gave a real-life situation of a communication channel without optimum coding in which the rate of transmission is significant. On the other hand, its optimization process opened the door for the theory of rebalanced portfolios with known underlying distributions. We also overview the work on universal portfolios with and without side information, which yield portfolio strategies that have the same exponential rate of growths as the ones achieved by the best state-constant and constant rebalanced portfolios chosen after the stock outcomes are revealed. We do not intend to be encyclopedic, the topics included are bounded to reflect our own research interest.

1 Information Theory

Information theory was founded in 1948 by Claude E. Shannon[1] in his landmark paper "A Mathematical Theory of Communication" [26]. The main object of classical

Hai Q. Dinh
Department of Mathematical Sciences, Kent State University,
4314 Mahoning Avenue, Warren, Ohio 44483, USA
e-mail: hdinh@kent.edu

[1] Claude Elwood Shannon (April 30, 1916 - February 24, 2001) was an American mathematician, electronic engineer, and cryptographer, who is refered to as "the father of information theory" [9]. Shannon is also generally known as the founder of both digital computer and digital circuit design theory [9, 18], when, as a 21-year-old master's student at M.I.T in 1937, he wrote a thesis establishing that electrical application of Boolean algebra could construct and resolve any logical, numerical relationship [25]. It has been claimed that this was the most important master's thesis of all time. Shannon contributed to the field of cryptanalysis during World War II and afterwards, including basic work on code breaking.

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 61–77.
DOI: 10.1007/978-3-642-35443-4_5

information theory is the engineering problem of the transmission of information over a noisy channel. The most fundamental results of this theory are Shannon's source coding theorem and Shannon's noisy-channel coding theorem. The former obtains that, on average, the number of bits needed to represent the result of an uncertain event is given by its entropy; and the later establishes that reliable communication is possible over noisy channels provided that the rate of communication is below a certain threshold, called the channel capacity. The channel capacity can be approached in practice by using appropriate encoding and decoding systems.

1.1 Coding Theory

The existence of noise in communication channels is an unavoidable fact of life. A response to this problem has been the creation of error-correcting codes. Coding Theory is the study of the properties of codes and their properties for a specific application. Codes are used for data compression, cryptography, error-correction, and more recently for network coding.

The common feature of communication channels is that the original information is sent across a noisy channel to a receiver at the other end. The channel is ”noisy” in the sense that the received message is not always the same as what was sent. The fundamental problem is to detect if there is an error, and in such case, to determine what message was sent based on the approximation that was received. An example that motivated the study of coding theory is telephone transmission. It is impossible to avoid errors that occur as messages pass through long telephone lines and are corrupted by things such as lightening and crosstalk. The transmission and reception capabilities of many modems are increased by error handling capability in hardware. Another area in which coding theory has been applied successfully is deep space communication. The message source is the satellite, the channel is the out space and hardware that sends and receives date, the receiver is the ground station on earth, and the noise are outside problems such as atmospheric conditions and thermal disturbance. Data from space missions has been coded for transmission, since it is normally impractical to retransmit. It is also important to protect communication across time from inaccuracies. Data stored in computer banks or on tapes is subject to the intrusion of gamma rays and magnetic interference. Personal computers are exposed to much battering, their hard disks are often equipped with an error correcting code called "cyclic redundancy check" (CRC)[2] designed to detect accidental changes to raw computer data.

The study of codes has grown into an important subject that intersects various scientific disciplines, such as information theory, electrical engineering, mathematics, and computer science, for the purpose of designing efficient and reliable data transmission methods. This typically involves the removal of redundancy and the detection and correction of errors in the transmitted data. There are essentially two aspects to coding theory, namely, source coding (i.e, data compression) and channel coding (i.e, error correction). These two aspects may be studied in combination.

Source coding attempts to compress the data from a source in order to transmit it more efficiently. This process can be found every day on the internet where the common Zip data compression is used to reduce the network bandwidth and make files smaller. The second aspect, channel coding, adds extra data bits to make the transmission of data more robust to disturbances present on the transmission channel. The ordinary users usually are not aware of many applications using channel coding. A typical music CD uses the Reed-Solomon code to correct damages caused by scratches and dust. In this application the transmission channel is the CD itself. Cellular phones also use coding techniques to correct for the fading and noise of high frequency radio transmission. Data modems, telephone transmissions, and NASA all employ channel coding techniques to get the bits through, for example the turbo code and LDPC codes.

Algebraic coding theory studies the subfield of coding theory where the properties of codes are expressed in algebraic terms. Algebraic coding theory is basically divided into two major types of codes, namely block codes and convolutional codes. It analyzes the following three important properties of a code: code length, total number of codewords, and the minimum distance between two codewords, using mainly the Hamming[3] distance, sometimes also other distances such as the Lee distance, Euclidean distance.

---

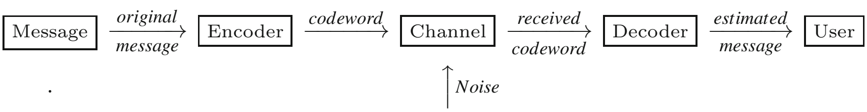
[2] A cyclic redundancy check (CRC) is an error-detecting code designed to detect accidental changes to raw computer data, and is commonly used in digidelltal networks and storage devices such as hard disk drives. The CRC was first introduced by Peterson and Brown in 1961 [24], the 32-bit polynomial used in the CRC function of Ethernet and many other standards is the work of several researchers and was published in 1975. Blocks of data entering these systems get a short check value attached, derived from the remainder of a polynomial division of their contents; on retrieval the calculation is repeated, and corrective action can be taken against presumed data corruption if the check values do not match. CRCs are so called because the check (data verification) value is a redundancy (it adds zero information to the message) and the algorithm is based on cyclic codes. CRCs are popular because they are simple to implement in binary hardware, are easy to analyze mathematically, and are particularly good at detecting common errors caused by noise in transmission channels. Because the check value has a fixed length, the function that generates it is occasionally used as a hash function.

[3] The Hamming distance is named after Richard Hamming, who first introduced it in his fundamental paper on Hamming codes in 1950 [8]. It is used in telecommunication to count the number of flipped bits in a fixed-length binary word as an estimate of error, and hence it is sometimes referred to as the *signal distance*.

Given an alphabet $\mathscr{A}$ with q symbols, a block code C of length n over the alphabet $\mathscr{A}$ is simply a subset of $\mathscr{A}^n$. The q-ary n-tuples from C are called the codewords of the code C. One normally envisions K, the number of codewords in C, as a power of q, i.e., $K = q^k$, thus replacing the parameter K with the dimension $k = \log_q K$. This dimension k is the smallest integer such that each message for C can be assigned its own individual message k-tuple from the q-ary alphabet $\mathscr{A}$. Thus, the dimension k can be considered as the number of codeword symbols that are carrying message rather than redundancy. Hence, the number $n-k$ is sometimes called the redundancy of the code C. The error correction performance of a block code is described by the minimum Hamming distance d between each pair of code words, and is normally referred as the distance of the code.

In a block code, each input message has a fixed length of $k < n$ input symbols. The redundancy added to a message by transforming it into a larger codeword enables a receiver to detect and correct errors in a transmitted code word, and to recover the original message by using a suitable decoding algorithm. The redundancy is described in terms of its information rate, or more simply, for a block code, in terms of its code rate, k/n.

At the receiver end, a decision is made about the codeword transmitted based on the information in the received n-tuple. This decision is statistical, that is, it is a best guess on the basis of available information. A good code is one where k/n, the rate of the code, is as close to one as possible (so that, without too much redundancy, information may be transmitted efficiently) while the codewords are far enough from one another that the probability of an incorrect interpretation of the received message is very small. The following diagram describes a communication channel that includes an encoding/decoding scheme:



Shannon's noisy-channel coding theorem ensures that our hopes of getting the correct messages to the users will be fulfilled a certain percentage of the time. Based on the characteristics of the communication channel, it is possible to build the right encoders and decoders so that this percentage, although not 100%, can be made as high as we desire. However, the proof of Shannon's noisy-channel coding theorem is probabilistic and only guarantees the existence of such good codes. No specific codes were constructed in the proof that provides the desired accuracy for a given channel. The main goal of Coding Theory is to establish good codes that fulfill the assertions of Shannon's noisy-channel coding theorem. During the last 50 years, while many good codes have been constructed, but only from 1993, with the introduction of turbo codes[4], the rediscoveries of LDPC codes[5], and the study of related codes and associated iterative decoding algorithms, researchers started to see codes that approach the expectation of Shannon's noisy-channel coding theorem in practice.

1.2 Information Theory without Optimum Coding

The rate of transmission over a noisy communication channel was defined by Shannon [26] in terms of various probabilities. This rate plays an important role in the Shannon's noisy-channel coding theorem that ensures that a system of encoders and decoders can be built to transmit over the channel at this rate with the probability of error made as small as we want.

In 1956, Kelly [10] considered the situations with communication systems where optimum coding is not being used, such as radar. Basically, in order to attach significance to the rate of transmission, one needs to attach a value measure to the system. That means a cost function must be constructed on pairs of symbols that can give the information on how bad it is to receive a certain symbol when another specified signal is transmitted. Moreover, the expected value of this cost function needs to have significance, i.e., a system must be preferable to another if its average cost is less. In general, this cost function should only depend on things external to the system and not on the probabilities which describe the system, so that its average value could not be identified with the rate as defined by Shannon. On the other hand, the ultimate receiver of a communication system is in a position to profit from any

[4] Turbo codes were first introduced and developed in 1993 by Berrou, Glavieux, and Thitimajshima [4]. Turbo codes are a class of high-performance forward error correction codes, which were the first practical codes to closely approach the channel capacity, a theoretical maximum for the code rate at which reliable communication is still possible given a specific noise level. Turbo codes are widely used in deep space communications and other applications where designers seek to achieve reliable information transfer over bandwidth-constrained or latency-constrained communication links in the presence of data-corrupting noise. The first class of turbo code was the parallel concatenated convolutional code. Since the introduction of the original parallel turbo codes in 1993, many other classes of turbo code have been discovered, including serial versions and repeat-accumulate codes. Iterative Turbo decoding methods have also been applied to more conventional forward error correction systems, including Reed-Solomon corrected convolutional codes.

[5] LDPC (low-density parity-check) codes were first introduced in 1963 by Robert G. Gallager in his doctoral dissertation at M.I.T [7]. At that time, it was impractical to implement and LDPC codes were forgotten, but they were rediscovered in 1996. A LDPC code is a linear error correcting code, a method of transmitting a message over a noisy transmission channel, and is constructed using a sparse bipartite graph. LDPC codes are capacity-approaching codes, which means that practical constructions exist that allow the noise threshold to be set arbitrarily close on the binary erasure channel to the Shannon limit for a symmetric memoryless channel. The noise threshold defines an upper bound for the channel noise, up to which the probability of lost information can be made as small as desired. Using iterative belief propagation techniques, LDPC codes can be decoded in time linear to their block length.

knowledge of the input symbols or even from a better estimate of their probabilities. Thus, a cost function needs to somehow reflect this feature as well.

Kelly then analized such a communication system with no optimum coding of a real-life situation without a cost function, namely, the gambling situation in which a gambler uses knowledge of the received symbols of a communication channel to make profitable bets on the transmitted symbols. He posed the question of whether a mathematical formula could be derived to ensure success in betting on horse races. His solution, known as Kelly gambling strategy, or Kelly bet, or Kelly criterion, or Kelly formula, became the basic of many gambling system. In most gambling scenarios, and several investing scenarios under some simplifying assumptions, the Kelly gambling strategy will do better than any essentially different strategy in the long run[6].

2 Information Theory and Gambling

2.1 Kelly Gambling Strategy

Kelly showed in [10] that if each horse race outcome can be represented as an independent and identically distributed copy of a random variable X and the gambler has some side information Y relevant to the outcome of the race, then under some conditions on the odds, the mutual information $I(X;Y)$ is the difference between growth rates of the optimal gambler's wealth with and without side information Y. Thus, Kelly's result gave an interpretation that mutual information $I(X;Y)$ is the value of side information Y for the horse race X.

The side information can be considered to be a set of symbols received over a communication channel that the gambler uses to make bets on the transmitted symbols. The exponential rate of growth is defined as

$$G = \lim_{N\to\infty} \frac{1}{N} \log \frac{V_N}{V_0},$$

where V_N is the gambler's capital after N bets and V_0 is the starting capital. An optimal gambler always tries to maximize the exponential rate of growth, i.e., to achieve $G_{\max}$. This is established by placing the bets in proportion to some probability distribution. Here, the channel has several input symbols, not necessarily equally likely, which represent the outcome of chance events. Denote

X : Random variable representing the transmitted symbols (outcome of the races),
Y : Random variable representing the received symbols (side information),
$p(s)$: probability that s is transmitted,
$p(r|s)$: conditional probability that r is received when s is transmitted,
$p(s,r)$: joint probability of s and r,

[6] Over the years, Kelly gambling strategy has become a part of mainstream investment theory. It has been claimed that well known successful investors including Warren Buffett and Bill Gross use Kelly methods [19, 30].

$q(r)$: received symbol probability,
$q(s|r)$: conditional probability of transmitted symbol given the received symbol,
α_s : odds paid on the occurrence of s, i.e., α_s is the number of dollars returned for a one dollar bet including that one dollar,
$a(s|r)$: fraction of capital the gambler bets on the occurrence of s after observing r.

In case there is no track take, i.e., the gambler does not pay anything to the track, and the odds are fair, i.e., $\alpha_s = \frac{1}{p(s)}$, then the exponential rate of growth is

$$G = \sum_{rs} p(s,r) \log \alpha_s a(s|r) = \sum_{rs} p(s,r) \log \frac{a(s|r)}{p(s)} = \sum_{rs} p(s,r) \log a(s|r) + H(X),$$

where $H(X)$ is the source rate as defined by Shannon. Thus, G is maximized when $a(s|r) = q(s|r)$, and

$$G_{\max} = H(X) - H(X|Y) = I(X;Y),$$

which is the rate of transmission defined by Shannon.

In the case there is no track take, and the odds are not fair, i.e., $\alpha_s \neq \frac{1}{p(s)}$, then

$$G = \sum_{rs} p(s,r) \log a(s|r) + \sum_{s} p(s) \log \alpha_s = \sum_{rs} p(s,r) \log a(s|r) + H(X),$$

where $H(X) = \sum_s p(s) \log \alpha_s$. In this case, G is also maximized when $a(s|r) = q(s|r)$, and

$$G_{\max} = H(\alpha) - H(X|Y) = I(X;Y).$$

It follows that the gambler can maximize his profit by ignoring the posted odds while placing his bets. Moreover, anything other than fair odds gives an advantage to the gambler.

Consider the case there is track take, i.e., the gambler has to pay a certain positive amount to the track. Denote by b_r the fraction not bet when the received symbol is r, i.e., $b_r = 1 - \sum_s a(s|r)$. Then

$$G = \sum_{rs} p(s,r) \log[b_r + \alpha_s a(s|r)],$$

subject to the constraints

$$b_r + \sum_s a(s|r) = 1.$$

In this case, Kelly's strategy to maximize the exponential rate of growth G is the following procedure:

(1) Permute indices so that $p(s)\alpha_s \geq p(s+1)\alpha_{s+1}$.
(2) Set

$$p_t = \sum_1^t p(s), \quad \sigma_t = \sum_1^t \frac{1}{\alpha_s}.$$

Let $t = T$ be the smallest value of t so that $\frac{1-p_t}{1-\sigma_t}$ achieves its minimum positive value, and set b equal to such value, i.e., $b = \frac{1-p_T}{1-\sigma_T}$.
(3) Set $a(s) = \max\{p(s) - \frac{b}{\alpha_s}, 0\}$.

The maximum value of G is

$$G_{\max} = \sum_1^T p(s) \log p(s)\alpha_s + (1-p_T) \log \frac{1-p_T}{1-\sigma_T}.$$

2.2 *Directed Information and Causal Conditioning*

In 1973, Marko [14] initiated the concept of *directed information*, which gave a meaningful notion of directivity to the information flow through a communication channel. The directed information $I(X^n \to Y^n)$ from a sequence X^n to a sequence Y^n was then specifically defined by Massey [17] in 1990 as

$$I(X^n \to Y^n) = \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}).$$

It can be seen readily that, in general, $I(X_n \to Y_n) \neq I(Y_n \to X_n)$. Massey [17] showed that when feedback is present, the directed information $I(X^n \to Y^n)$ gives a better upper bound on the information that the output sequence Y^n gives about the source sequence U^k than does the conventional mutual information $I(X^n; Y^n)$, i.e,

$$I(U^k; Y^n) \leq I(X^n \to Y^n) \leq I(X^n; Y^n).$$

Subsequently, it was established by several researchers that Massey's directed information and its variants indeed characterize the capacity of feedback and two-way channels [11, 12, 13, 22, 23, 27, 28, 29], and the rate distortion function with feedforward [31].

The *causal conditioning* notion $(\cdot||\cdot)$ was introduced by Kramer [12] in 1998. The probability mass function of $X^n = (X_1, X_2, \ldots, X_n)$ causally conditioned on Y^{n-d}, for some non-negative integer d, is denoted by $p(x^n||y^{n-d})$, and is defined as

$$p(x^n||y^{n-d}) = \prod_{i=1}^n p(x^i|x^{i-1}, y^{i-d}),$$

with the convention that if $i - d \leq 0$, then y^{i-d} is set to null. In light of the chain rule, it is easy to see that

$$p(x^n|y^n) = p(x^n||y^n)\, p(y^n||x^{n-1}).$$

The *causally conditional entropy* $H(X^n||Y^n)$ is defined as the expected logarithm of $p(X^n||Y^n)$, i.e.,

$$H(X^n||Y^n) = E[\log p(X^n||Y^n)] = \sum_{i=1}^{n} H(X_i|X^{i-1}, Y_i).$$

With this notion, the directed information from Y^n to X^n can be represented as

$$I(Y^n \to X^n) = \sum_{i=1}^{n} I(X_i; Y^i|X^{i-1}) = H(X^n) - H(X^n||Y^n).$$

2.3 Gambling with Causal Side Information

In 2008, Permuter *et al* [20, 21] investigated the problem of gambling in horse races and show that Massey's directed information characterizes the increment in the maximum achievable capital growth rate due to the availability of side information. This result gave a natural interpretation of directed information $I(Y^n \to X^n)$ as the amount of information that Y^n causally provides about X^n.

Denote

$X_i \in \mathscr{X} = [1, 2, \ldots, m]$: the horse that wins at time i,
Y_i : the side information at time i,
$o(X_i|X^{i-1})$: the payoffs at time i for horse X_i given that in the previous race, the horses X^{i-1} won,
$b(X_i|Y^i, X^{i-1})$: the fractions of the gambler's wealth invested in horse X_i at time i, given that the side information available at time i is Y^i and the horses X^{i-1} won in the previous races,
$S(X^n||Y^n)$: the gambler's wealth after n races when the outcomes of the races are X^n and the side information Y^n is causually available,
$W(X^n||Y^n)$: the growth, defined as the expected logarithm of the gambler's wealth

$$W(X^n||Y^n) = E[\log S(X^n||Y^n)],$$

$\frac{1}{n}W(X^n||Y^n)$: the growth rate.

Suppose that at any time n the gambler invest all his capital, then the total wealth after n rounds is

$$S(X^n||Y^n) = b(X_n|X^{n-1}, Y^n)\, o(X_n|X^{n-1})\, S(X^{n-1}||Y^{n-1}) = \prod_{i=1}^{n} b(X_i|X^{i-1}, Y^i)\, o(X_i|X^{i-1}).$$

The maximum growth rate is obtained if the gambler invests the money proportional to the causal conditioning distribution, i.e., for all $i \leq n$,

$$b^*(x_i|x^{i-1}, y^i) = p(x_i|x^{i-1}, y_i),$$

or equivalently,

$$b^*(x^n||y^n) = p(x^n||y^n),$$

and the optimal growth in this case is

$$W^*(X^n||Y^n) = E[\log o(X^n)] - H(X^n||Y^n).$$

If, in addition, the odds are fair and uniform, then $o(X_i|X^{i-1}) = \frac{1}{|\mathscr{X}|}$, and

$$\frac{1}{n}W^*(X^n||Y^n) = \log|\mathscr{X}| - \frac{1}{n}H(X^n||Y^n).$$

That means the sum of the optimal growth rate $\frac{1}{n}W^*(X^n||Y^n)$ and the entropy rate $\frac{1}{n}H(X^n||Y^n)$ of the horse race process causally conditioned on the side information, is a constant.

Now consider the case that the gambler invests only part of his capital. Let $b_0(y^i, x^{i-1})$ be the portion of capital not invested at time i given that the side information is y^i and the horses x^{i-1} won in the previous races. The wealth after n races is

$$S(X^n||Y^n) = \prod_{i=1}^{n}\left[b_0(Y^i, X^{i-1}) + b(X_i|X^{i-1}, Y^i)o(X_i|X^{i-1})\right].$$

The growth $W(X_n||Y_n)$ obeys a chain rule as

$$W(X^n||Y^n) = \sum_{i=1}^{n} W(X_i|X^{i-1}, Y^i),$$

where

$$W(X_i|X^{i-1}, Y^i) = E\left[b_0(Y^i, X^{i-1}) + b(X_i|X^{i-1})o(X_i|X^{i-1}, Y^i)\right].$$

It can be shown that the optimization problem of maximizing the growth $W(X^n||Y^n)$ is equivalent to the convex problem of maximizing $\sum_{x\in\mathscr{X}} p(x)\log[b_0 + b(x)o(x)]$ with the constrain $b_0 + \sum_{x\in\mathscr{X}} b(x) = 1$, which is exactly the same as the problem arised in Kelly strategy discussed in **2.1**.

3 Information Theory and Economics

The essential requirements of Kelly gambling strategy are the posibility of reinvestment of profits and the ability to control the amount of money invested in different categories. The channel may be extended to a real communication channel or just simply corresponds to the side information available to the investor. The optimization process in Kelly strategy is similar to that for rebalanced portfolios with known underlying distributions. Thus, Kelly strategy can be generalized to apply to portfolio theory to establish optimal portfolio strategies for investment in the stock market with or without side information.

3.1 Portfolio Theory

Portfolio Theory is a theory of finance that concentrates on maximizing the portfolio expected return for a given amount of portfolio risk, or equivalently, minimizing risk for a given level of expected return, by carefully choosing the proportions of various assets. Portfolio Theory was first introduced in 1952 by Markowitz [15] (see also [16]). The fundamental concept of Portfolio Theory is that the assets in an investment portfolio should not be chosen independently, each on their own merits. Rather, it is significant to consider how a change in price of each asset relates to that of every other asset in the portfolio.

Investing is a tradeoff between risk and expected return. In general, assets with higher expected returns have more risk. Given an amount of risk, Portfolio Theory describes how to choose a portfolio with the highest possible expected return. Conversely, for a given expected return, Portfolio Theory explains how to choose a portfolio with the lowest possible risk. Thus, Portfolio Theory is a form of diversification. Under certain assumptions and for specific quantitative definitions of risk and return, Portfolio Theory explains how to find the best possible diversification strategy.

3.2 Stock Market

The behavior of the stock market in n trading days is represented by non-negative price-relative sequence of stock market vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. Each stock market vector $\mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{im})^T$ represents the performance of the stock market on day i, where m is the number of stocks, and x_{ij} is the price relative of the j^{th} stock on day i, i.e., the ratio of closing to opening price of stock j on day i. A portfolio $\mathbf{b}_i = (b_{i1}, b_{i2}, \ldots, b_{im}) \in \mathscr{B}$, where the simplex $\mathscr{B}$ is the $(m-1)$-dimensional portfolios given by

$$\mathscr{B} = \left\{ \mathbf{b} \in \mathbb{R}^m : b_i \geq 0, \sum_{i=1}^{m} b_i = 1 \right\},$$

represents the propotion of wealth invested in each stock on day i. The wealth after n trading days for the portfolio strategy $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)$ is

$$S_n(\mathbf{b}) = \prod_{i=1}^{n} \mathbf{b}_i^T \mathbf{X}_i.$$

The objective of an optimal portfolio strategy is to maximize this wealth.

3.3 Constant Rebalanced Portfolios

A constant rebalanced portfolio is a sequential investment strategy that maintains fixed through time, trading period by trading period, the wealth distribution among a set of assets. In the other words, a constant rebalanced portfolio strategy uses the same portfolio $\mathbf{b}$ for each trading period. Let $\mathbf{X}^n = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ be a sequence of stock vectors, the constant rebalanced portfolio strategy using portfolio $\mathbf{b}$ achieves wealth $S_n(\mathbf{b}, \mathbf{X}^n)$, or just simply $S_n(\mathbf{b})$, given by

$$S_n(\mathbf{b}, \mathbf{X}^n) = S_n(\mathbf{b}) = \prod_{i=1}^{n} \mathbf{b}^T \mathbf{X}_i.$$

It requires a great deal of trading to maintain a constant rebalanced portfolio starategy because at the end of each trading period i, the proportion of wealth invested in each stock has changed from $b_1, b_2, \ldots, b_m$ to $\frac{x_{i1}b_1}{\mathbf{b}^T\mathbf{X}_i}, \frac{x_{i2}b_2}{\mathbf{b}^T\mathbf{X}_2}, \ldots, \frac{x_{im}b_m}{\mathbf{b}^T\mathbf{X}_m}$; and therefore, stocks must be bought and sold to restore the proportions of wealth back to $b_1, b_2, \ldots, b_m$ for the next trading period.

Given a sequence of stock vectors $\mathbf{X}^n$, the maximum wealth is denoted by $S_n^*(\mathbf{X}^n)$, or just simply S_n^*:

$$S_n^*(\mathbf{X}^n) = S_n^* = \max_{\mathbf{b} \in \mathscr{B}} S_n(\mathbf{b}).$$

The best constant rebalanced portfolio is the one that achieves this wealth S_n^*, and it is denoted by $\mathbf{b}^*(\mathbf{X}^n)$, or simply $\mathbf{b}^*$:

$$\mathbf{b}^*(\mathbf{X}^n) = \mathbf{b}^* = \arg\max_{\mathbf{b} \in \mathscr{B}} S_n(\mathbf{b}).$$

The exponential growth rate of wealth achieved by the best constant rebalanced portfolio at time n is denoted by $W_n^*(\mathbf{X}^n)$, and it is given by

$$W_n^*(\mathbf{X}^n) = \frac{1}{n} \log S_n^*(\mathbf{X}^n).$$

3.4 Side Information

Investors often use various sources of side information to adjust and update their portfolios. This side information is modeled as a finite-valued variable y made available at the start of each investment period. The knowledge of y is then used in the process of choosing the portfolio. Hence, the formal domain of the stock market model is a sequence of oredered pairs $\{(\mathbf{X}_i, y_i)\}$, where $\mathbf{X}_i$ is the stock vector of day i, and $y_i \in \mathscr{Y} = \{1, 2, \ldots, k\}$ is the state of the side information at time i.

3.5 State-Constant Rebalanced Portfolios

The constant rebalanced portfolio is designed to handle the portfolio selection problem when there is no additional information concerning the stock market. To overcome this limitation, Cover and Ordentlich [6] proposed the state constant rebalanced portfolio, which is capable of appropriately exploiting the available side information of the stock market. The constant rebalanced portfolio is generalized to the state-constant rebalanced portfolio by allowing the portfolio decisions to vary with the side information y. A state-constant rebalanced portfolio specifies portfolios $\mathbf{b}(1), \mathbf{b}(2), \ldots, \mathbf{b}(k) \in \mathscr{B}$, and uses portfolio $\mathbf{b}(y_i)$ at time i when the side information state takes on value $y_i \in \mathscr{Y} = \{1, 2, \ldots, k\}$. Note that, in the special case of $k = 1$, i.e., $|\mathscr{Y}| = 1$, the corresponding state-constant rebalanced portfolio is just simply the constant rebalanced portfolio.

For a sequence of stock vectors $\mathbf{X}^n = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$, and side information y^n, the choice of portfolios $\mathbf{b} : \mathscr{Y} \to \mathscr{B}$ results in the wealth

$$S_n(\mathbf{b}(\cdot), \mathbf{X}^n | y^n) = \prod_{i=1}^{n} \mathbf{b}(y_i)^T \mathbf{X}_i.$$

Let $\mathscr{B}^k$ be the set of all state-constant rebalanced portfolios with k states. Given a sequence of stock vectors $\mathbf{X}^n$ and side information y^n, the maximum wealth is denoted by $S_n^*(\mathbf{X}^n | y^n)$ and is given by

$$S_n^*(\mathbf{X}^n | y^n) = \max_{\mathbf{b}(\cdot) \in \mathscr{B}^k} S_n(\mathbf{b}(\cdot), \mathbf{X}^n | y^n).$$

The best state-constant rebalanced portfolio is the one that achieves this wealth $S_n^*(\mathbf{X}^n | y^n)$, and it is denoted by $\mathbf{b}^*(\cdot)$:

$$\mathbf{b}^*(\cdot) = \arg \max_{\mathbf{b}(\cdot) \in \mathscr{B}^k} S_n(\mathbf{b}(\cdot), \mathbf{X}^n | y^n).$$

The exponential growth rate of wealth achieved by the best state-constant rebalanced portfolio is

$$W_n^*(\mathbf{X}^n | y^n) = \frac{1}{n} \log S_n^*(\mathbf{X}^n | y^n).$$

3.6 Universal Portfolios

The universal portfolio was introduced in 1991 by Cover [5]. Let $\mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{im})^T$ be the stock market vector for one investment period on day i, where x_{ij} is the price relative of the j^{th} stock on day i, i.e., the ratio of closing to opening price of stock j on day i. An investment at time i in the stock market is represented by a portfolio vector $\mathbf{b}_i = (b_{i1}, b_{i2}, \ldots, b_{im})^T$, where all $b_{ij} \geq 0$ and $\sum_{j=1}^{m} b_{ij} = 1$. The components b_{ij} of $\mathbf{b}_i$ are the proportions of the current wealth invested in each of the m stocks on day i. The wealth after n trading days for the portfolio strategy $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)$ is

$$S_n(\mathbf{b}) = \prod_{i=1}^{n} \mathbf{b}_i^T \mathbf{X}_i,$$

where the initial wealth is normalized to 1, i.e. $S_0(\mathbf{b}) = 1$. Consider as the goal $S_n^*(\mathbf{b})$, the maximum wealth achievable over all constant rebalanced portfolio strategies $\mathbf{b}$, including those obtained by assuming perfect knowledge of future stock prices:

$$S_n^*(\mathbf{b}) = \max_{\mathbf{b}} S_n(\mathbf{b}).$$

It can be shown that S_n^* exceeds the best stocks, the Dow Jones average, and the value line index at time n. Cover [5] constructively showed that there is an *universal portfolio* strategy $\hat{\mathbf{b}}_k$, where $\hat{\mathbf{b}}_k$ is based only on past $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{k-1}$, such that the performance of $\hat{\mathbf{b}}_k$ is as good as the best portfolio based on the prior knowledge of the sequence of price relatives. This universal portfolio strategy $\hat{\mathbf{b}}_k$ is a performance weighted strategy given by,

$$\hat{\mathbf{b}}_1 = \left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right),$$

$$\hat{\mathbf{b}}_{k+1} = \frac{\int_{\mathscr{B}} \mathbf{b}\, S_k(\mathbf{b})\, d\,\mathbf{b}}{\int_{\mathscr{B}} S_k(\mathbf{b})\, d\,\mathbf{b}}$$

where

$$S_k(\mathbf{b}) = \prod_{i=1}^{k} \mathbf{b}^T \mathbf{X}_i,$$

and integration is over the set of $(m-1)$-dimensional portfolios

$$\mathscr{B} = \left\{ \mathbf{b} \in \mathbb{R}^m : b_i \geq 0, \sum_{i=1}^{m} b_i = 1 \right\}.$$

Hence, the initial universal portfolio $\hat{\mathbf{b}}_1$ is uniform over the stocks, and the portfolio $\hat{\mathbf{b}}_k$ at time k is the performance average of all portfolios $\mathbf{b}$ in $\mathscr{B}$. The wealth $\hat{S}_n$ generated by the universal portfolio is given by,

$$\hat{S}_n = \prod_{i=1}^{n} \hat{\mathbf{b}}_i^T \mathbf{X}_i.$$

It can be shown that, for arbitrary bounded stock sequences $\mathbf{X}_1, \mathbf{X}_2, \dots$, $\hat{S}_n$ and S_n^* have the same exponent to first order as

$$\frac{1}{n} \ln \frac{\hat{S}_n}{S_n^*} \longrightarrow 0.$$

A more refined analysis of the wealth generated by universal portfolio strategy shows that, in the special case of two assets, i.e., for $m = 2$ stocks,

$$\hat{S}_n \cong \sqrt{\frac{2\pi}{nJ_n}} S_n^*,$$

and in the general case of m stocks, under certain conditions,

$$\hat{S}_n \cong \frac{S_n^* (m-1)! \left(\frac{2\pi}{n}\right)^{\frac{m-1}{2}}}{|J_n|^{\frac{1}{2}}}.$$

where J_n is a $(m-1) \times (m-1)$ matrix and represents the sensitivity of the stock market.

3.7 Universal Portfolios with Side Information

In 1996, Cover and Ordentlich [6] studied Universal Portfolios with side information for $|\mathscr{Y}| = k > 1$. The μ-weighted universal portfolio with side information is given by

$$\hat{\mathbf{b}}_i(y) = \frac{\int_{\mathscr{B}} \mathbf{b} S_{i-1}(\mathbf{b}\,|y)\, d\mu(\mathbf{b})}{\int_{\mathscr{B}} S_{i-1}(\mathbf{b}\,|y)\, d\mu(\mathbf{b})}, \qquad i = 1, 2, \ldots, y \in \mathscr{Y},$$

with

$$\int_{\mathscr{B}} d\mu(\mathbf{b}) = 1,$$

where $S_i(\mathbf{b}\,|y)$ is the wealth obtained by the constant rebalanced portfolio $\mathbf{b}$ along the sequence $\{j \le i : y_j = y\}$ and is given by

$$S_i(\mathbf{b}\,|y) = \prod_{j \le i : y_j = y} \mathbf{b}^T \mathbf{X}_j \qquad \text{with}\, S_0(\mathbf{b}\,|y) = 1.$$

The corresponding wealth generated by the universal portfolio with side information is

$$\hat{S}_n(\mathbf{X}^n|y^n) = \prod_{i=1}^{n} \hat{\mathbf{b}}_i(y_i)\mathbf{X}_i$$

which can be expressed more compactly as,

$$\hat{S}_n(\mathbf{X}^n|y^n) = \prod_{y=1}^{k} \int_{\mathscr{B}} S_n(\mathbf{b}\,|y)\, d\mu(\mathbf{b}).$$

Recall that $S_n^*(\mathbf{X}^n|y^n)$ is the wealth achieved by the best μ-weighted state-constant rebalanced portfolio with side information. It can be shown that there exists measure μ for which the μ-weighted universal portfolio with side information $\hat{\mathbf{b}}(\cdot)$ is universal for the state-constant rebalanced portfolios $\mathscr{B}^k$ in the sense that

$$\lim_{n\to\infty} \sup_{\mathbf{X}^n, y^n} \frac{1}{n} \log \frac{S_n^*(\mathbf{X}^n|y^n)}{\hat{S}_n(\mathbf{X}^n|y^n)} = 0.$$

In other words, there exists μ such that $\hat{S}_n(\mathbf{X}^n|y^n)$ and $S_n^*(\mathbf{X}^n|y^n)$ have the same exponent in first order, just like the case of universal portfolios with no side information discussed in **3.6**.

This is established for two choices of μ, namely, the uniform (i.e., the Dirichlet $(1, \ldots, 1)$) distribution, and the Dirichlet $(\frac{1}{2}, \ldots, \frac{1}{2})$ distribution on the portfolio simplex $\mathscr{B}$ by showing that if μ equals to the uniform distribution then

$$\sup_{\mathbf{X}^n, y^n} \frac{1}{n} \log \frac{S_n^*(\mathbf{X}^n|y^n)}{\hat{S}_n(\mathbf{X}^n|y^n)} = \sup_{\mathbf{X}^n, y^n} \left(W_n^*(\mathbf{X}^n|y^n) - \hat{W}_n(\mathbf{X}^n|y^n) \right) \le \frac{k(m-1)}{n} \log(n+1),$$

and if μ is chosen to be the Dirichlet $(\frac{1}{2}, \ldots, \frac{1}{2})$ distribution then

$$\sup_{\mathbf{X}^n, y^n} \frac{1}{n} \log \frac{S_n^*(\mathbf{X}^n|y^n)}{\hat{S}_n(\mathbf{X}^n|y^n)} = \sup_{\mathbf{X}^n, y^n} \left(W_n^*(\mathbf{X}^n|y^n) - \hat{W}_n(\mathbf{X}^n|y^n)\right) \leq \frac{k(m-1)}{2n} \log(n+1) + \frac{k}{n} \log 2.$$

Since both bounds tend to 0 as $n \to \infty$, the universality for these choices of μ follows.

Acknowledgment. A part of this paper was written during the author's visit to Chiang Mai University, Thailand, in January 2012. The author is thankful to the Faculty of Economics and the Department of Mathematics of Chiang Mai University for their hospitalities. The author would like to sincerely thank Professors Songsak Sriboonchitta and Sompong Dhompongsa for helpful discussions as well as invaluable comments and suggestions, which indeed initiated this paper. The author is grateful to the referee for a very meticulous reading of this manuscript, and for valuable suggestions which help to create an improved final version.

References

1. Bell, R., Cover, T.M.: Competitive optimality of logarithmic investment. Math. Oper. Res. 5, 161–166 (1980)
2. Bell, R., Cover, T.M.: Game-theoretic optimal portfolios. Manage. Sci. 34, 724–733 (1988)
3. Breiman, L.: Optimal gambling systems for favourable games. In: Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 65–78 (1961)
4. Berrou, C., Glavieux, A., Thitimajshima, P.: Near Shannon Limit Error-correcting Coding and Decoding: Turbo-codes. In: IEEE Inter. Conf. Comm., pp. 1064–1070 (1993)
5. Cover, T.M.: Universal portfolios. Math. Finance 1, 1–29 (1991)
6. Cover, T.M., Ordentlich, E.: Universal portfolios with side information. IEEE Trans. Inf. Theory 42, 348–363 (1996)
7. Gallager, R.G.: Low Density Parity-Check Codes. MIT Press, Cambridge (1963)
8. Hamming, R.W.: Error detecting and error correcting codes. Bell Sys. Tech. J. 29, 147–160 (1950)
9. James, I.: Claude Elwood Shannon 30 April 1916 - 24 February 2001. Biographical Memoirs of Fellows of the Royal Society 55, 257–265 (2009)
10. Kelly, J.: A new interpretation of information rate. Bell Syst. Tech. J. 35, 917–926 (1956)
11. Kim, Y.-H.: A coding theorem for a class of stationary channels with feedback. IEEE Trans. Inf. Theory. 54, 1488–1499 (2008)
12. Kramer, G.: Directed information for channels with feedback, Ph.D. Dissertation, Swiss Federal Institute of Technology (ETH), Zurich (1998)
13. Kramer, G.: Capacity results for the discrete memoryless network. IEEE Trans. Inf. Theory 49, 4–21 (2003)
14. Marko, H.: The bidirectional communication theory - a generalization of information theory. IEEE Trans. Comm. 21, 1345–1351 (1973)
15. Markowitz, H.M.: Portfolio Selection. Journal of Finance 7, 77–91 (1952)
16. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments. John Wiley & Sons, New York (1959)
17. Massey, J.: Causality, feedback and directed information. In: Proc. Int. Symp. Inf. Theory Applic., ISITA 1990 (1990)
18. MIT News, MIT Professor Claude Shannon dies; was founder of digital communications, http://web.mit.edu/newsoffice/2001/shannon.html

19. Pabrai, M.: The Dhandho Investor: The Low-Risk Value Method to High Returns. Willey (2007)
20. Permuter, H.H., Kim, Y.-H., Weissman, T.: On directed information and gambling. CoRR (February 2008)
21. Permuter, H.H., Kim, Y.-H., Weissman, T.: Interpretations of Directed Information in Portfolio Theory, Data Compression, and Hypothesis Testing. IEEE Trans. Inform. Theory 57, 3248–3259 (2011)
22. Permuter, H.H., Weissman, T., Chen, J.: Capacity region of the finite-state multiple-access channel with and without feedback. IEEE Trans. Inf. Theory. 55, 2455–2477 (2009)
23. Permuter, H.H., Weissman, T., Goldsmith, A.J.: Finite state channels with time-invariant deterministic feedback. IEEE Trans. Inf. Theory. 55, 644–662 (2009)
24. Peterson, W.W., Brown, D.T.: Cyclic Codes for Error Detection. Proceedings of the IRE 49, 228 (1961)
25. Shannon, C.E.: A Symbolic Analysis of Relay and Switching Circuits. M.Sc. Thesis, Massachusetts Institute of Technology (1937)
26. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423, 623–656 (1948)
27. Shrader, B., Permuter, H.H.: On the compound finite state channel with feedback. In: Proc. Internat. Symp. Inf. Theory, Nice, France, pp. 396–400 (2007)
28. Tatikonda, S.: Control under communication constraints, Ph.D. disertation, Massachusetts Institute of Technology, Cambridge, MA (2000)
29. Tatikonda, S., Mitter, S.: The capacity of channels with feedback. IEEE Trans. Inf. Theory. 55, 323–349 (2009)
30. Thorp, E.O.: The Kelly Criterion: Part II. Wilmott Magazine (September 2008)
31. Venkataramanan, R., Pradhan, S.S.: Source coding with feedforward: Rate-distortion theorems and error exponents for a general source. IEEE Trans. Inf. Theory 53, 2154–2179 (2007)

Why Clayton and Gumbel Copulas: A Symmetry-Based Explanation

Vladik Kreinovich, Hung T. Nguyen, and Songsak Sriboonchitta

Abstract. In econometrics, many distributions are non-Gaussian. To describe dependence between non-Gaussian variables, it is usually not sufficient to provide their correlation: it is desirable to also know the corresponding copula. There are many different families of copulas; which family shall we use? In many econometric applications, two families of copulas have been most efficient: the Clayton and the Gumbel copulas. In this paper, we provide a theoretical explanation for this empirical efficiency, by showing that these copulas naturally follow from reasonable symmetry assumptions. This symmetry justification also allows us to provide recommendations about which families of copulas we should use when we need a more accurate description of dependence.

Keywords: Archimedean copulas, econometrics, symmetries.

1 Formulation of the Problem

Copulas are needed. Traditionally, in statistics the dependence between random variables η, ν, ..., is described by their correlation. This description is well

Vladik Kreinovich
Department of Computer Science,
University of Texas at El Paso,
500 W. University, El Paso, TX 79968, USA
e-mail: vladik@utep.edu

Hung T. Nguyen
Department of Mathematical Sciences,
New Mexico State University, Las Cruces,
New Mexico 88003, USA
e-mail: hunguyen@nmsu.edu

Songsak Sriboonchitta · Hung T. Nguyen
Department of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: songsak@econ.chiangmai.ac.th

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 79–90.
DOI: 10.1007/978-3-642-35443-4_6

justified in the frequent cases when the joint distribution is Gaussian: in this case, to describe the joint distribution, i.e., to describe the corresponding cumulative distribution function $P(\eta \le x \,\&\, \nu \le y \,\&\, \ldots)$, it is sufficient to describe the marginal distribution $F_\eta(x) \stackrel{\text{def}}{=} P(\eta \le x)$, $F_\nu(y) \stackrel{\text{def}}{=} P(\nu \le y)$, $\ldots$, of each of the variables, and the correlations between each pairs of variables.

In many practical situations, e.g., in economics, the distributions are often non-Gaussian; see, e.g., [5]. For non-Gaussian pair of variables (η, ν), in general, it is not enough to know the distribution of each variables and the correlations between them, we need more information about the dependence. Such information is provided, e.g., by a *copula*, i.e., by a function $C(u,v)$ for which $P(\eta \le x \,\&\, \nu \le y) = C(F_\eta(x), F_\nu(y))$; see, e.g., [4, 5, 6].

Usually, Archimedean copulas are used, i.e., copulas of the form $C(u,v) = \psi(\psi^{-1}(u) + \psi^{-1}(v))$ for some decreasing *generator function* $\psi(x)$ that maps $[0,\infty)$ into $(0,1]$.

Most efficient Archimedean copulas. In econometric applications, the following two classes of Archimedean copulas turned out to be most efficient [5]: the Frank copulas $C(u,v) = -\dfrac{1}{\theta} \cdot \ln\left(1 - \dfrac{(1-\exp(-\theta\cdot u))\cdot(1-\exp(-\theta\cdot v))}{1-\exp(-\theta)}\right)$, the Clayton copulas $C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$ and the Gumbel copulas $C(u,v) = \exp\left(\left((-\ln(u))^{-\theta} + (-\ln(v))^{-\theta} - 1\right)^{-1/\theta}\right)$. The efficiency of Frank's copulas is clear: Frank copulas are the only Archimedean copulas which satisfy the natural condition $C(u,v) + C(u,1-v) + C(1-u,v) + C(1-u,1-v) = 1$ that describes the intuitive idea that for every two events U and V, we should have $P(U \,\&\, V) + P(U \,\&\, \neg V) + P(\neg U \,\&\, V) + P(\neg U \,\&\, \neg V) = 1$ (see, e.g., [4, 5, 6]). But why Clayton and Gumbel copulas?

Main question: why Clayton and Gumbel copulas? In principle, there are many different copulas. So why did the above two classes turned out to be the most efficient in econometrics?

Auxiliary question: what if these copulas are not sufficient? While at present, the above two classes of copulas provide a good description of all observed dependencies, in the future, we will need to describe this dependence in more detail, so we will larger classes of copulas. Which classes should we use?

What we do in this paper. In this paper, we provide answers to both questions. Specifically, we show that natural symmetry-based ideas indeed explain the efficiency of the above classes of copulas, and that these same ideas can lead us, if necessary, to more general classes.

2 Why Symmetries

Symmetries as a fundamental description of knowledge: brief reminder. Symmetries are one of the fundamental concepts of modern physics. The reason for their ubiquity is that most of our knowledge is based on symmetry; see, e.g., [2].

Indeed, how do we gain any knowledge about the physical world? Let us start with a simple example: we observe many times that the Sun rises every morning, and we conclude that it will rise again. This conclusion is based on the implicit assumption that the dynamics of the Solar system does not change when we move from one day to another.

Similarly, we drop a rock, and it falls down with an acceleration of 9.81 m/sec^2. We repeat this experiment at different locations on the Earth, we repeat it at the same location turning to different places, and we always get the same acceleration. We therefore conclude that at all locations on the Earth surface, no matter what our orientation is, the rock will drop with the same acceleration. This means that no matter how we shift or rotate, the fundamental laws of physics do not change.

In general, when we formulate a physical law based on observations, we thus implicitly assume that new situations are similar to the already observed ones, so the regularities that we observed earlier will happen in future situations as well. This idea has been formalized in modern physics, to the extend that may physical theories (starting with the quarks theory) are formulated not in terms of differential equations as before, but explicitly in terms of appropriate symmetries [2]. Moreover, it was discovered that many fundamental physical equations – e.g., Maxwell equations of electrodynamics, Schrödinder's equations of quantum mechanics, Einstein's equations of General Relativity etc. – can be uniquely derived from the corresponding symmetries; see, e.g., [3].

In view of efficiency of symmetries in physics, it is reasonable to use them in other disciplines as well; for example, in [7], we have shown that symmetries can be efficiently applied to computing.

Basic symmetries. In this paper, we will use the *basic* symmetries that come from the fact that the numerical value of a physical quantity depends on the choice of the measuring unit and on the choice of a starting point.

Let us start with the choice of the measuring unit. For example, when we measure lengths and instead of using meters, start using centimeters – a unit which is $\lambda = 100$ times smaller than the meter – instead of the original numerical values x, we get new values $x' = \lambda \cdot x$ which are λ times larger. Many fundamental physical processes do not have any preferred unit of length; for such processes, it is reasonable to require that the corresponding equations do not change if we simply change the units. The corresponding transformations are called *scalings*, and invariance under such transformation is known as *scale-invariance.* It is worth mentioning that scale-invariance is an important part of symmetry-based derivation of the fundamental physical equations presented in [3].

Another basic symmetry is the possibility to select different starting points for measurements. For example, when we measure time, we can arbitrarily select the starting point: instead of the usual calendar that starts at Year 0, we can start, as the French Revolution proposed, so start with the date of the Revolution. In this case, instead of the original numerical value x, we get a new value $x' = x + s$, where s is the difference between the starting points (e.g., $s = -1789$ for the French Revolution). Many fundamental physical quantities like time do not have any preferred starting point; for such processes, it is reasonable to require that the corresponding equations do not change if we simply change the starting point. The corresponding transformations are called *shifts*, and invariance under such transformation is known as *shift-invariance.*

3 Invariant Functions Corresponding to Basic Symmetries

Example of invariance. A power law $y = x^a$ has the following invariance property: if we change a unit in which we measure x, then in the new units, we get the exact same formula – provided that we also appropriately changing a measuring unit for y. Let us explain this property in detail.

If we replace the original unit for measuring x by a new measuring unit which is λ times larger, then all the numerical values get decreased by a factor of λ. In other words, instead of the original values x, we have new values $x' = \frac{x}{\lambda}$. How will the dependence of y on x look in the new units?

From the above dependence of x' on x, we conclude that $x = \lambda \cdot x'$. Substituting this expression into the formula $y = x^a$, we conclude that $y = (\lambda \cdot x')^a = \lambda^a \cdot (x')^a$. This new formula is different from the original formula – it has not only x' raised to the power a, it also has a multiplicative constant $\mu \stackrel{\text{def}}{=} \lambda^a$. However, we can make this formula exactly the same if we also select a new unit for y: namely, a unit which is μ times larger than the original one. Now, instead of the original values y, we get new values $y' = \frac{y}{\mu}$. In these new units, due to $y = \mu \cdot (x')^a$, the dependence of y on x takes the form $y' = (x')^a$ – i.e., exactly the same form as in the previous units.

Towards a general description of invariant functions corresponding to basic symmetries. Let us now provide a general description of invariant functions corresponding to basic symmetries. As we have mentioned, there are two types of basic symmetries: scaling (corresponding to a change in measuring unit) and shift (corresponding to the change in the starting point). When we are looking for invariant functions $y = f(x)$, we have 2 possible symmetries for x and 2 possible symmetries for y, so we need to consider all $2 \times 2 = 4$ possible combinations of these symmetries. Let us describe what happens in all these 4 cases: scale $\to$ scale, scale $\to$ shift, shift $\to$ scale, and shift $\to$ shift.

Definition 1. *A differentiable function $f(x)$ is called* scale-to-scale invariant *if for every λ, there exists a μ for which $f(\lambda \cdot x) = \mu \cdot f(x)$.*

Comment. In this case, if we replace x with $x' = \lambda \cdot x$, we can get the same dependence $y' = f(x')$ if we replace y with $y' = \frac{y}{\mu}$.

Proposition 1. *A function $f(x)$ is scale-to-scale invariant if and only if it has the form $f(x) = A \cdot x^a$ for some real numbers A and a.*

For readers' convenience, all the proofs are placed in a special (last) section.

Definition 2. *A differentiable function $f(x)$ is called* scale-to-shift invariant *if for every λ, there exists an s for which $f(\lambda \cdot x) = f(x) + s$.*

Comment. In this case, if we replace x with $x' = \lambda \cdot x$, we can get the same dependence $y' = f(x')$ if we replace y with $y' = y - s$.

Proposition 2. *A function $f(x)$ is scale-to-shift invariant if and only if it has the form $f(x) = A \cdot \ln(x) + b$ for some real numbers A and b.*

Definition 3. *A differentiable function $f(x)$ is called* shift-to-scale invariant *if for every s, there exists a λ for which $f(x+s) = \lambda \cdot f(x)$.*

Comment. In this case, if we replace x with $x' = x + s$, we can get the same dependence $y' = f(x')$ if we replace y with $y' = \dfrac{y}{\lambda}$.

Proposition 3. *A function $f(x)$ is shift-to-scale invariant if and only if it has the form $f(x) = A \cdot \exp(k \cdot x)$ for some real numbers A and k.*

Definition 4. *A differentiable function $f(x)$ is called* shift-to-shift invariant *if for every s, there exists a b for which $f(x+s) = f(x) + b$.*

Comment. In this case, if we replace x with $x' = x + s$, we can get the same dependence $y' = f(x')$ if we replace y with $y' = y - b$.

Proposition 4. *A function $f(x)$ is shift-to-shift invariant if and only if it has the form $f(x) = A \cdot x + c$ for some real numbers A and c.*

Definition 5. *A function is called* invariant *if it is either scale-to-scale invariant, or scale-to-shift invariant, or shift-to-scale invariant, or shift-to-shift invariant.*

Discussion. Not all physical dependencies are invariant. Specifically, when the mappings $y = f(x)$ from x to y and $z = g(y)$ are both invariant with respect to the same symmetries, then their composition $z = g(f(x))$ is also invariant with respect to the same symmetries. In general, however, the mappings $y = f(x)$ and $z = g(y)$ correspond to different symmetries; in this case, their composition is not necessarily invariant.

In view of this observation, if we want to use symmetries but cannot find an invariant function, we should be look for functions which are compositions of two invariant functions (if necessary, compositions of three, etc.) Let us apply this approach to our problem of finding appropriate copulas.

4 Why Scalings and Shifts Can Be Applied to Probabilities

Symmetries are actively used in statistics. One of the main objectives of mathematical statistics is to process data, in particular, physical data. As we have mentioned, in physics, symmetries are very important, in particular basic symmetries such as scalings and shifts. Not surprisingly, invariance with respect to different symmetries is one of the main tools in traditional mathematical statistics – especially scale- and shift-invariance; see, e.g., [1].

We plan to use symmetries in copula techniques as well. Our objective is go beyond traditional statistical techniques and to derive the formulas for copulas, i.e., the formulas that transform probabilities into probabilities.

Due to the success of symmetries in physics and in traditional mathematical statistics, it seems desirable to apply symmetries for copulas as well. Specifically, we would like to use the simplest – and most widely used – symmetries: scalings and shifts.

Challenge: for copulas, we need a new justification of symmetries. In physics (and in traditional mathematical statistics), the use of these symmetries is justified by the possibility to select different measuring units and different starting points. This justification cannot be *directly* applied to copulas – functions from probabilities to probabilities: since probabilities are limited by the interval $[0,1]$, for probabilities, we have a natural starting point (0) and a natural measuring unit (1).

We will show, however, that, by using arguments which are slightly more complex than in the general case, we can still justify the use of scalings and shifts in the (copula-related) probabilistic context.

Main idea of the new justification: considering conditional probabilities. We will show that the possibility of scaling naturally comes from the fact that most of the probabilities that we analyze are, in effect, *conditional* probabilities, and the numerical values of these probabilities change if we change the context.

Let us give two econometric examples that illustrate the possibility of scaling and shift of probabilities. For clarity, these examples are made as simple as possible.

An example illustrating the possibility of scalings. Let us assume that we want to invest in stable stocks, and we have selected several such stocks, i.e., stocks that only experience a drastic change in price when the market as a whole starts changing. We want to gauge stability of several such stocks.

One way to estimate such a stability is to divide the number of days c when this stock drastically changed by the total number of days N during which we kept the records. Alternatively, since the stock only changes when the market itself changes, we can divide c by the total number of days $n < N$ when the market drastically changed. The two resulting probabilities $p = \frac{c}{N}$ and $p' = \frac{c}{n}$ differ by a multiplicative constant $p' = \lambda \cdot p$, where $\lambda \stackrel{\text{def}}{=} \frac{n}{N}$.

Thus, in econometric applications, scaling makes sense for probabilities.

Possibility of scalings: general idea. In general, all the probabilities $P(E)$ of different events E are, in effect, conditional, i.e., have the form $P(E\,|\,U)$ for some universal event U. The numerical value of each such probability depends on the selection of the universal event U that contains all desired events E. If we replace the original universal event U with a new universal event U', then the original condition probability $P(E) \stackrel{\text{def}}{=} P(E\,|\,U) = \frac{P(E\,\&\,U)}{P(U)} = \frac{P(E)}{P(U)}$ is replaced by a new value $P'(E) \stackrel{\text{def}}{=} P(E\,|\,U') = \frac{P(E\,\&\,U')}{P(U')} = \frac{P(E)}{P(U')}$ which is related to the original value by the scaling formula $P'(E) = \lambda \cdot P(E)$, where $\lambda \stackrel{\text{def}}{=} \frac{P(U)}{P(U')}$.

An example illustrating the possibility of shifts. Suppose that we have a stock which always fluctuates when the market changes *and* also sometimes experiences drastic changes of its own. How can we estimate the stability of this stock? One way is to divide the number of days c when this stock drastically changed by the total number of days N during which we kept the record of this stock, and get an estimate $p = \frac{c}{N}$. However, since we know that this stock always changes when the market changes, it makes sense to only consider days when the market itself was stable, i.e., to use the estimate $p' = \frac{c-n}{N-n}$. Since $n \ll N$, we have $p' \approx \frac{c-n}{N} = p + s$, where $s \stackrel{\text{def}}{=} -\frac{n}{N}$.

Thus, in econometric applications, shifts also make sense for probabilities.

Discussion

- Similarly to the scaling example, this shift example can also be easily in a general form.
- In our analysis of copulas, we used the simplest (basic) symmetries: scalings and shifts. By using these basic symmetries, we came up with a reasonable explanation of the empirical success of specific copulas. In view of this result, we believe that it will be beneficial to perform a deeper analysis of the application of basic (and other) symmetries to copulas.

5 Archimedean Copulas with Whose Generators Are Either Invariant or Compositions of Two Invariant Functions

Now that we have given arguments that symmetries – including basic symmetries such as scalings and shifts – can be applied to econometric copulas, let us describe the corresponding results. Let us start by describing all the Archimedean copulas in which the generator is invariant.

Proposition 5. *The only Archimedean copula with an invariant generator is the copula $C(u,v) = u \cdot v$ corresponding to independence.*

Discussion. This result shows that to describe dependence, it is not sufficient to use invariant generators, we need to consider *compositions* of invariant generators.

Proposition 6. *The only Archimedean copulas in which a generator is a composition of two invariant functions are the following:*

- *Clayton copulas* $C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$;
- *the the Gumbel copulas* $C(u,v) = \exp\left(\left((-\ln(u))^{-\theta} + (-\ln(v))^{-\theta} - 1\right)^{-1/\theta}\right)$;
- *the copulas* $C(u,v) = \frac{1}{L}\exp\left(\frac{1}{\ell} \cdot \ln(u \cdot L) \cdot \ln(v \cdot L)\right)$, *with* $\ell = \ln(L)$.

Comment. Please notice that we have an additional family of copulas.

Discussion. This approach leads to a natural answer to a question of which copulas we should use when we the approximation provided by the existing Archimedean copulas is no longer sufficient:

- we should first try Archimedean copulas whose generator function is a composition of *three* invariant functions,
- if needed, we should move to Archimedean copulas whose generator function is a composition of *four* invariant functions,
- etc.

Example. The generator $\psi(x) = -\frac{1}{\theta} \cdot \ln(1-(1-\exp(-\theta)) \cdot \exp(-x))$ of Frank copulas can be obtained as a composition of three invariant transformations: first, we apply an invariant function $y = f(x) = (1-\exp(-\theta)) \cdot \exp(-x)$, then an invariant function $z = g(y) = 1-y$, and finally, an invariant function $t = h(z) = -\frac{1}{\theta} \cdot \ln(z)$.

6 Proofs

Proof of Proposition 1. It is easy to check that every function $f(x) = A \cdot x^a$ is scale-to-scale invariant. Vice versa, let $f(x)$ be a scale-to-scale invariant function. By definition, this means that for every λ, there exists a μ (depending on this λ) for which $f(\lambda \cdot x) = \mu(\lambda) \cdot f(x)$.

This property is trivially true when $f(x) = 0$ for all x. It is therefore sufficient to consider the cases when the function $f(x)$ is not identically 0. Let us prove, by contradiction, that in such cases, the function $f(x)$ cannot attain zero values for $x \neq 0$. Indeed, if $f(x_0) = 0$ for some $x_0 \neq 0$, then, for every other x, we will get $f(x) = \mu\left(\frac{x}{x_0}\right) \cdot f(x_0) = 0$. So, $f(x) \neq 0$ for $x \neq 0$.

Here, the function $f(x)$ is differentiable, the function $f(\lambda \cdot x)$ is also differentiable, and thus, their ratio $\mu(\lambda) = \frac{f(\lambda \cdot x)}{f(x)}$ is also differentiable. Differentiating both sides of the equation $f(\lambda \cdot x) = \mu(\lambda) \cdot f(x)$ by λ, we conclude that $x \cdot f'(\lambda \cdot x) = \mu'(\lambda) \cdot f(x)$. In particular, for $\lambda = 1$, we get $x \cdot f'(x) = \mu_0 \cdot f(x)$, where we denoted $\mu_0 \stackrel{\text{def}}{=} \mu'(1)$. This equation can be rewritten as $x \cdot \frac{df}{dx} = \mu_0 \cdot f$. In this equation, we can separate variables by moving all the terms containing df and f to the left side and all the terms containing dx and x to the right side. As a result, we get $\frac{df}{f} = \mu_0 \cdot \frac{dx}{x}$. Integrating both sides, we get $\ln(f) = \mu_0 \cdot \ln(x) + C$ for some constant C. Thus, we conclude that $f(x) = \exp(\ln(f(x))) = \exp(\mu_0 \cdot \ln(x) + C) = \exp(C) \cdot x^{\mu_0}$, which is exactly the desired form for the transformation $f(x)$, with $A = \exp(C)$ and $a = \mu_0$. The proposition is proven.

Proof of Proposition 2. It is easy to check that every function $f(x) = A \cdot \ln(x) + b$ is scale-to-shift invariant. Vice versa, let $f(x)$ be a scale-to-shift invariant function. By definition, this means that for every λ, there exists an s (depending on this λ) for which $f(\lambda \cdot x) = f(x) + s(\lambda)$.

Here, the function $f(x)$ is differentiable, the function $f(\lambda \cdot x)$ is also differentiable, and thus, their difference $s(\lambda) = f(\lambda \cdot x) - f(x)$ is also differentiable. Differentiating both sides of the equation $f(\lambda \cdot x) = f(x) + s(\lambda)$ by λ, we conclude that

$x \cdot f'(\lambda \cdot x) = s'(\lambda)$. In particular, for $\lambda = 1$, we get $x \cdot f'(x) = s_0$, where we denoted $s_0 \stackrel{\text{def}}{=} s'(1)$. This equation can be rewritten as $x \cdot \frac{df}{dx} = s_0$. In this equation, we can separate variables by moving all the terms containing df and f to the left side and all the terms containing dx and x to the right side. As a result, we get $df = s_0 \cdot \frac{dx}{x}$. Integrating both sides, we get $f = s_0 \cdot \ln(x) + C$ for some constant C. This is exactly the desired form for the transformation $f(x)$. The proposition is proven.

Proof of Proposition 3. It is easy to check that every function $f(x) = A \cdot \exp(k \cdot x)$ is shift-to-scale invariant. Vice versa, let $f(x)$ be a shift-to-scale invariant function. By definition, this means that for every shift s, there exists a λ (depending on this s) for which $f(x+s) = \lambda(s) \cdot f(x)$.

This property is trivially true when $f(x) = 0$ for all x. It is therefore sufficient to consider the cases when the function $f(x)$ is not identically 0. Let us prove, by contradiction, that in such cases, the function $f(x)$ cannot attain zero values at any x. Indeed, if $f(x_0) = 0$ for some x_0, then, for every other x, we will get $f(x) = \lambda(x - x_0) \cdot f(x_0) = 0$. So, $f(x) \neq 0$ for all x.

Here, the function $f(x)$ is differentiable, the function $f(x+s)$ is also differentiable, and thus, their ratio $\lambda(s) = \frac{f(x+s)}{f(x)}$ is also differentiable. Differentiating both sides of the equation $f(x+s) = \lambda(s) \cdot f(x)$ by s, we conclude that $f'(x+s) = \lambda'(s) \cdot f(x)$. In particular, for $s = 0$, we get $f'(x) = \lambda_0 \cdot f(x)$, where we denoted $\lambda_0 \stackrel{\text{def}}{=} \lambda'(0)$. This equation can be rewritten as $\frac{df}{dx} = \lambda_0 \cdot f$. In this equation, we can separate variables by moving all the terms containing df and f to the left side and all the terms containing dx and x to the right side. As a result, we get $\frac{df}{f} = \lambda_0 \cdot dx$. Integrating both sides, we get $\ln(f) = \lambda_0 \cdot x + C$ for some constant C. Thus, we conclude that $f(x) = \exp(\ln(f(x))) = \exp(\lambda_0 \cdot x + C) = \exp(C) \cdot \exp(\lambda_0 \cdot x)$, which is exactly the desired form for the transformation $f(x)$, with $A = \exp(C)$ and $k = \lambda_0$. The proposition is proven.

Proof of Proposition 4. It is easy to check that every function $f(x) = A \cdot x + b$ is shift-to-shift invariant. Vice versa, let $f(x)$ be a shift-to-shift invariant function. By definition, this means that for every shift s, there exists a b (depending on this s) for which $f(x+s) = f(x) + b(s)$.

Here, the function $f(x)$ is differentiable, the function $f(x+s)$ is also differentiable, and thus, their difference $b(s) = f(x+s) - f(x)$ is also differentiable. Differentiating both sides of the equation $f(x+s) = f(x) + b(s)$ by s, we conclude that $f'(x+s) = b'(s)$. In particular, for $a = 0$, we get $f'(x) = b_0$, where we denoted $b_0 \stackrel{\text{def}}{=} b'(0)$. Integrating this equation, we get $f = b_0 \cdot x + C$ for some constant C. This is exactly the desired form for the transformation $f(x)$. The proposition is proven.

Proof of Proposition 5. A generator $\psi(x)$ of an Archimedean copula should map 0 into 1 and ∞ into 0: $\psi(0) = 1$ and $\psi(\infty) \stackrel{\text{def}}{=} \lim_{x \to \infty} \psi(x) = 0$. One can easily check that most invariant functions do not satisfy this property:

- the function $f(x) = A \cdot x^a$ does not satisfy the property $f(0) = 1$;
- the function $f(x) = A \cdot \ln(x) + b$ does not satisfy the property $f(0) = 1$, and
- the function $f(x) = A \cdot x + b$ does not satisfy the property $f(\infty) = 0$.

The only remaining invariant function is $f(x) = A \cdot \exp(k \cdot x)$. For this function, from $f(0) = 1$, we conclude that $A = 1$, and from $f(\infty) = 0$, that $k < 0$. One can check that for this generator function $\psi(x) = \exp(-|k| \cdot x)$, the inverse is equal to $\psi^{-1}(u) = -\frac{1}{|k|} \cdot \ln(u)$, and thus, the corresponding copula has the form $C(u,v) = \psi(\psi^{-1}(u) + \psi^{-1}(v)) = u \cdot v$. The proposition is proven.

Proof of Proposition 6. We have 4 types of invariant functions $f(x)$ and 4 types of invariant functions $g(y)$, so we have $4 \times 4 = 16$ possible compositions $\psi(x) = g(f(x))$. Let us consider them one by one.

1°. Let us first consider the case when $f(x) = A \cdot x^a$.

1.1°. If $g(y)$ is of the same type $g(y) = B \cdot y^b$, then the composition is also of the same type, and we already know, from the proof of Proposition 5, that a function of this type cannot be a generator.

1.2°. If $g(y) = B \cdot \ln(y) + b$, then the composition has the form

$$\psi(x) = g(f(x)) = B \cdot \ln(A \cdot x^a) + b = (B \cdot a) \cdot \ln(x) + (B \cdot \ln(A) + b).$$

In this case, we cannot have $\psi(0) = 1$.

1.3°. If $g(y) = B \cdot \exp(k \cdot y)$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot \exp(k \cdot A \cdot x^a)$. The condition $\psi(0) = 1$ leads to $B = 1$ and $a > 0$, the condition $\psi(\infty) = 0$ leads to $k \cdot A < 0$. One can easily check that in this case, we get the Gumbel copula.

1.4°. If $g(y) = B \cdot y + b$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot A \cdot x^a + b$. The condition $\psi(\infty) = 0$ leads to 0, so $\psi(x) = (B \cdot A) \cdot x^a$, and the equality $\psi(0) = 1$ is not possible.

2°. Let us first consider the case when $f(x) = A \cdot \ln(x) + b$.

2.1°. If $g(y) = B \cdot y^a$, then the composition takes the form

$$\psi(x) = g(f(x)) = B \cdot (A \cdot \ln(x) + b)^a.$$

This function cannot satisfy the property $\psi(0) = 1$.

2.2°. If $g(y) = B \cdot \ln(y) + a$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot \ln(A \cdot \ln(x) + b) + a$, then we also cannot have the property $\psi(0) = 1$.

2.3°. If $g(y) = B \cdot \exp(k \cdot y)$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot \exp(k \cdot A \cdot \ln(x) + k \cdot b) = (B \cdot \exp(k \cdot b)) \cdot x^{k \cdot A}$, so we cannot have $\psi(0) = 1$.

2.4°. If $g(y) = B \cdot y + a$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot A \cdot \ln(x) + (B \cdot A + b)$, we also cannot have $\psi(0) = 1$.

3°. Let us consider the case when $f(x) = A \cdot \exp(k \cdot x)$.

3.1°. If $g(y) = B \cdot y^a$, then the composition has the form

$$\psi(x) = g(f(x)) = (B \cdot A^a) \cdot \exp((k \cdot a) \cdot x).$$

We already know, from the proof of Proposition 5, that such generator functions lead to the independence copula.

3.2°. If $g(y) = B \cdot \ln(y) + a$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot \ln(A \cdot \exp(k \cdot x)) + a = (B \cdot k) \cdot x + (B \cdot \ln(A) + a)$, so we cannot have $\psi(\infty) = 0$.

3.3°. If $g(y) = B \cdot \exp(a \cdot y)$, then the composition takes the form

$$\psi(x) = B \cdot \exp(\ell \cdot \exp(k \cdot x)),$$

for $\ell = a \cdot A$. The condition $\psi(0) = 1$ leads to $B \cdot \exp(\ell) = 1$, so $B = \exp(-\ell)$, and $\psi(x) = \exp(\ell \cdot (\exp(k \cdot x) - 1))$.

Let us describe the copula corresponding to this generator. For that, let us first find an explicit expression for the inverse function $\psi^{-1}(u)$. From the condition that $\psi(x) = u$, we conclude that $\ell \cdot (\exp(k \cdot x) - 1) = \ln(u)$, hence $\exp(k \cdot x) = 1 + \dfrac{\ln(u)}{\ell} = \dfrac{\ln(u \cdot L)}{\ell}$, where we denoted $L \stackrel{\text{def}}{=} \exp(\ell)$. Thus, $k \cdot x = \ln\left(\dfrac{\ln(u \cdot L)}{\ell}\right)$, and

$$x = \frac{1}{k} \cdot \ln\left(\frac{\ln(u \cdot L)}{\ell}\right).$$

To find $C(u,v)$, we compute $x + y$, where $x = \psi^{-1}(u)$ and $y = \psi^{-1}(v)$, then compute $z = x + y$ and $C(u,v) = \psi(z)$. Here,

$$z = x + y = \frac{1}{k} \cdot \left[\ln\left(\frac{\ln(u \cdot L)}{\ell}\right) + \ln\left(\frac{\ln(v \cdot L)}{\ell}\right)\right],$$

hence $k \cdot z = \ln\left(\dfrac{\ln(u \cdot L)}{\ell}\right) + \ln\left(\dfrac{\ln(v \cdot L)}{\ell}\right)$ and $\exp(k \cdot z) = \dfrac{\ln(u \cdot L)}{\ell} \cdot \dfrac{\ln(v \cdot L)}{\ell}$. Thus, $\ell \cdot (\exp(k \cdot z) - 1) = \dfrac{\ln(u \cdot L) \cdot \ln(v \cdot L)}{\ell} - \ell$, so for

$$C(u,v) = \exp(\ell \cdot (\exp(k \cdot z) - 1)),$$

we get the desired expression.

3.4°. If $g(y) = B \cdot y + a$, then the composition takes the form

$$\psi(x) = g(f(x)) = (B \cdot A) \cdot \exp(k \cdot x) + a.$$

The condition $\psi(\infty) = 0$ leads to $a = 0$, so we get an exponential generator function which, as we have mentioned, leads to the independence copula.

4°. Finally, let us consider the case when $f(x) = A \cdot x + b$.

4.1°. If $g(y) = B \cdot y^a$, then the composition takes the form

$$\psi(x) = g(f(x)) = B \cdot (A \cdot x + b)^a.$$

This generator function leads to the Clayton copulas.

4.2°. If $g(y) = B \cdot \ln(y) + a$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot \ln(A \cdot x + b) + a$. For this function, the condition $\psi(\infty) = 0$ cannot be satisfied.

4.3°. If $g(y) = B \cdot \exp(k \cdot y)$, then the composition takes the form $\psi(x) = g(f(x)) = B \cdot \exp(k \cdot A \cdot \ln(x) + k \cdot b) = (B \cdot \exp(k \cdot b) \cdot x^{k \cdot A}$. This function cannot satisfy the condition $\psi(0) = 1$.

4.4°. If $g(y) = B \cdot y + a$, then the composition $\psi(x) = g(f(x))$ is also a linear function, so we cannot have $\psi(0) = 1$.

The proposition is proven.

Acknowledgments. This work was supported in part by the National Science Foundation grants HRD-0734825, HRD-1242122, and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by a grant on F-transforms from the Office of Naval Research. The authors are greatly thankful to the anonymous referees for valuable suggestions.

References

1. Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (2010)
2. Feynman, R., Leighton, R., Sands, M.: The Feynman Lectures on Physics. Addison Wesley, Boston (2005)
3. Finkelstein, A.M., Kreinovich, V., Zapatrin, R.R.: Fundamental physical equations uniquely determined by their symmetry groups. Springer Lecture Notes in Mathematics, vol. 1214, pp. 159–170. Springer, Heidelberg (1986)
4. Jaworski, P., Durante, F., Härdle, W.K., Ruchlik, T. (eds.): Copula Theory and Its Applications. Springer, Heidelberg (2010)
5. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques, Tools. Princeton University Press, Princeton (2005)
6. Nelsen, R.B.: An Introduction to Copulas. Springer, Heidelberg (1999)
7. Nguyen, H.T., Kreinovich, V.: Applications of Continuous Mathematics to Computer Science. Kluwer, Dordrecht (1997)

Size Distortion in the Analysis of Volatility and Covolatility Effects

Christian Gourieroux and Joann Jasiak

Abstract. Let us assume that $\hat{A}_T$ is a consistent, asymptotically normal estimator of a matrix A (where T is the sample size), this paper shows that test statistics used in empirical work to test 1) the noninvertibility of A, i.e. $detA = 0$, 2) the positivite semi-definiteness $A >> 0$, have a different asymptotic distribution in the case where $A = 0$ than in the case where $A \neq 0$. Moreover, the paper shows that an estimator of A constrained by symmetry or reduced rank has a different asymptotic distribution when $A = 0$ than when $A \neq 0$. The implication is that inference procedures that use critical values equal to appropriate quantiles from the distribution when $A \neq 0$ may be size distorted. The paper points out how the above statistical problems arise in standard models in Finance in the analysis of risk effects.A Monte Carlo study explores how the asymptotic results are reflected in finite sample.

Keywords: Multivariate Volatility, Risk Premium, BEKK Model, Volatility Transmission, Identifiability, Boundary, Invertibility Test.

JEL number: C10, C32, G10, G12.

1 Introduction

In financial models, the risk on a set of assets is commonly represented by a volatility-covolatility matrix, while the risk effect on expected returns and future volatilities is often specified as an affine function of current and lagged realized volatilities and covolatilities. Under some specific hypotheses, the regularity

Christian Gourieroux
CREST, CEPREMAP and University of Toronto
e-mail: gouriero@ensae.fr

Joann Jasiak
York University, Canada
e-mail: jasiakj@yorku.ca

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 91–118.
DOI: 10.1007/978-3-642-35443-4_7

conditions may not hold in this particular framework, and, as a consequence, the limiting distributions of some commonly used estimators and test statistics may differ from the standard ones.

There are two strands of literature that are directly concerned: the literature on risk premium, and the literature on multivariate ARCH models.

Let us first consider a simple risk premium model with 2 assets, called asset 1 and asset 2, and the following volatility matrix

$$\Sigma_t = \begin{pmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{22,t} \end{pmatrix}.$$

The expected return on asset 1 can be written as:

$$E_t(r_{1,t+1}) = r_{f,t+1} + a_1\sigma_{11,t} + 2b_1\sigma_{12,t} + c_1\sigma_{22,t} + a_1^*\sigma_{11,t-1} + 2b_1^*\sigma_{12,t-1} + c_1^*\sigma_{22,t-1},$$

where $r_{f,t}$ is the riskfree return, and coefficients (a_1,b_1,c_1), (a_1^*,b_1^*,c_1^*) are the elements of matrices $A = \begin{pmatrix} a_1 & b_1 \\ b_1 & c_1 \end{pmatrix}$ and A^*, respectively. This model allows us to estimate the ex-ante equity risk premium and to test the statistical significance and positivity of the risk premium. Technically, these two tests concern the significance and sign of matrix A (resp. A^*). Regarding the sign, there exists evidence that suggests that risk premium can be either positive or negative. In particular, Boudoukh et al (1993), Ostdiek (1985), Arnott, Ryan (2001), Arnott, Bernstein (2002), Chen, Guo, Zhang (2006), Walsh (2006) tested the positivity of the conditional risk premium using the method of instrumental variables and showed that risk premium can be of either sign, depending on the environment. The rank of risk premium is also unclear. The theory underlying the CAPM model suggests the existence of a relationship between the expected return and the variance of a single market portfolio that captures the entire effect of variances and covariances of all assets. This would imply that matrix A, in the above risk premium model, is not of full rank.

A similar ambiguity concerning the sign and rank of risk premium arises in foreign exchange markets [see e.g. Domowitz, Hakkio (1985), Macklem (1991) , Hakkio, Sibert (1995)]. The literature suggests that the sign of the foreign exchange real risk premium can vary depending on the ratio of market volatilities in both countries. The significance of risk premium is of economic interest too, as it has an important interpretation in the context of exchange rates. In particular, if the risk premium is zero, the forward exchange rate becomes an unbiased predictor of the future spot exchange rate.

In multivariate ARCH models, the expected future volatility is defined by linear functions of volatility-covolatility (see, e.g. Engle, Granger, Kraft (1984), Bollerslev, Engle, Wooldridge (1988), Bollerslev, Chou, Kroner (1992)). For example, the so-called vech-representation is:

$$V_t(r_{1,t+1}) = d + a_1\tilde{\sigma}_{11,t} + 2b_1\tilde{\sigma}_{12,t} + c_1\tilde{\sigma}_{22,t} + a_1^*\tilde{\sigma}_{11,t-1} + 2b_1^*\tilde{\sigma}_{12,t-1} + c_1^*\tilde{\sigma}_{22,t-1},$$

where $\tilde{\sigma}_{ij,t} = r_{i,t}r_{j,t},\ i,j = 1,2$, and as before, the remaining coefficients are elements of matrices A and A^*.

In this model, it is interesting to test the significance of lagged realized volatility, and the existence of a factor representation of realized volatility, as in the BEKK model (Baba, Engle, Kraft, Kroner (1990)). These tests directly concern the rank and sign of matrix A (resp. A^*), as it was the case in the risk premium model discussed in the previous paragraphs.

In this paper, we assume that matrix A is estimated from a sample of asset returns of size T, and that estimator $\hat{A}_T$ is a consistent, asymptotically normal estimator of A. Our study is focused on the tests of various hypotheses concerning matrix A, mainly for A of dimension 2×2, for clarity of exposition.

The hypotheses of interest discussed in this paper are:

1) the hypothesis of noninvertibility of matrix A;
2) the hypothesis that matrix A is positive semi-definite.

This last hypothesis is equivalent to the hypothesis of nonnegativity of the linear form $Tr(A\Sigma)$ (see Appendix, Lemma 1). Indeed, the linear form in volatilities-covolatilities that appears in the risk premium model and the vech-representation above can be rewritten as:

$$a\sigma_{11} + 2b_1\sigma_{12} + c\sigma_{22} = Tr\left[\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right] = Tr(A\Sigma), \text{ say},$$

where Tr is the trace operator.

Moreover, we will also investigate
3) constrained estimation of A
when the constraint implies that the rank of A is less or equal to 1.

At a first sight, the above hypotheses tests and constrained estimation [1] seem quite standard. Indeed, the invertibility of matrix A (hypothesis 1) is usually tested from the singular value decomposition of the (asymptotically) Gaussian random matrix $\hat{A}_T$ [see Anderson (1989), Gourieroux, Monfort, Renault (1995), Bilodeau, Brenner (1999)]. As for hypothesis 2), the tests of matrix positivity are based on asymptotic tests of the following inequality restrictions $ac - b^2 \geq 0$, $a \geq 0$ [see e.g. Gourieroux, Monfort (1989), Wolak (1991)]. Finally, estimation of A under the hypothesis of reduced rank is commonly performed by a quasi-maximum likelihood method, as in the BEKK model [Engle, Kroner (1995), Jeantheau (1998), Comte, Lieberman (2000)].

The purpose of this paper is to point out the identifiability problems, boundary problems and degeneracies that may be encountered while performing the aforementioned tests and estimation in the framework of risk premium and multivariate ARCH models. In the presence of these effects, the asymptotic distributions of estimators and test statistics can be non-standard. This can render the outcomes of standard inference misleading and the stylized facts questionable. The degeneracies

[1] Similar problems arise in the so-called vech-diagonal multivariate ARCH models, such as $\sigma_{ij,t} = d_{ij} + a_{ij}\tilde{\sigma}_{ij,t}$, i,j=1,2, $i \leq j$. It is easy to check that the expected volatility-covolatility matrix is positive semi-definite, if and only if, the matrix $A = (a_{ij})$ is positive semi-definite. This condition is sufficient only for matrices of larger dimension (Silberberg, Pafka (2001)).

discussed in the paper concern some commonly used estimators and test statistics for which the true asymptotic distributions are derived. In particular, the asymptotic admissibility of the test statistics and their potential improvements are out of the scope of the present paper.

The paper is organized as follows. Section 2 considers the Wald test of non-invertibility of matrix A based on the estimated determinant $\det \hat{A}_T$. It shows that in the degenerate case $A = 0$ the Wald test statistic has a non-Gaussian distribution and that this distribution depends on the asymptotic variance of random matrix $\hat{A}_T$. Section 3 discusses the constrained estimation of A when it is not of full rank. We point out that when $A = 0$, the distribution of the constrained estimator is non-standard. Section 4 considers the test of positive semi-definiteness, that is, of the hypothesis defined by inequality constraints $a \geq 0, c \geq 0, ac - b^2 \geq 0$. We show that when $A = 0$, the standard asymptotic theory is no longer valid. The necessary adjustments are given for an unconstrained A [respectively, for A of reduced rank] under the maintained hypothesis. Finite sample properties of the standard test statistics in the degenerate case are presented in Section 5. Section 6 concludes.

2 Invertibility Tests

Let us consider the test of invertibility of matrix A, based on the significance of its determinant. The null hypothesis is:

$$H_0 : \ (\det A = 0). \tag{1}$$

2.1 The Unconstrained Model

Let us consider a $n \times n$ matrix of parameters A, and its consistent, asymptotically Gaussian estimator $\hat{A}_T$. $vecA$ denotes a vector of length n^2 obtained by stacking the columns of matrix A. We assume that:

$$\sqrt{T}[vec(\hat{A}_T) - vec(A)] \xrightarrow{d} N(0, \Omega), \tag{2}$$

where Ω is a $(n^2 \times n^2)$ invertible matrix and $\xrightarrow{d}$ denotes the convergence in distribution. $\hat{A}_T$ conveys all relevant information about A contained in the data. From now on, model (2) is referred to as the unconstrained (asymptotic) model.

2.2 Wald Test Statistic

A standard method for testing the null hypothesis H_0 in (1) is based on the estimated determinant $det\hat{A}_T$ and its asymptotic distribution obtained by applying the δ-method.

Since $\frac{\partial(\det A)}{\partial(vecA)} = vec[cof(A)]$, where $cof(A)$ is the (n $\times n$) matrix whose elements are the cofactors[2] of elements of A, we get:

$$\sqrt{T}(det\hat{A}_T - detA) \xrightarrow{d} N(0, vec[cof(A)]'\Omega\, vec[cof(A)]). \tag{3}$$

The Wald test statistic for testing the null hypothesis (1) is:

$$\hat{\xi}_T = \frac{\sqrt{T}det\hat{A}_T}{[vec[cof(\hat{A}_T)]'\hat{\Omega}_T\, vec[cof(\hat{A}_T)]]^{1/2}}, \tag{4}$$

where $\hat{\Omega}_T$ is a consistent estimator of Ω. If $vec[cof(A)] \neq 0$, this Wald statistic follows asymptotically a standard normal distribution and a critical region of the type $\{|\hat{\xi}_T| > 1.96\}$ defines a test at asymptotic level 5%.

2.3 The Degenerate Case

The standard asymptotic properties of the test are valid as long as $vec[cof(A)] \neq 0$, that is, if $A \neq 0$. Otherwise, the asymptotic properties of the Wald test statistic are significantly altered.
i) Asymptotic Properties of the Estimated Determinant

When $A = 0$, we have $\sqrt{T}vec(\hat{A}_T) \xrightarrow{d} vec(A_\infty) \sim N(0, \Omega)$, say. Thus we have: $det(\sqrt{T}\hat{A}_T) \xrightarrow{d} det(A_\infty)$, or equivalently

$$T^{n/2}det\hat{A}_T \xrightarrow{d} det(A_\infty). \tag{5}$$

When $n \geq 2$, the asymptotic behavior differs from the standard behavior, since:

i) the speed of convergence is $1/(T^{n/2})$ instead of $1/\sqrt{T}$, that is greater;
ii) the limiting distribution is not Gaussian, but instead, it is a determinant transformation of a multivariate Gaussian distribution.
ii) Asymptotic Properties of the Wald Test Statistic

Similarly, we can examine the test statistic $\hat{\xi}_T$ when A=0. Since $cof(\sqrt{T}\hat{A}_T) = T^{(n-1)/2}cof(\hat{A}_T)$, we see that $\hat{\xi}_T \xrightarrow{d} \xi(A_\infty)$, where

$$\xi(A_\infty) = \frac{det(A_\infty)}{\{vec[cof(A_\infty)]'\Omega\, vec[cof(A_\infty)]\}^{1/2}}. \tag{6}$$

iii) Comparison with the Literature
The degenerate case considered here does not belong to those discussed in Andrews (2001), in which some parameters are not identifiable under the null. In our framework, matrix A is always identifiable. This explains why the asymptotic distribution of the Wald statistic differs from the distribution derived by Andrews (2001).

This degeneracy cannot be disregarded or circumvented, for example, by introducing a sequence of null hypotheses indexed by the number T of observations,

[2] A cofactor is a determinant obtained by deleting the row and column of a given element of a matrix preceded by a + or − sign depending whether the element is in a + or − position.

such as $H_{0,T}: [detA = 0, ||A|| > h(T)]$, where $||A||^2$ denotes the largest eigenvalue of AA' and $h(T)$ is strictly positive and tends to zero at an appropriate rate, when T tends to infinity [3]. Indeed, the hypothesis $H_0 : \{A = 0\}$ does not belong in the union of this sequence of hypothesis $H_{0,T}$, and hypothesis H_0 has often structural interpretations whereas the sequence $H_{0,T}$ does not. For instance, the test of H_0 allows for determining the autoregressive order of a multivariate ARCH model [4]. In the application to risk premium, the condition $A = 0$ characterizes the hypothesis of nonpredictability of asset returns that is of economic interest.

2.4 Critical Values

1) Asymptotic Size of the Test

The multiplicity of limiting distributions of the Wald test statistic under the null hypothesis suggests that a detailed analysis of the type I error is needed, as the condition of asymptotic similarity on the boundary condition is violated [see Hansen (2003)]. For instance, suppose that the null hypothesis is rejected when the Wald statistic $\hat{\xi}_T$ is larger in absolute value than the critical value c, and let us denote $\mathscr{A}_I$ the set of noninvertible matrices A.

The size of the test for a finite sample of length T is equal to:

$$\alpha_T(c) = sup_{A\in\mathscr{A}_I} P_A(\hat{\xi}_T > c),$$

and is reached for matrix A_T^* in $\mathscr{A}_I$.

Then, it is possible to define

(*) the asymptotic null rejection probability as:

$$\alpha_\infty(c) = sup_{A\in\mathscr{A}_I} \lim_{T\to\infty} P_A(\hat{\xi}_T > c);$$

(**) the asymptotic size of the test as:

$$\tilde{\alpha}_\infty(c) = \lim_{T\to\infty} \alpha_T(c) = \lim_{T\to\infty} sup_{A\in\mathscr{A}_I} P_A(\hat{\xi}_T > c).$$

In the sequel, we assume that $\lim_{T\to\infty}$ and $sup_{A\in\mathscr{A}_I}$ can commute, which implies a "uniform convergence" condition of the finite sample distribution of $\hat{A}_T$ towards its asymptotic Gaussian distribution.

Assumption A.1

The asymptotic size of the test is equal to the asymptotic null rejection probability.

In the applications, Assumption A.1. has to be verified case by case according to the type of asymptotically Gaussian estimator $\hat{A}_T$ which is used.

---

[3] Such a methodology is followed in the test of switching regimes, for the parameter representing the unknown switching date [Andrews (1993)].

[4] See Andrews (2001), Francq, Zakoian (2006) for tests concerning the orders of univariate GARCH processes.

Under Assumption A.1, the asymptotic size of the test is:

$$\begin{aligned}\alpha_\infty(c) &= sup_{H_0} \lim_{T\to\infty} P_A[|\hat{\xi}_T| > c] \\ &= sup[sup_{A:detA=0,A\neq 0} \lim_{T\to\infty} P_A(|\hat{\xi}_T| > c), sup_{A=0} \lim_{T\to\infty} P_A(|\hat{\xi}_T| > c)] \\ &= sup[(P(|X| > c), P(|\xi(A_\infty)| > c)] \text{ (where } X \sim N(0,1)).\end{aligned}$$

By inverting this relationship, we deduce the critical value for an asymptotic size $\alpha_\infty = \alpha$:

$$c(\alpha) = Max[\Phi^{-1}(1-\alpha/2), Q(\alpha,\Omega)],$$

where Φ is the cdf of the standard normal, and $Q(\alpha,\Omega)$ is the quantile computed from:

$$P[|\xi(A_\infty)| > Q(\alpha,\Omega)] = \alpha, \tag{7}$$

where $vec(A_\infty) \sim N(0,\Omega)$. Function Q is too complicated to be calculated analytically, but the value $Q(\alpha,\Omega)$ can be easily approximated by Monte-Carlo simulations. Let us denote by $\hat{\Omega}_{0T}$ an estimator of Ω, which is consistent under the null and by $\hat{Q}(\alpha,\hat{\Omega}_{0T})$ the associated value of Q derived by simulations. The critical value will be chosen as $\hat{c}(\alpha) = Max[\Phi^{-1}(1-\alpha/2), \hat{Q}(\alpha,\hat{\Omega}_{0T})]$.

ii) Comparison with Sequential Procedures

Under Assumption A.1, the procedure above provides the correct asymptotic size of the test of the null hypothesis $H_0 : \{\det A = 0\}$. It is an alternative to the sequential procedures described below, which are asymptotically size distorted.

i) A two-step procedure can be as follows. In the first step, we consider a Fisher statistic F for testing the hypothesis $H_0^* : \{A = 0\}$ with critical value f_{α_0}, say, corresponding to level α_0. If $F < f_{\alpha_0}$, the null hypothesis H_0 is accepted. Otherwise, in the second step we perform a test based on the determinant at level α_1, and accept H_0, if $\xi_T < \Phi^{-1}(1-\alpha_1/2)$. The critical region of the sequential test is:

$$W = \{F > f_{\alpha_0}, \xi_T > \Phi^{-1}(1-\alpha_1/2)\}.$$

For a given choice of α_0, α_1, the asymptotic size[5] of this test is equal to

$$Sup_{A:det(A)=0} \lim_{T\to\infty} P[W > f_{\alpha_0}, \xi_T > \Phi^{-1}(1-\alpha_1/2)].$$

The asymptotic size can be bounded by a known function of α_0, α_1, but depends on α_0, α_1 and Ω, in general. Thus, this sequential test can be asymptotically size distorted.

ii) Another sequential procedure can be based on the analysis of the rank of matrix A [see e.g. Anderson (1989), Gill, Lewbel (1992), Cragg, Donald (1996), (1997), Bilodeau, Bremer (1999), Robin, Smith (2000)]. Indeed, for a matrix of dimension $(n \times n)$, we know that

[5] Assumed equal to the asymptotic null rejection probability.

$$H_0 : \{detA = 0\} = \{rank(A) = 0\} \cup \{rank(A) = 1\} \cup \ldots \{rank(A) = n-1\}$$
$$= \{rank(A) < n\}.$$

Thus, we can first test if $rankA = 0$; then, if this hypothesis is rejected, we test if $rankA = 1$, etc. As above, the asymptotic size of this sequential test can be easily bounded, but its exact value is difficult to derive. The interpretation in terms of rank shows that a) a reason for the degenerate asymptotic behavior of statistic $\hat{\xi}_T$ is that the null hypothesis H_0 is a union of elementary null hypotheses $\{rank(A) = p\}$; b) the only elementary hypothesis that causes the degeneracy is $\{rank(A) = 0\}$, whereas the other elementary hypotheses $\{rank(A) = p\}$, $p = 1,...,n-1$ have been jointly accommodated in the single statistic $\hat{\xi}_T$.

2.5 Symmetric Matrix A of Dimension (2,2)

Let us consider the estimator of a square (2,2) symmetric matrix $\hat{A}_T = \begin{pmatrix} \hat{a}_T & \hat{b}_T \\ \hat{b}_T & \hat{c}_T \end{pmatrix}$ and its convergence limit $A_\infty = \begin{pmatrix} a_\infty & b_\infty \\ b_\infty & c_\infty \end{pmatrix}$. The aim of this section is to derive the asymptotic critical values of the Wald test for any possible matrix Ω. Matrix Ω contains 6 different elements. First, we show that the critical values depend on Ω by only three parameters and the sign. This finding allow us to simplify the display of critical values.

The test statistic $\xi(A_\infty)$ is such that $\xi(PA_\infty P') = \xi(A_\infty)$, for any matrix P of dimension $(n \times n)$ (see Proposition A.1, ii) in Appendix). We infer that the quantiles $Q(\alpha, \Omega)$ and $Q(\alpha, \Omega(P))$ are identical if $\Omega = V[vec(A_\infty)]$ and $\Omega(P) = V[vec(PA_\infty P')]$, for any P. By choosing an appropriate linear transformation P, we show in Appendix, b) that the quantiles $Q(\alpha, \Omega)$, $\forall \Omega$, depend in fact, on a number of parameters much smaller than the number of elements in Ω.

Proposition 1: Up to a transformation $A_\infty \longrightarrow PA_\infty P'$, matrix Ω can be defined as:

$$\Omega = Var \begin{pmatrix} a_\infty \\ b_\infty \\ c_\infty \end{pmatrix} = \begin{pmatrix} 1 & 0 & \varepsilon\rho^2 \\ 0 & \gamma^2 & 0 \\ \varepsilon\rho^2 & 0 & 1 \end{pmatrix},$$

where parameters ρ and γ are nonnegative, $\rho < 1$, and ε is equal to +1 or -1, according to the sign of correlation between a_∞ and c_∞.

Thus, the set of admissible quantiles $\{Q(\alpha, \Omega), \Omega >> 0\}$, where $\Omega >> 0$ means that the matrix is symmetric, positive semi-definite, coincides with the set of quantiles $\{Q[\alpha, \Omega(\varepsilon, \rho, \gamma)], \varepsilon = \pm 1, 0 < \rho < 1, \gamma > 0\}$.

The Wald test statistic with $\Omega(\varepsilon, \rho, \gamma)$ is:

$$\xi(A_\infty) = \frac{a_\infty c_\infty - b_\infty^2}{\sqrt{(c_\infty, -2b_\infty, a_\infty)\Omega(\varepsilon, \rho, \gamma)(c_\infty, -2b_\infty, a_\infty)'}}$$
$$= \frac{a_\infty c_\infty - b_\infty^2}{\sqrt{c_\infty^2 + a_\infty^2 + 2\varepsilon\rho^2 c_\infty a_\infty + 4b_\infty^2\gamma^2}}. \quad (8)$$

Table 1 provides the upper quantiles at 10%, 5% and 1% of the variable $|\xi(A_\infty)|$ for different values of parameters ρ, γ and $\varepsilon = +/-1$. The quantiles are obtained from Monte-Carlo experiments with 5000 replications. They can be directly compared to the critical values 1.64, 1.96, 2.57 of the standard normal distribution, which correspond to the case when $detA = 0$ with $A \neq 0$. We observe that all these values are smaller than their Gaussian counterparts. This implies that, under Assumption A.1. for a (2,2) symmetric matrix A the asymptotic size of the standard Wald test does not need to be corrected for degeneracy $A = 0$, but likely, the magnitude of size distortion depends on the dimension of matrix A. Moreover, we show that such a size correction is needed for other inference on matrix A.

3 Constrained Estimation of A

3.1 The Example of BEKK Model

To ensure the positivity of volatility matrix $H_t = V_t(r_{t+1})$, the multivariate GARCH literature (Engle, Kroner (1995)) proposed the following constrained specification[6]:

$$H_t = C_0 + \sum_{j=1}^{p} M_j H_{t-j} M_j' + \sum_{k=1}^{q} N_k r_{t-k} r_{t-k}' N_k', \text{ say},$$

where M_j, N_k, C_0 are (n,n) matrices and $C_0 >> 0$. Accordingly, the volatility of asset i is:

$$h_{iit} = c_{0,ii} + \sum_{j=1}^{p} M_{ij} H_{t-j} M_{ji}' + \sum_{k=1}^{q} N_{ik} r_{t-k} r_{t-k}' N_{ki}',$$

where M_{ij} (resp. N_{ik}) is the i^{th} row of M_j (resp. N_k). A component of the first sum on the right-hand side is of the form:

$$M_i H M_i' = Tr(M_i H M_i') = Tr(M_i' M_i H) = Tr(A_i H), \text{ say},$$

where $A_i = M_i' M_i$ is of rank less or equal to 1.

Under a BEKK specification, the estimation of matrix A_i has to be performed under constraints. A common approach consists in optimizing a quasi-likelihood function with respect to parameters M (and N) [see e.g. Engle, Kroner (1995), Comte, Lieberman (2003), Iglesias, Phillips (2005)]. Let us consider matrix A of dimension two[7], $A = \begin{pmatrix} m_1^2 & m_1 m_2 \\ m_1 m_2 & m_2^2 \end{pmatrix}$. Due to a lack of identifiability of parameter M, the following two difficulties arise:

[6] For ease of exposition, we introduced only 1 positive component per lag.

[7] The results can be easily extended to matrix A of dimension $(n \times n)$ and of rank less or equal to 1.

Table 1 Critical Values of the Wald Test Statistic for Positive and Negative ε

ρ	γ	ε Positive			ε Negative		
		10	5	1	10	5	1
0.0	0.5	0.945	1.092	1.428	0.903	1.074	1.399
0.0	1.0	0.989	1.141	1.472	0.922	1.092	1.491
0.0	1.5	0.935	1.074	1.359	0.898	1.048	1.423
0.0	2.0	0.896	1.044	1.336	0.871	1.014	1.337
0.1	0.5	0.942	1.081	1.412	0.930	1.064	1.393
0.1	1.0	0.980	1.135	1.465	0.927	1.096	1.457
0.1	1.5	0.936	1.083	1.356	0.894	1.046	1.424
0.1	2.0	0.894	1.046	1.318	0.869	1.015	1.326
0.2	0.5	0.929	1.098	1.411	0.941	1.066	1.384
0.2	1.0	0.975	1.125	1.420	0.946	1.089	1.480
0.2	1.5	0.930	1.076	1.335	0.905	1.037	1.416
0.2	2.0	0.890	1.046	1.309	0.868	1.016	1.331
0.3	0.5	0.928	1.054	1.391	0.930	1.074	1.361
0.3	1.0	0.955	1.119	1.404	0.959	1.102	1.470
0.3	1.5	0.923	1.064	1.306	0.919	1.044	1.396
0.3	2.0	0.896	1.038	1.294	0.876	1.002	1.336
0.4	0.5	0.915	1.062	1.407	0.926	1.079	1.323
0.4	1.0	0.939	1.104	1.380	0.970	1.114	1.441
0.4	1.5	0.904	1.063	1.270	0.921	1.052	1.391
0.4	2.0	0.889	1.026	1.287	0.878	1.015	1.340
0.5	0.5	0.910	1.030	1.367	0.927	1.075	1.354
0.5	1.0	0.930	1.075	1.339	0.991	1.125	1.437
0.5	1.5	0.898	1.049	1.272	0.937	1.058	1.380
0.5	2.0	0.866	1.027	1.268	0.880	1.017	1.340
0.6	0.5	0.899	1.047	1.337	0.934	1.095	1.360
0.6	1.0	0.923	1.058	1.300	1.008	1.147	1.440
0.6	1.5	0.889	1.042	1.259	0.951	1.082	1.358
0.6	2.0	0.858	1.018	1.260	0.888	1.038	1.329
0.7	0.5	0.878	1.052	1.281	0.936	1.108	1.353
0.7	1.0	0.885	1.020	1.276	1.019	1.168	1.414
0.7	1.5	0.857	1.028	1.262	0.965	1.106	1.377
0.7	2.0	0.850	1.005	1.263	0.906	1.057	1.318
0.8	0.5	0.866	1.041	1.255	0.939	1.101	1.372
0.8	1.0	0.859	1.008	1.228	1.034	1.196	1.424
0.8	1.5	0.840	1.020	1.268	0.986	1.121	1.387
0.8	0.2	0.833	0.996	1.265	0.923	1.072	1.333
0.9	0.5	0.841	0.970	1.238	0.941	1.094	1.389
0.9	1.0	0.8433	0.992	1.249	1.049	1.189	1.515
0.9	1.5	0.837	0.995	1.268	0.999	1.155	1.415
0.9	2.0	0.833	0.995	1.265	0.938	1.072	1.327

i) First, there is a problem of global identifiability since the same matrix A is obtained for M and $-M$. To solve this problem, it is common to use the following change of parameters:

$$A = m_1^2 \begin{pmatrix} 1 & m_2/m_1 \\ m_2/m_1 & (m_2/m_1)^2 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ \beta \end{pmatrix} (1,\beta), \text{ say,} \tag{9}$$

where $\alpha = m_1^2 \geq 0, \beta = m_2/m_1$ (whenever $m_1 \neq 0$, or equivalently $\alpha \neq 0$).

ii) Second, there is a problem of local identifiability at $A = 0$. The reason is that the Jacobian

$$\frac{\partial\, vechA}{\partial(m_1,m_2)'} = \begin{pmatrix} 2m_1 & 0 \\ m_2 & m_1 \\ 0 & 2m_2 \end{pmatrix}$$

is of rank 2, except when $A = 0$.

The asymptotic theory established for multivariate BEKK models doesn't hold for the estimators of parameters α and β defined in (9), because it assumes the identifiability of parameter M [see Assumption A.4 in Comte, Lieberman (2003)]. To overcome this difficulty Engle, Kroner (1995) (Proposition 2.1) introduce the identifiability condition $m_1 > 0$. This condition eliminates both the global and local identifiability problems.

In the next section, we derive the true asymptotic distributions of the minimum distance estimators of α and β based on a consistent, and asymptotically normal estimator of A. For the application to BEKK model, we assume that the unconstrained quasi-maximum likelihood estimator of A is asymptotically normal. This, in turn, requires some additional assumptions on the BEKK model, such as the presence of at least one non-zero ARCH effect [$N_{ki} \neq 0$ for at least one index k] to avoid another degeneracy pointed out in Andrews (2001).

3.2 The Constrained Estimator

Let us now assume that matrix A is symmetric and of reduced rank. Then we can write $A = \alpha \begin{pmatrix} 1 \\ \beta \end{pmatrix} (1,\beta)$, where α and β are unconstrained [8].

The constrained estimator of A based on $\hat{A}_T$ is the solution of the following minimization:

$$(\hat{\alpha}_T, \hat{\beta}_T) = \arg min_{\alpha,\beta} (\hat{a}_T - \alpha, \hat{b}_T - \alpha\beta, \hat{c}_T - \alpha\beta^2) \tilde{\Omega}_T^{-1} \begin{pmatrix} \hat{a}_T - \alpha \\ \hat{b}_T - \alpha\beta \\ \hat{c}_T - \alpha\beta^2 \end{pmatrix}. \tag{10}$$

The objective function (10) is defined for all values of parameters α, β. However, the stochastic coefficients involved in the objective function cannot be normalized uniformly with respect to the true matrix A. Therefore, Assumption 3 in Andrews (1999), p. 1349, is not satisfied and new asymptotic results need to be derived.
The objective function can be concentrated with respect to α. Then, the solution in α for a given β is:

[8] We do not assume a priori that A is positive semi-definite.

$$\alpha(\beta) = < vech\hat{A}_T, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} > / < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >,$$

where $<,>$ denotes the inner product associated with $\tilde{\Omega}_T^{-1}$.

The concentrated objective function is:

$$\Psi_T(\beta) = < vech\hat{A}_T, vech\hat{A}_T > - \left[(< vech\hat{A}_T, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >) \right]^2 / < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >. \quad (11)$$

The optimization of the concentrated objective function yields a finite solution (see Appendix).

Since the first-order condition is:

$$< vech\hat{A}_T, \begin{pmatrix} 0 \\ 1 \\ 2\beta \end{pmatrix} >< \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} > - < vech\hat{A}_T, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >< \begin{pmatrix} 0 \\ 1 \\ 2\beta \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} > = 0, \quad (12)$$

the solution that minimizes (11) is a root of a polynomial of degree 5.

3.3 Asymptotic Distribution of the Constrained Estimator

When A is not equal to zero (i.e. if $\alpha \neq 0$), the standard asymptotic theory holds and we have:

$$\sqrt{T} \left[\begin{pmatrix} \hat{\alpha}_T \\ \hat{\beta}_T \end{pmatrix} - \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right] \xrightarrow{d} N[0, (J(\alpha,\beta)\tilde{\Omega}^{-1} J(\alpha,\beta)')^{-1}],$$

where the Jacobian matrix is $J(\alpha,\beta) = \begin{pmatrix} 1 & \beta & \beta^2 \\ 0 & \alpha & 2\alpha\beta \end{pmatrix}$.

When $A = 0$, the Jacobian matrix is of rank 1, and the standard asymptotic theory is no longer valid. Let us now consider this case. It follows from (12) that $\hat{\beta}_T$ is a solution of

$$Max_\beta \left[< vech\hat{A}_T, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} > \right]^2 / < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >$$

$$\iff Max_\beta \left[< vech(\sqrt{T}\hat{A}_T), \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} > \right]^2 / < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >.$$

As a consequence, $\hat{\beta}_T$ tends to a limit β_∞, which is a solution to the optimization:

$$Max_{\beta}\left[< vech(A_{\infty}), \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >\right]^2 / < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} > . \tag{13}$$

Similarly, we note that:

$$\begin{aligned} \sqrt{T}\hat{\alpha}_T &= \sqrt{T}\alpha(\hat{\beta}_T) \\ &= < vech(\sqrt{T}\hat{A}_T), \begin{pmatrix} 1 \\ \hat{\beta}_T \\ \hat{\beta}_T^2 \end{pmatrix} > / < \begin{pmatrix} 1 \\ \hat{\beta}_T \\ \hat{\beta}_T^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \hat{\beta}_T \\ \hat{\beta}_T^2 \end{pmatrix} > \end{aligned}$$

tends to a limit

$$\alpha_{\infty} =< vech(A_{\infty}), \begin{pmatrix} 1 \\ \beta_{\infty} \\ \beta_{\infty}^2 \end{pmatrix} > / < \begin{pmatrix} 1 \\ \beta_{\infty} \\ \beta_{\infty}^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta_{\infty} \\ \beta_{\infty}^2 \end{pmatrix} > . \tag{14}$$

Proposition 3 summarizes the above discussion.

Proposition 3

If $A = 0$, then $(\sqrt{T}\hat{\alpha}_T, \hat{\beta}_T) \xrightarrow{d} (\alpha_{\infty}, \beta_{\infty})$, where $(\alpha_{\infty}, \beta_{\infty})$ is a complicated nonlinear transformation of the Gaussian vector, derived from (13), (14).

Note that parameter β is not identifiable when $A = 0$. Nevertheless its estimator $\hat{\beta}_T$ admits a limiting distribution.

The asymptotic limiting distributions of test statistics for α and β are non-standard too. For instance, the t-statistic for the test of significance of parameter α is:

$$\hat{\eta}_T^{\alpha} = \sqrt{T}\hat{\alpha}_T / \hat{\sigma}_{\alpha,T},$$

where $\hat{\sigma}_{\alpha,T}$ is the square root of the first diagonal element of the matrix

$$[J(\hat{\alpha}_T, \hat{\beta}_T)\tilde{\Omega}_T^{-1}J(\hat{\alpha}_T, \hat{\beta}_T)']^{-1}.$$

When $A \neq 0$, this statistic tends in distribution to a standard normal. When $A = 0$, statistic $\hat{\eta}_T^{\alpha}$ tends to:

$$\eta_{\infty}^{\alpha} =< vech(A_{\infty}), \begin{pmatrix} 1 \\ \beta_{\infty} \\ \beta_{\infty}^2 \end{pmatrix} > / \sigma_{\alpha,\infty}, \tag{15}$$

where $\sigma_{\alpha,\infty}$ is the square root of the first diagonal element of the random matrix $\Sigma_{\infty} = [J(\alpha_{\infty}, \beta_{\infty})\tilde{\Omega}^{-1}J(\alpha_{\infty}, \beta_{\infty})']^{-1}$.

Similarly, the t-statistic for the test of significance of parameter β,

$$\hat{\eta}_T^{\beta} = \sqrt{T}\hat{\beta}_T / \hat{\sigma}_{\beta,T}$$

Table 2 Upper Quantiles of the Student Statistic for α and β

ρ	γ	$\eta_\alpha(10\%)$	$\eta_\alpha(5\%)$	$\eta_\alpha(1\%)$	$\eta_\beta(10\%)$	$\eta_\beta(5\%)$	$\eta_\beta(1\%)$
0.000	0.500	1.523	1.880	2.535	0.997	1.175	1.587
0.000	1.000	1.621	1.963	2.614	1.258	1.500	2.069
0.000	1.500	1.656	1.979	2.625	1.318	1.610	2.221
0.000	2.000	1.657	1.973	2.611	1.331	1.630	2.276
0.100	0.500	1.535	1.866	2.533	1.005	1.183	1.604
0.100	1.000	1.625	1.965	2.593	1.270	1.527	2.064
0.100	1.500	1.650	1.971	2.607	1.327	1.608	2.235
0.100	2.000	1.658	1.966	2.609	1.333	1.635	2.273
0.200	0.500	1.529	1.857	2.574	1.021	1.202	1.639
0.200	1.000	1.631	1.965	2.634	1.300	1.561	2.080
0.200	1.500	1.645	1.982	2.635	1.374	1.666	2.247
0.200	2.000	1.651	1.974	2.621	1.380	1.691	2.315
0.300	0.500	1.538	1.855	2.562	1.054	1.237	1.668
0.300	1.000	1.640	1.959	2.614	1.335	1.622	2.171
0.300	1.500	1.669	1.998	2.631	1.432	1.748	2.360
0.300	2.000	1.664	1.976	2.612	1.455	1.772	2.398
0.400	0.500	1.540	1.870	2.567	1.094	1.284	1.740
0.400	1.000	1.667	1.987	2.683	1.416	1.737	2.319
0.400	1.500	1.691	2.017	2.673	1.539	1.877	2.477
0.400	2.000	1.691	2.010	2.668	1.566	1.939	2.599
0.500	0.500	1.555	1.866	2.548	1.159	1.359	1.843
0.500	1.000	1.670	1.994	2.648	1.537	1.869	2.522
0.500	1.500	1.710	2.025	2.697	1.681	2.047	2.732
0.500	2.000	1.710	2.020	2.673	1.722	2.135	2.882
0.600	0.500	1.551	1.851	2.505	1.248	1.469	1.998
0.600	1.000	1.697	2.006	2.669	1.727	2.085	2.869
0.600	1.500	1.743	2.043	2.721	1.920	2.300	3.131
0.600	2.000	1.724	2.017	2.687	1.982	2.426	3.290
0.700	0.500	1.565	1.901	2.503	1.392	1.641	2.254
0.700	1.000	1.714	2.032	2.691	1.974	2.374	3.297
0.700	1.500	1.741	2.065	2.728	2.252	2.701	3.682
0.700	2.000	1.723	2.054	2.688	2.350	2.855	3.793
0.800	0.500	1.573	1.885	2.516	1.639	1.947	2.670
0.800	1.000	1.726	2.037	2.660	2.444	2.935	4.053
0.800	1.500	1.737	2.074	2.694	2.827	3.359	4.622
0.800	2.000	1.730	2.054	2.700	2.956	3.600	4.834
0.900	0.500	1.607	1.904	2.556	2.281	2.697	3.666
0.900	1.000	1.729	2.049	2.668	3.524	4.200	5.655
0.900	1.500	1.731	2.028	2.694	4.044	4.843	6.567
0.900	2.000	1.739	2.008	2.645	4.291	5.152	6.917

tends to

$$\eta_{\infty}^{\beta} = \beta_{\infty}/\sigma_{\beta,\infty}, \tag{16}$$

where $\sigma_{\beta,\infty}$ is the square root of the second diagonal element of Σ_{∞}.
Table 2 presents the quantiles at 10%, 5%, 1% of the distribution of variables $|\eta_{\infty}^{\alpha}|$ and $|\eta_{\infty}^{\beta}|$, respectively, calculated for Gaussian matrices introduced in Section 2. The quantiles have been obtained by simulations with 5000 replications.

The quantiles associated with the t-statistic for α are less sensitive to parameters ρ and γ than the quantiles associated with the t-statistics for β. Both sets of quantiles are much more sensitive to parameter γ than to other parameters. Moreover, the quantiles differ significantly from the Gaussian quantiles 1.64, 1.96, 2.57, especially for parameter β. In particular, the critical values exceed significantly the critical values from the standard normal distribution.

Figure 1 shows the distribution of β_{∞} for $\rho = 0, \gamma = 1$. For $\rho = 0, \gamma = 1$, β_{∞} is the solution of $Max_{\beta}(a_{\infty} + b_{\infty}\beta + c_{\infty}\beta^2)^2/(1+\beta^2+\beta^4)$, where $a_{\infty}, b_{\infty}, c_{\infty}$ are independent standard normal. Since

$$\beta_{\infty}(-a_{\infty}, -b_{\infty}, -c_{\infty}) = \beta_{\infty}(a_{\infty}, b_{\infty}, c_{\infty}),$$
$$\beta_{\infty}(c_{\infty}, b_{\infty}, a_{\infty}) = 1/\beta_{\infty}(a_{\infty}, b_{\infty}, c_{\infty}),$$

the distribution of β_{∞} is symmetric and invariant with respect to transformation $\beta_{\infty} \to 1/\beta_{\infty}$. This explains the shape of the distribution displayed in Figure 1, with a mode at 0 and very heavy tails.

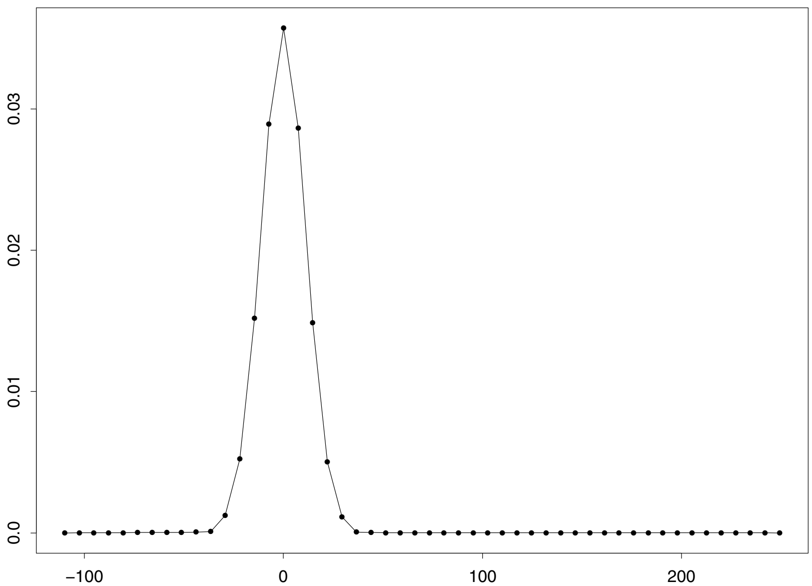


Fig. 1 Distribution of β_{∞}

4 Positivity Test

Let us now focus on the test of positivity for a symmetric matrix A. This test depends on the maintained hypothesis, that is, on whether we assume "A unconstrained", or "A of reduced rank". Both cases are discussed below.

4.1 A Unconstrained

A common approach to testing matrix positivity is as follows. The null hypothesis is written as $H_0 : \{a \geq 0, c \geq 0, ac - b^2 \geq 0\}$, and the test of these inequality restrictions is performed along the lines developped [9] by [Gourieroux, Holly, Monfort (1980), (1982), Kodde, Palm (1986), Gourieroux, Monfort (1989), Wolak (1991)]. However, in the presence of a degeneracy due to $A = 0$, this standard technique cannot be applied. The reason is that it requires the Jacobian of the constraints, that is, $(a,b,c) \to (a, ac - b^2)$ to be of full rank on the boundaries of the null hypothesis. For $A = 0$, however, the Jacobian $\begin{pmatrix} 1 & 0 & 0 \\ c & -2b & a \end{pmatrix}$ is of reduced rank.

Intuitively, the degeneracy can be explained as follows. The positivity condition involves three restrictions and the null hypothesis should be written as $H_0 : \{a \geq 0, c \geq 0, ac - b^2 \geq 0\}$. If either a (resp. c) is strictly positive, than condition $ac - b^2 \geq 0$ implies that c (resp. a) is nonnegative. Thus, one of the two first inequalities seems to be redundant. In fact, this is not the case. For instance, the restrictions $a \geq 0, ac - b^2 \geq 0$ are satisfied for $A = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$, which is not positive semi-definite.

Let us now consider the asymptotic properties of the likelihood ratio test. The log-likelihood function of the (asymptotic) unconstrained model is :

$$L_T(A) = T[-\log 2\pi - \frac{1}{2} \log det \tilde{\Omega}_T - \frac{1}{2} vech(\hat{A}_T - A)' \tilde{\Omega}_T^{-1} vech(\hat{A}_T - A)], \quad (17)$$

where *vech* denotes the vec-half operator. The likelihood ratio statistic for testing the positivity hypothesis is:

$$\begin{aligned} \xi_T^P &= 2(Max_A L_T(A) - Max_{A:A>>0} L_T(A)) \\ &= Min_{A:A>>0} T vech(\hat{A}_T - A)' \tilde{\Omega}_T^{-1} vech(\hat{A}_T - A). \end{aligned} \quad (18)$$

The estimator of matrix A constrained by the positivity condition can be equal to either of the three following:

i) $\hat{A}_T$, when $\hat{A}_T >> 0$;
a solution to (18), which can be either:
ii) a positive semi-definite matrix of rank 1;
iii) 0.
Under standard regularity conditions, the maximum value of type I error under the null is attained for $A = 0$, and is computed from a weighted mixture of chi-square

[9] see e.g. example iv) in Andrews, (1996), p. 705.

distributions, with weights equal to the probabilities of the three outcomes i), ii), iii) evaluated under $A = 0$.

For $A = 0$ however, some identification problems arise, as shown in the previous sections. Let us consider the asymptotic behavior of the likelihood ratio statistic when $A = 0$. Since the set of positive semi-definite matrices is a positive cone, we get:

$$\begin{aligned} \xi_T^P &= Min_{A:A>>0} T vech(\hat{A}_T - A)' \tilde{\Omega}_T^{-1} vech(\hat{A}_T - A) \\ &= Min_{A:A>>0} vech(\sqrt{T}\hat{A}_T - A)' \tilde{\Omega}_T^{-1} vech(\sqrt{T}\hat{A}_T - A) \\ \xrightarrow{d} \xi_\infty^P &= Min_{A:A>>0} vech(A_\infty - A)' \tilde{\Omega}^{-1} vech(A_\infty - A). \end{aligned} \quad (19)$$

Thus, (19) defines an asymptotic optimization criterion under $A = 0$. There are 3 regimes distinguished by the admissible values of that objective function:

Value in regime i) : $\xi_\infty^{1,P} = 0$;
Value in regime ii) : $\xi_\infty^{2,P} = vech(A_\infty - A_\infty^0)' \tilde{\Omega}^{-1} vech(A_\infty - A_\infty^0)$,
where $vech(A_\infty^0)' = (\alpha_\infty, \alpha_\infty \beta_\infty, \alpha_\infty \beta_\infty^2)$;
Value in regime iii) : $\xi_\infty^{3,P} = vech(A_\infty)' \tilde{\Omega}^{-1} vech(A_\infty)$.

The asymptotic probabilities of these regimes are denoted by $\pi_\infty^1, \pi_\infty^2, \pi_\infty^3$.

Let us now consider the type I error. We get

$$sup_{A>>0} \lim_{T\to\infty} P[\xi_T^P > c] = sup[sup_{A>>0, A\neq 0} \lim_{T\to\infty} P[\xi_T^P > c], P_{A=0}[\xi_\infty^P > c]].$$

By standard asymptotic theory underlying the tests of inequality constraints [see e.g. Gourieroux, Holly, Monfort (1980), Gourieroux, Monfort (1989), Wolak (1991)], the first component $sup_{A>>0, A\neq 0} \lim_{T\to\infty} P[\xi_T^P > c]$ is bounded from above by the survival function corresponding to a mixture of chi-square [10]:

$$\pi_\infty^1 \chi^2(0) + \pi_\infty^2 \chi^2(2) + \pi_\infty^3 \chi^2(3).$$

This survival function has to be compared with the survival function of ξ_∞^P under $A = 0$. This survival function is of the type:

$$\pi_\infty^1 \chi^2(0) + \pi_\infty^2 Q_\infty + \pi_\infty^3 \chi^2(3),$$

where Q_∞ denotes the asymptotic distribution of $\xi_\infty^{2,P}$. As in the previous sections, the limiting distribution Q_∞ and the probabilities of the regimes can be easily obtained from simulations.

4.2 *A of Reduced Rank*

Section 3 considered the estimation of A when the rank of matrix A is less or equal to 1. In this parametric framework, the positivity hypothesis can be written as

[10] Under regime ii), the standard theory implies a mixture of $\chi^2(1)$ and $\chi^2(2)$, which is bounded from above by a $\chi^2(2)$.

Table 3 Lower Quantiles of the Student Statistic for α

ρ	γ	$\eta_\alpha(1\%)$	$\eta_\alpha(5\%)$	$\eta_\alpha(10\%)$
0.000	0.500	-2.250	-1.510	-1.103
0.000	1.000	-2.283	-1.606	-1.248
0.000	1.500	-2.298	-1.615	-1.263
0.000	2.000	-2.312	-1.636	-1.264
0.100	0.500	-2.263	-1.518	-1.109
0.100	1.000	-2.313	-1.608	-1.239
0.100	1.500	-2.312	-1.642	-1.264
0.100	2.000	-2.327	-1.639	-1.266
0.200	0.500	-2.248	-1.520	-1.122
0.200	1.000	-2.314	-1.622	-1.234
0.200	1.500	-2.334	-1.635	-1.265
0.200	2.000	-2.327	-1.627	-1.262
0.300	0.500	-2.230	-1.528	-1.146
0.300	1.000	-2.267	-1.630	-1.260
0.300	1.500	-2.296	-1.652	-1.279
0.300	2.000	-2.326	-1.654	-1.281
0.400	0.500	-2.225	-1.534	-1.159
0.400	1.000	-2.290	-1.648	-1.279
0.400	1.500	-2.331	-1.686	-1.290
0.400	2.000	-2.317	-1.679	-1.300
0.500	0.500	-2.231	-1.540	-1.157
0.500	1.000	-2.309	-1.663	-1.281
0.500	1.500	-2.394	-1.691	-1.301
0.500	2.000	-2.344	-1.698	-1.308
0.600	0.500	-2.203	-1.525	-1.170
0.600	1.000	-2.352	-1.681	-1.285
0.600	1.500	-2.416	-1.727	-1.317
0.600	2.000	-2.391	-1.722	-1.316
0.700	0.500	-2.218	-1.567	-1.215
0.700	1.000	-2.461	-1.711	-1.316
0.700	1.500	-2.506	-1.757	-1.335
0.700	2.000	-2.468	-1.744	-1.320
0.800	0.500	-2.241	-1.576	-1.238
0.800	1.000	-2.435	-1.737	-1.331
0.800	1.500	-2.510	-1.755	-1.360
0.800	2.000	-2.413	-1.747	-1.342
0.900	0.500	-2.265	-1.614	-1.242
0.900	1.000	-2.423	-1.736	-1.365
0.900	1.500	-2.427	-1.734	-1.331
0.900	2.000	-2.377	-1.752	-1.317

$H_0 : (\alpha \geq 0)$. It is usually tested by a one-sided test based on the t-statistic $\hat{\eta}_T^{\alpha}$. As shown in Section 3, the asymptotic distribution of this test statistic is standard normal, except when $\alpha = 0$ (that is $A = 0$). We provide in Table 3 the one-sided critical value, that is the lower quantile of η_{∞}^{α} at 1%, 5%, 10%, derived by simulation with 5000 replications.

5 Finite Sample Properties

The previous sections were focused on the asymptotic distributions of test statistics. These distributions can be significantly different from the finite sample distributions.

To study the finite sample properties of standard test statistics, we generate three samples of iid standard Gaussian returns $(r_{1,t}, r_{2,t})'$, that are $IIN(0, Id)$, where Id denotes the identity matrix. The number of observations in each sample is $T = 50, 100, 200$. Next, we consider the following regressions:

Regression 1: $r_{1,t} = d + ar_{1,t-1}^2 + 2br_{1,t-1}r_{2,t-1} + cr_{2,t-1}^2 + v_t$;

Regression 2: $r_{1,t}^2 = d + ar_{1,t-1}^2 + 2br_{1,t-1}r_{2,t-1} + cr_{2,t-1}^2 + v_t$.

The first regression is a model with a bivariate risk premium, while the second one is a volatility transmission model.

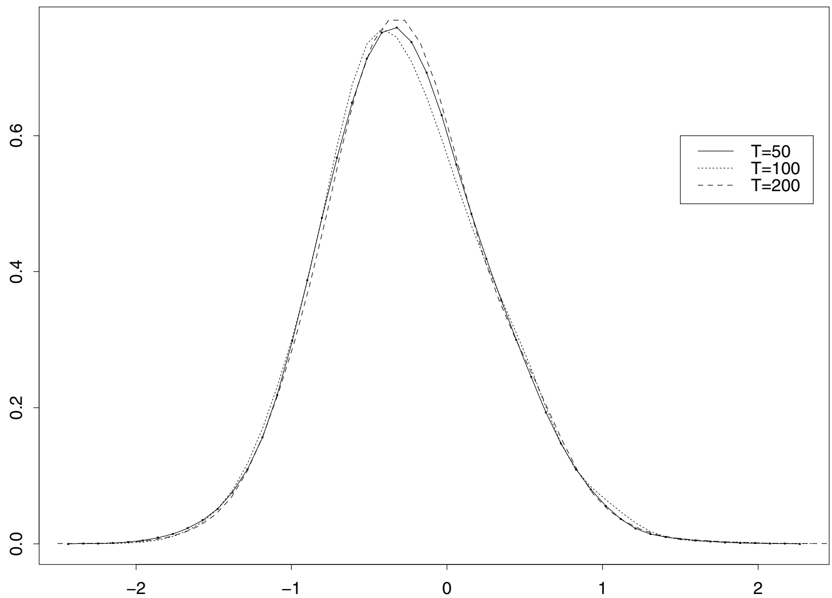


Fig. 2a Finite sample distribution of $\hat{\xi}_T$, Regression 1

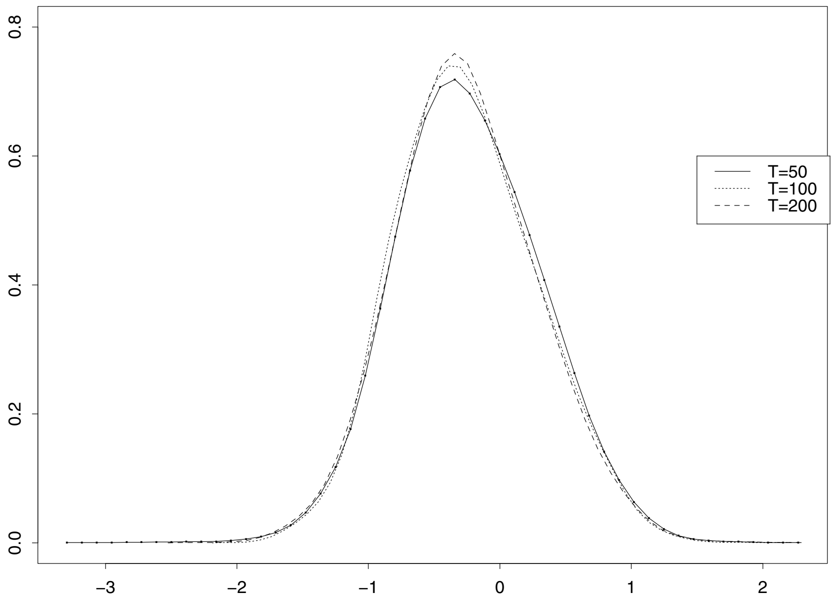


Fig. 2b Finite sample distribution of $\hat{\xi}_T$, Regression 2

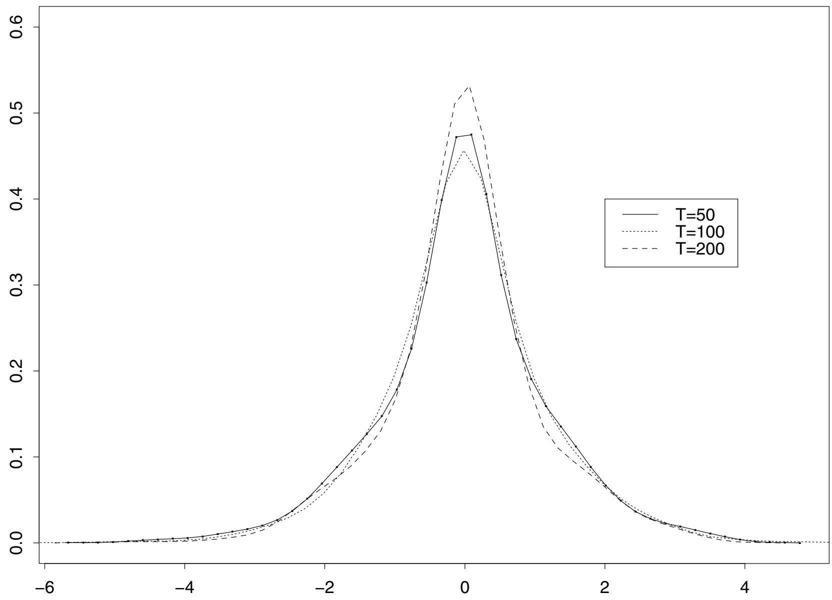


Fig. 3a Finite sample distribution of $\hat{\eta}_T^{\alpha}$, Regression 1

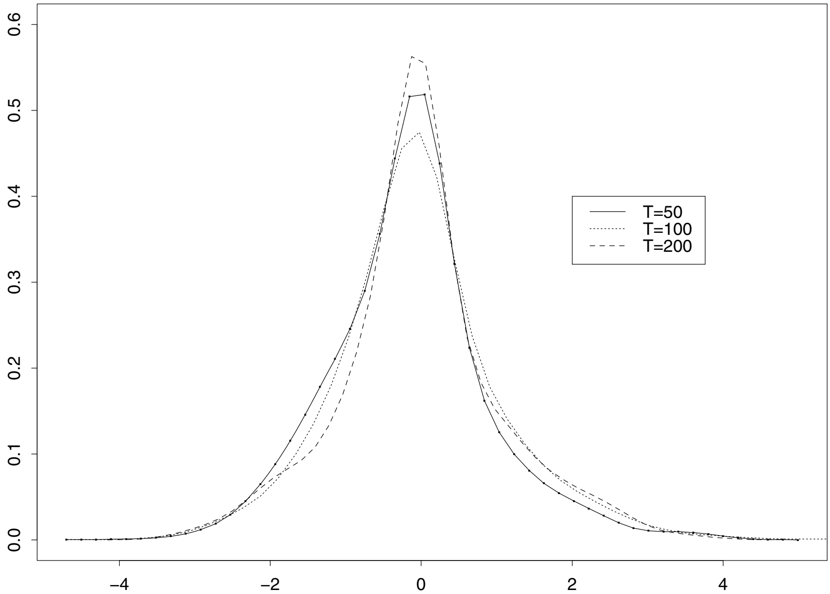


Fig. 3b Finite sample distribution of $\hat{\eta}_T^{\alpha}$, Regression 2

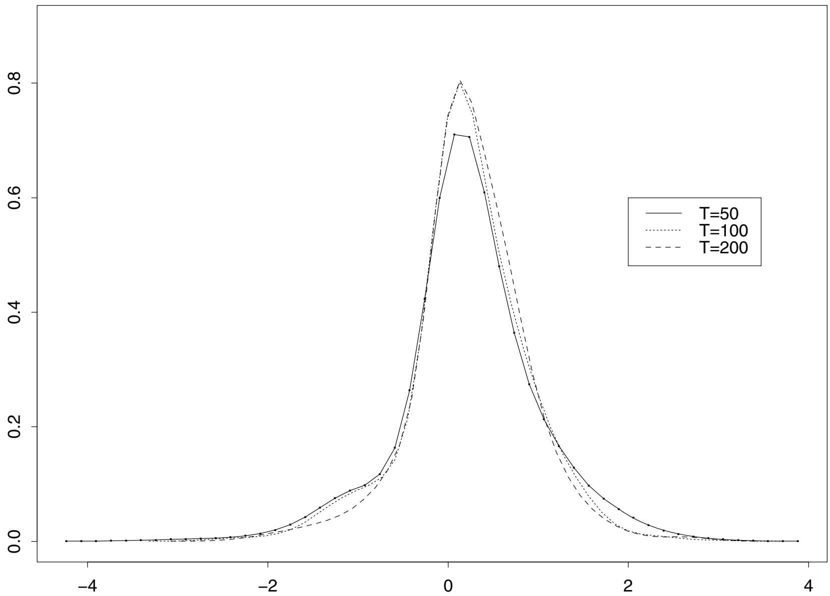


Fig. 4a Finite sample distribution of $\hat{\eta}_T^{\beta}$, Regression 1

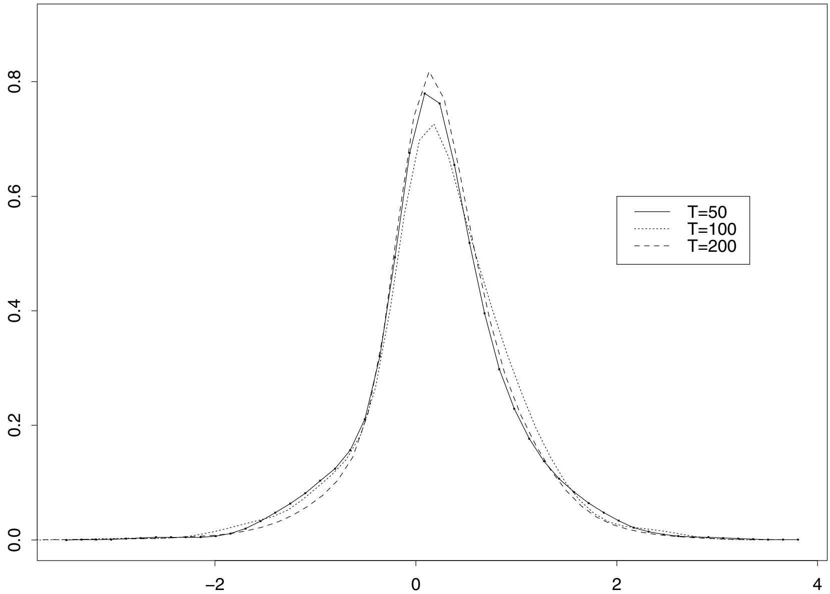


Fig. 4b Finite sample distribution of $\hat{\eta}_T^{\beta}$, Regression 2

For each regression, we determine the finite sample distributions of $\hat{\xi}_T, \hat{\eta}_T^{\alpha}, \hat{\eta}_T^{\beta}$, where the Wald statistics are derived from the OLS estimators of a, b, c with the OLS estimated variance-covariance matrix $\tilde{\Omega}_T$. The distributions of $\hat{\xi}_T$ for the two regressions are displayed in Figures 2a-2b. We observe fat tails, and different limiting distributions for each of the two regressions due to the differences between the limiting OLS covariance matrices for the two regressions (see Section 2.5).

Let us now consider the finite sample distributions of the t-ratios for α and β. All the distributions feature fat tails due to the stochastic variance in the denominator of the t-ratio.

6 Concluding Remarks

The paper derives the limiting distributions of standard estimators and test statistics for the analysis of return volatility and covolatility effects on the expected returns and future volatilities. When the volatility effects vanish, one can encounter difficulties that are due to non-identifiability of parameters, or to non-uniform convergence of the objective function used in estimation. Similar problems arise when the second-order causality is examined. Indeed, the null hypotheses of unidirectional second-order causality involve inequality restrictions, which entail identifiability problems of the type considered in this paper (see Gourieroux, Jasiak (2006), Gourieroux (2007), for the definition of causality hypotheses in terms of parameter restrictions).

Acknowledgement. The authors gratefully acknowledge financial support of the Natural Sciences and Engineering Council of Canada (NSERC). Gourieroux thanks the Chair AXA/Risk Foundation: "Large Risk in Insurance".

Appendix

Positivity Condition

Let us consider a linear form in symmetric positive semi-definite (2,2) matrices:

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \longrightarrow h(\Sigma) = a\sigma_{11} + 2b\sigma_{12} + c\sigma_{22}.$$

This linear form can be equivalently written as:

$$h(\Sigma) = Tr[A\Sigma],$$

where $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ and Tr is the trace operator, which computes the sum of diagonal elements of a square matrix.

Lemma 1: The linear form takes nonnegative values for any positive semi-definite matrix Σ, if and only if, matrix A is positive semi-definite.

Proof

Since the set of symmetric positive semi-definite matrices is a positive convex cone, it is equivalent to check the positivity condition on the boundary of the set. This boundary corresponds to the non invertible Σ matrices. These matrices can be written as

$$\Sigma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} (\alpha\ \beta) = \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{pmatrix}.$$

We get

$$h(\Sigma) = a\alpha^2 + 2b\alpha\beta + c\beta^2 \geq 0, \ \forall \alpha, \beta.$$

Let us assume $\alpha \neq 0$. The condition becomes:

$$a + 2b(\beta/\alpha) + c(\beta/\alpha)^2 \geq 0, \ \forall \alpha, \beta,$$

which is equivalent to $b^2 - ac \leq 0$ (the discriminant of the polynomial of degree 2 is nonpositive), and $a \geq 0$.

By considering the other case $\alpha = 0$, we see that $c \geq 0$.

The set of conditions: $a \geq 0, c \geq 0, ac - b^2 \geq 0$ is exactly the set of conditions for positive semi-definiteness of matrix A. QED

For dimension n larger than 2, it is known that the linear form $\Sigma \to Tr(A\Sigma)$ takes nonnegative values for any positive semi-definite matrix Σ, when A is symmetric positive semi-definite. However, this condition is no longer necessary.

From the proof of Lemma 1, the positivity semi-definiteness condition on A is also required if the linear form has to be nonnegative for any degenerate positive matrix Σ. This is important in ARCH modeling where the realized volatility matrix is generally approximated by squared returns $\Sigma_t = \begin{pmatrix} r_{1t}^2 & r_{1t}r_{2t} \\ r_{1t}r_{2t} & r_{2t}^2 \end{pmatrix}$, that has rank 1. Thus, it is not necessary for us to assume that Σ_t is invertible and to average square returns over a fixed window, for this reason (as suggested, for instance, in Tse, Tsui (2002)).

Proof of Proposition 1

a) Some Invariance Properties

Proposition A.1 below illustrates invariance properties of $det(A_\infty)$ and $\xi(A_\infty)$ with respect to linear transformations of matrix A_∞.

Proposition A.1 Invariance Properties: For any (n $\times$ n) invertible matrices P, Q,

we have:

i) $det(PA_\infty Q) = det(P)det(A_\infty)det(Q)$;
ii) $\xi(PA_\infty P') = \xi(A_\infty)$.

Proof: The proof is based on a succession of Lemmas

Lemma 2: If P and Q are (n,n) invertible matrices, we get:

$$cof(PAQ) = det(P)det(Q)Q^{-1}cof(A)P^{-1}.$$

Proof

From the identity $A\,cof(A) = det(A)\,Id$, it follows that

$$(PAQ)Q^{-1}cof(A)P^{-1}det(P)det(Q) = det(A)det(P)det(Q)\,Id = det(PAQ)\,Id.$$

The result follows. QED

Lemma 3: There exists a (n,n) permutation matrix Δ such that $vec(A') = \Delta vec(A)$. This matrix satisfies $\Delta = \Delta' = \Delta^2$.

Lemma 4: i) $vec(PA) = diag(P)vec(A)$, where $diag(P)$ denotes the bloc-diagonal matrix, with diagonal block P.
ii) $vec(AQ) = \Delta diag(Q')\Delta vec(A)$.
iii) $vec(PAQ) = diag(P)\Delta diag(Q')\Delta vec(A)$.

Proof

i) We have

$$\begin{aligned} PA &= P(a_1,...,a_n) \text{ (where } a_j \text{ denotes the } j^{th} \text{ column of} A) \\ &= (Pa_1,...,Pa_n). \end{aligned}$$

Thus, $vec(PA) = \begin{pmatrix} Pa_1 \\ \vdots \\ Pa_n \end{pmatrix} = diag(P)vecA.$

ii) We have

$$\begin{aligned} vec(AQ) &= \Delta vec[(AQ)'] \text{ (by Lemma 3)} \\ &= \Delta vec(Q'A') \\ &= \Delta diag(Q')vec(A') \text{ (from part } i)) \\ &= \Delta diag(Q')\Delta vecA \text{ (by Lemma 3).} \end{aligned}$$

iii) This follows directly from parts i) and ii).

QED

Lemma 5: For any (n,n) invertible matrix P, we have $\xi(PA_\infty P) = \xi(A_\infty)$.

Proof:

Let us consider the transformation:

$$A_\infty \longrightarrow A_\infty^* = PA_\infty Q,$$

where P and Q are deterministic (n,n) invertible matrices. We have:

$$\begin{aligned} vec(A_\infty^*) &= diag(P)\Delta diag(Q')\Delta vec(A_\infty) \text{ (by Lemma 4)}, \\ \Omega^* &= Var[vec(A_\infty^*)] = diag(P)\Delta diag(Q')\Delta\Omega\Delta diag(Q)\Delta diag(P'), \\ det(A_\infty^*) &= det(P)det(Q)det(A_\infty), \\ cof(A_\infty^*) &= det(P)det(Q)Q^{-1}cof(A_\infty)P^{-1}, \\ vec[cof(A_\infty^*)] &= det(P)det(Q)diag(Q^{-1})\Delta diag[(P')^{-1}]\Delta vec[cof(A_\infty)]. \end{aligned}$$

If $detP\,detQ > 0$, we find that:

$$\xi(A_\infty^*) = \frac{det(A_\infty^*)}{\sqrt{vec[cof(A_\infty^*)]'\Omega^* vec[cof(A_\infty^*)]}} = \frac{det(A_\infty)}{B_\infty},$$

where

$$\begin{aligned} B_\infty = {} & vec[cof(A_\infty)]'\Delta diag[(P)^{-1}]\Delta diag[(Q')^{-1}]diagP\Delta diag(Q')\Delta\Omega\Delta diag(Q)\Delta diag(P') \\ & diag(Q^{-1})\Delta diag[(P')^{-1}]\Delta vec[cof(A_\infty)]. \end{aligned}$$

It follows directly that, if $Q = P'$, we have

$$detP\, detQ = (detP)^2 > 0,$$

$B_\infty = vec[cof(A_\infty)]'\Omega vec[cof(A_\infty)]$ and $\xi(PA_\infty P) = \xi(A_\infty)$. The result follows.

b) **Proof of Proposition 1**

We use the invariance properties of $\xi(A_\infty)$ and $det(A_\infty)$ to show Proposition 1.

i) Let us consider a matrix $P = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$. We get:

$$PA_\infty P' = \begin{pmatrix} a_\infty \lambda^2 & b_\infty \lambda \mu \\ b_\infty \lambda \mu & c_\infty \mu^2 \end{pmatrix}.$$

Thus, it is always possible to standardize a_∞ and c_∞ to get $V(a_\infty) = V(c_\infty) = 1$.

ii) Let us now prove that we can find a linear transformation in order to have

$$Cov(a_\infty, b_\infty) = Cov(c_\infty, b_\infty) = 0.$$

For $P = \begin{pmatrix} 1 & \alpha \\ \beta & 1 \end{pmatrix}$, the matrix $A^* = PAP'$ is such that

$$\begin{aligned} a^*_\infty &= a_\infty + 2b_\infty \alpha + c_\infty \alpha^2, \\ b^*_\infty &= a_\infty \beta + (1 + \alpha\beta) b_\infty + c_\infty \alpha, \\ c^*_\infty &= a_\infty \beta^2 + 2b_\infty \beta + c_\infty. \end{aligned}$$

The condition $Cov(b^*_\infty, c^*_\infty) = 0$ implies

$$\alpha = -\frac{Cov(a_\infty \beta + b_\infty, a_\infty \beta^2 + 2b_\infty \beta + c_\infty)}{Cov(b_\infty \beta + c_\infty, a_\infty \beta^2 + 2b_\infty \beta + c_\infty)}.$$

By substituting this expression for α in the condition $Cov(a^*_\infty, b^*_\infty) = 0$, we get a polynomial in β of degree 5 (almost surely). This polynomial has at least one real root, which needs to be selected in order to obtain zero covariances.

The Solution in β Is Finite

When β tends to infinity, the quantity

$$\mu_T(\beta) =< vech\hat{A}_T, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >^2 / < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >$$

tends to $\hat{c}_T^2$. Moreover, the condition $\mu_T(\beta) > \hat{c}_T^2$ is equivalent to:

$$< vech\hat{A}_T, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >^2 - \hat{c}_T^2 < \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix}, \begin{pmatrix} 1 \\ \beta \\ \beta^2 \end{pmatrix} >> 0.$$

It is satisfied for a finite beta value, since the left-hand side of the inequality is a polynomial of degree 3.

References

1. Anderson, T.: An Introduction to Multivariate Statistical Analysis, 2nd edn. Wiley, New York (1984)
2. Andrews, D.: Tests for Parameter Instability and Structural Change with Unknown Change Point. Econometrica 61, 821–856 (1993)
3. Andrews, D.: Admissibility of the Likelihood Ratio Test when the Parameter Space is Restricted Under the Alternative. Econometrica 64, 705–718 (1996)
4. Andrews, D.: Estimation when a Parameter is on a Boundary. Econometrica 67, 1341–1383 (1999)
5. Andrews, D.: Testing when a Parameter is on the Boundary of the Maintained Hypothesis. Econometrica 69, 683–734 (2001)
6. Arnott, R., Bernstein, P.: What Risk Premium is Normal? Financial Analysts Journal 58, 64–85 (2002)
7. Arnott, R., Ryan, R.: The Death of the Risk Premium. Journal of Portfolio Management 27, 61–74 (2001)
8. Baba, Y., Engle, R., Kraft, D., Kroner, K.: Multivariate Simultaneous Generalized ARCH Models, DP, University of California, San Diego (1990)
9. Bilodeau, M., Brenner, D.: Theory of Multivariate Statistics. Springer (1999)
10. Bollerslev, T., Chou, R., Kroner, K.: ARCH Modelling in Finance. Journal of Econometrics 52, 5–59 (1992)
11. Bollerslev, T., Engle, R., Wooldridge, J.: A Capital Asset Pricing Model with Time-Varying Covariances. Journal of Political Economy 96, 116–131 (1988)
12. Boudoukh, J., Richardson, M., Smith, T.: Is the Ex-Ante Risk Premium Always Positive? A New Approach to Testing Conditional Asset Pricing Models. Journal of Financial Economics 34, 541–553 (1993)
13. Chen, L., Guo, H., Zhang, L.: Equity Market Volatility and Expected Risk Premium. Federal Reserve Bank of St. Louis (2006)
14. Comte, F., Lieberman, O.: Asymptotic Theory for Multivariate GARCH Processes. Journal of Multivariate Analysis 84, 61–84 (2003)
15. Cragg, J., Donald, S.: On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix. JASA 91, 1301–1309 (1996)
16. Cragg, J., Donald, S.: Inferring the Rank of a Matrix. Journal of Econometrics 76, 223–250 (1997)
17. Domowitz, I., Hakkio, C.: Conditional Variance and the Risk Premium in the Foreign Exchange Market. Journal of International Economics 22, 301–315 (1985)
18. Engle, R., Granger, C., Kraft, D.: Combining Competing Forecasts of Inflation Using A Bivariate ARCH Model. Journal of Economic Dynamics and Control 8, 151–165 (1984)
19. Engle, R., Kroner, K.: Multivariate Simultaneous Generalized ARCH. Econometric Theory 11, 122–150 (1995)
20. Francq, C., Zakoian, J.H.: Quasi-Likelihood Inference in GARCH Processes when Some Coefficients are Equal to Zero, CREST, DP (2006)

21. Gill, L., Lewbel, A.: Testing the Rank and Definiteness of Estimated Matrices with Applications to Factor, State-Space and ARMA Models. JASA 87, 766–776 (1992)
22. Gourieroux, C.: Positivity Condition for a Bivariate Autoregressive Volatility Specification. Journal of Financial Econometrics 5, 624–636 (2007)
23. Gourieroux, C., Holly, A., Monfort, A.: Kuhn-Tucker, Likelihood Ratio and Wald Tests for Nonlinear Models with Inequality Constraints on the Parameters. Harvard D.P. 770 (1980)
24. Gourieroux, C., Holly, A., Monfort, A.: Likelihood Ratio Test, Wald Test and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. Econometrica 50, 63–80 (1982)
25. Gourieroux, C., Jasiak, J.: Nonlinear Causality, with Application to Liquidity and Stochastic Volatility. CREST Working Paper (2006)
26. Gourieroux, C., Jasiak, J., Sufana, R.: The Wishart Autoregressive Process of Multivariate Stochastic Volatility. Journal of Econometrics (in Press)
27. Gourieroux, C., Monfort, A.: Statistique et Modeles Econometriques. Economica. English Translation 2 (1989); Statistics and Econometric Models, vol. 2. Cambridge Univ. Press, Cambridge (1995)
28. Gourieroux, C., Monfort, A., Renault, E.: Inference in Factor Models. In: Maddala, G., Phillips, P.C.B., Srinivasan, T. (eds.) Advances in Econometrics and Quantitative Economics, pp. 311–353. Brockwell, Oxford (1995)
29. Guggenberger, P., Hahn, J., Kim, K.: Specification Testing Under Moment Inequalities. UCLA, DP (2006)
30. Hakkio, C., Sibert, A.: The Foreign Exchange Risk Premium: Is it Real? Journal of Money, Credit and Banking 27, 301–317 (1995)
31. Hansen, P.: Asymptotic Tests of Composite Hypotheses. Brown University, DP 2003/09 (2003)
32. Iglesias, E., Phillips, G.: Bivariate ARCH Models: Finite Sample Properties of QML Estimators and an Application to an LM-type Tests. Econometric Theory 21, 1058–1086 (2005)
33. Jeantheau, T.: Strong Consistency of Estimators for Multivariate ARCH Models. Econometric Theory 14, 70–86 (1998)
34. Kodde, D., Palm, F.: Wald Criteria for Jointly Testing Equality and Inequality Restrictions. Econometrica 54, 1243–1248 (1986)
35. Macklem, T.: Forward Exchange Rates and Risk Premium in Artificial Economies. Journal of International Money and Finance 10, 365–391 (1991)
36. Moon, R., Schorfheide, F.: Empirical Likelihood Estimation with Inequality Moment Restrictions. University of Pennsylvania D.P (2005)
37. Ostdiek, B.: The World Ex-Ante Risk Premium: An Empirical Investigation. Journal of International Money and Finance 17, 967–999 (1998)
38. Robin, J., Smith, R.: Tests of Rank. Econometric Theory 16, 151–175 (2000)
39. Silberberg, G., Pafka, S.: A Sufficient Condition for the Positive Definiteness of the Covariance Matrix of the Multivariate GARCH Model. CEU, 7/2001 (2001)
40. Tse, Y., Tsui, A.: A Multivariate GARCH Model with Time Varying Correlations. Journal of Business and Economic Statistics 20, 351–362 (2002)
41. Walsh, K.: Is the Ex-Ante Risk Premium Always Positive? Further Evidence. Australian Journal of Management 31, 93–113 (2006)
42. Wolak, F.: The Local and Global Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models. Econometrica 59, 981–995 (1991)

Maximum Entropy Test for Autoregressive Models

Sangyeol Lee and Siyun Park

Abstract. In this paper, we apply the maximum entropy test developed for a goodness of fit in iid samples by [11] to autoregressive time series models including non-stationary unstable models. Its asymptotic distribution is derived under the null hypothesis. A bootstrap version of the test is also discussed and its performance is evaluated through Monte Carlo simulations. A real data analysis is conducted for illustration.

Keywords: Maximum entropy measure, goodness of fit test, time series models, unstable autoregressive models.

1 Introduction

The maximum entropy principle (cf. [7]) is well known as a criterion for selecting a priori probabilities. Maximum entropy modeling has been successfully applied to diverse research fields such as computer vision, spatial physics, natural language processing, and many other fields. For a probability density function $f(x)$, [5] provided the Boltzmann-Shannon entropy is defined as

$$H(f) = -\int_{-\infty}^{\infty} f(x)\log(f(x))dx. \tag{1}$$

[6] proposed the function $-\sum p_i \log(p_i/(x_i - x_{i-1}))$ as a discrete analogue of (1.1), where $p_i = P[x_{i-1} < X \leq x_i] = \int_{x_{i-1}}^{x_i} f(x)dx$, $i = 1, \ldots, n-1$ and $a = x_0 < \ldots < x_n = b$. Based on this, [11] constructed the maximum entropy test described in Section 2.

Sangyeol Lee
Department of Statistics, Seoul National University, Seoul, Korea
e-mail: sylee@stats.snu.ac.kr

Siyun Park
Business School, Korea University, Seoul, Korea
e-mail: siyunpark@korea.ac.kr

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 119–128.
DOI: 10.1007/978-3-642-35443-4_8

Goodness of fit (gof) tests measure the degree of agreement between the distribution of an observed random sample and a theoretical statistical distribution. Over the years, a number of gof methods including the chi-squared test and various empirical distribution function (edf) tests (cf. [3]) have been developed. In time series analysis, the gof test problem has been a crucial issue for modeling time series. In particular, the normality test attracted much attention from many researchers, since the normality of time series ensures several advantageous properties that non-normal time series do not possess. On the other hand, a prior information of non-normality is also beneficial in practice since, for instance, a heavy-tailed distribution modeling is required in the analysis of financial time series in practice. As relevant references, we refer to [12] who considered the empirical process gof test in autoregressive models, [8] who considered the Bickel-Rosenblatt test in stationary time series, and [10] who considered the Jarque-Bera test (cf. [1]) in GARCH models. Recently, [11] developed a maximum entropy test in iid settings and demonstrated its usefulness. Since the test outperforms several existing gof tests, in this study, we consider applying the maximum entropy test to autoregressive time series models (cf. [2] and [12]).

In Section 2, we introduce the test statistic and summarize its asymptotic distribution and properties in iid settings. In Section 3, we apply this test to autoregressive models. In Section 4, we perform a simulation study in order to explore the capabilities of the proposed test statistic. Particularly, a bootstrap method is employed to cope with small samples. In Section 5, we conduct a real data analysis for illustration. Concluding remarks are provided in Section 6.

2 Maximum Entropy Test

Let Y_i, $i=1,\dots,n$ be a random sample from a distribution with unknown distribution function F and consider the following test of fit:

$$H_0 : F = F_0 \quad \text{vs.} \quad F \neq F_0. \tag{1}$$

[11] considered the following generalization of [6] entropy:

$$S^w(F) = -\sum_{i=1}^{m} w_i \left(F(s_i) - F(s_{i-1})\right) \log\left(\frac{F(s_i) - F(s_{i-1})}{s_i - s_{i-1}}\right), \tag{2}$$

where the w's are appropriate weight functions with $0 \le w_i \le 1$ and $\sum_{i=1}^m w_i = 1$, m is the number of disjoint intervals for partitioning the data range, and $-\infty < a \le s_1 \le \ldots \le s_m \le b < \infty$ are preassigned partition points. For a properly selected constant c, the null hypothesis is rejected if $\sup_w |S^w(F_n) - S^w(F_0)| \ge c$, where $F_n(x) = n^{-1}\sum_{i=1}^n I(Y_i \le x)$.

Observe that if F_0 is the uniform distribution in $[0,1]$, then $S^w(F_0) = 0$. This fact enables us to concentrate on the uniform distribution on $[0,1]$ without loss of generality, so that the hypothesis problem is reduced to a uniform test. Specifically, the probability integral transform can be applied to construct the values $F_0(Y_i)$ denoted by U_i, and a test is then made of whether a uniform distribution is appropriate for the U_i's. To incorporate with various possible alternatives though, we consider the test with weights (cf. (2.2) and Theorem 2.1 below). Indeed, the role of weights are vital especially before the implementation of the uniform transformation. In our case, however, the data is transformed into uniform r.v.s and $s_i = i/m$ can be used to make a uniform spacing of the unit interval. Further, the supremum over all weights is taken to cope with any possibilities characterized by specific alternatives. This approach eases the difficulty of choosing optimal weights irrespective of their existence. The following is due to [11].

Theorem 2.1. Let $Y_1, \ldots, Y_n$ be a random sample from a continuous distribution with cumulative distribution function F. Under H_0 given in (2.1), as $n \to \infty$, we have

$$\sqrt{n} \sup_{w \in W} |S^w(F_n)| \xrightarrow{d} \sup_{w \in W} \left| \sum_{i=1}^{m} w_i \left(\mathscr{B}(s_i) - \mathscr{B}(s_{i-1}) \right) \right|, \tag{3}$$

where F_n is the sample distribution based on $U_i = F_0(Y_i)$, $\mathscr{B}(s)$ is a Brownian bridge on [0,1] (i.e., a Gaussian process with mean zero and the covariance structure such that $Cov(\mathscr{B}(s), \mathscr{B}(t)) = s \wedge t - st$ for all $s,t \in [0,1]$), W denotes the space of bounded weights $w_i : [0,1] \to [0,1]$ with $\sum_{i=1}^{m} w_i = 1$, and $0 = s_0 \le s_1 \le \ldots \le s_m = 1$. Here, the symbol $\xrightarrow{d}$ indicates the convergence in distribution.

It is common to test the composite null hypothesis that the unknown distribution belongs to a parametric family $\{F_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta} \in \Theta}$, where Θ is an open subset in R^k. In this case, a consistent estimator $\hat{\boldsymbol{\theta}}$ to test $H_0 : F = F_{\boldsymbol{\theta}}$ vs. H_1 : not H_0 by using $F_{\hat{\boldsymbol{\theta}}}(Y_i) = \hat{U}_i$. As discussed in [4], the limiting distribution is affected by the estimation of $\boldsymbol{\theta}$. The effect, though, may diminish when m is large and $\max_i(s_i - s_{i-1})$ is small. In the next section, we will demonstrate that this phenomenon can be also seen in autoregressive models.

3 Autoregressive Models

In this section, we consider the maximum entropy test for time series models. Let us consider the autoregressive model:

$$X_t - \beta_1 X_{t-1} - \cdots - \beta_q X_{t-q} = \varepsilon_t, \tag{1}$$

where ε_t are iid random variables with $E\varepsilon_1 = 0$, $E\varepsilon_1^2 = \sigma^2$ and $E\varepsilon_1^4 < \infty$. We assume that the corresponding characteristic polynomial ϕ has a decomposition

$$\begin{aligned} \phi(z) &= 1 - \beta_1 z - \cdots - \beta_q z^q \\ &= (1-z)^a (1+z)^b \prod_{k=1}^{l} (1 - 2\cos\theta_k z + z^2)^{d_k} \psi(z), \end{aligned}$$

where a, b, l, d_k are nonnegative integers, θ_k belongs to $(0, \pi)$ and $\psi(z)$ is the polynomial of order $r = q - (a + b + 2d_1 + \ldots + 2d_l)$ that has no zeros on the unit disk in the complex plane. If a, b, l, d_k are all zeros, $\{X_t\}$ in (3.1) is a stationary process.

Let $\mathbf{X}_t = (X_t, \ldots, X_{t-q+1})'$, where $X_t = 0$ for all $t \leq 0$. Let

$$\hat{\boldsymbol{\beta}}_n = \left(\sum_{t=1}^{n} \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \right)^{-1} \sum_{t=1}^{n} \mathbf{X}_{t-1} X_t, \quad n > q,$$

be the least squares estimate (LSE) of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$ based on $X_1, \ldots, X_n$. We denote by $\boldsymbol{\beta}_0$ the true parameter of $\boldsymbol{\beta}$. Then the residuals are $\hat{\varepsilon}_t = X_t - \hat{\boldsymbol{\beta}}_n' \mathbf{X}_{t-1}, \quad t = 1, \ldots, n$. It is well known that the LSE has a limiting distribution of a functional form of standard Brownian motions (cf. [2]). The residual empirical process has been considered to test $H_0' : \varepsilon_t \sim F_0(\cdot/\sigma),\ \sigma > 0$ vs. H_1' : not H_0' (cf. [12] and [9]), where F_0 is strictly increasing and twice differentiable and satisfies $\int x dF_0(x) = 0$.

According to [12], under some regularity conditions, the residual empirical process

$$\hat{\mathscr{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s)$$

with

$$\hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\varepsilon}_t / \hat{\sigma}_n) \leq s)$$

can be expressed as

$$\begin{aligned} \hat{\mathscr{V}}_n(s) &= \mathscr{V}_n(s) - (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{X}_{t-1} f_0(F_0^{-1}(s)) \\ &\quad + \frac{\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)}{2\sigma_0^2} f_0(F_0^{-1}(s)) F_0^{-1}(s) + \Delta_n(s), \end{aligned} \tag{2}$$

where σ_0 and $\boldsymbol{\beta}_0$ denote the true values under H_0, $\mathscr{V}_n(s) = \sqrt{n}(F_n(s) - s)$, $F_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\varepsilon_t / \sigma_0) \leq s)$, $f_0 = F_0'$, and $\sup_s |\Delta_n(s)| = o_P(1)$. Further, it is well known that $(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{X}_{t-1} = O_P(1)$. Here, we assume that

$$\lim_{|x| \to \infty} |x f_0(x)| = 0 \text{ and } \sup_x |f_0'(x)| < \infty. \tag{3}$$

Then similarly to (2.3), we have the following result.

Theorem 3.1. Under H_0' and (3.3), if $\max_{1 \leq i \leq m} |s_i - s_{i-1}| \to 0$ as $m \to \infty$, we have that for large m, as $n \to \infty$,

$$\hat{T}_n := \sqrt{n} \sup_{w \in W} |S_{\max}^w(\hat{F}_n)| \overset{d}{\approx} \sup_{w \in W} \left| \sum_{i=1}^{m} w_i \left(\mathscr{B}(s_i) - \mathscr{B}(s_{i-1}) \right) \right|, \tag{4}$$

where the symbol $A_n \overset{d}{\approx} A$ *as* $n \to \infty$ *indicates that the limiting distribution of* A_n *is approximately the same as the distribution of* A *as* n *tends to* ∞.

Proof. Express

$$S^w_{\max}(\hat{F}_n) = -\sum_{i=1}^m w_i \left(\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})\right) \cdot \log\left(\frac{\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})}{s_i - s_{i-1}} - 1 + 1\right).$$

Recall that $\hat{F}_n(s) \to s$ in probability under the null. Then, by using the fact that $|\log(1+x) - x| \le x^2$ for $|x| < 1/2$, due to (3.2), we have

$$\begin{aligned} S^w_{\max}(\hat{F}_n) &= -\sum_{i=1}^m w_i \left(\frac{\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})}{s_i - s_{i-1}}\right) \\ &\quad \cdot \left[\left(\hat{F}_n(s_i) - s_i\right) - \left(\hat{F}_n(s_{i-1}) - s_{i-1}\right)\right] + o_P(1/\sqrt{n}) \\ &= \frac{-1}{\sqrt{n}} \sum_{i=1}^m w_i \left(\frac{\hat{F}_n(s_i) - \hat{F}_n(s_{i-1})}{s_i - s_{i-1}}\right)\left(\hat{\mathscr{V}}_n(s_i) - \hat{\mathscr{V}}_n(s_{i-1})\right) + o_P(1/\sqrt{n}), \end{aligned}$$

so that since $\mathscr{V}_n \overset{w}{\to} \mathscr{B}$, for large m, as $n \to \infty$,

$$\begin{aligned} &\sup_{w\in W}\left|\sqrt{n}S^w_{\max}(\hat{F}_n) - \sum_{i=1}^m w_i(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbf{X}_{t-1}\left\{f_0(F_0^{-1}(s_i)) - f_0(F_0^{-1}(s_{i-1}))\right\}\right. \\ &+ \left.\sum_{i=1}^m w_i \frac{\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)}{2\sigma_0^2}\left\{f_0(F_0^{-1}(s_i))F_0^{-1}(s_i) - f_0(F_0^{-1}(s_{i-1}))F_0^{-1}(s_{i-1})\right\}\right| \\ &\overset{d}{\approx} \sup_{w\in W}\left|\sum_{i=1}^m w_i(\mathscr{V}_n(s_i) - \mathscr{V}_n(s_{i-1}))\right| \overset{d}{\approx} \sup_{w\in W}\left|\sum_{i=1}^m w_i(\mathscr{B}(s_i) - \mathscr{B}(s_{i-1}))\right|. \end{aligned}$$

Hence, $\hat{T}_n \overset{d}{\approx} \sup_{w\in W}\left|\sum_{i=1}^m w_i\left(\mathscr{B}(s_i) - \mathscr{B}(s_{i-1})\right)\right|$ by virtue of (3.2). This completes the proof. □

Remark 3.1. In order to implement our test in practice, we consider $w_i^{(l)}, l = 1, \ldots, L$, independent and identically distributed random variables from $U[0,1]$, which are also independent from $U_i \sim U[0,1]$, where L is a fixed positive integer. Then, if we put

$$w_{li} = \frac{w_i^{(l)}}{w_1^{(l)} + \cdots + w_m^{(l)}},$$

we have that as $L \to \infty$,

$$\max_{1\le l\le L}\left|\sum_{i=1}^m w_{li}\left(\mathscr{B}(s_i) - \mathscr{B}(s_{i-1})\right)\right| \overset{d}{\to} \sup_{w\in W}\left|\sum_{i=1}^m w_i\left(\mathscr{B}(s_i) - \mathscr{B}(s_{i-1})\right)\right|.$$

Subsequently, instead of the test of the form in (3.4), by taking $s_i = i/m$, $i = 1,\ldots,m$ for convenience, we can use as the maximum entropy test the quantity

$$\begin{aligned}\hat{T}_n &= \sqrt{n} \max_{1\le l\le L} \Big| \sum_{i=1}^{m} w_{li}(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \\ &\quad \times \log m(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \Big| \\ &\overset{d}{\approx} \sup_{w\in W} \Big| \sum_{i=1}^{m} w_i(\mathscr{B}(i/m) - \mathscr{B}((i-1)/m)) \Big|.\end{aligned} \quad (5)$$

For more details, see [11].

Remark 3.2. Conventionally, m is chosen to be much less than n so that m/n is close to 0. Either a too small or a too large m will give unsatisfactory results. Thus, for given n, the choice of m can be an important issue in practice. Nevertheless, in a theoretical aspect, there is no such a rule to choose an optimal m. Furthermore, Theorem 3.1 only provides an asymptotic result and cannot be directly applied to small samples. Therefore, in this study, we recommend to use a bootstrap method as in [13]. The detailed procedure is as follows:

(1) Based on the data $X_1,\ldots,X_n$, obtain the LSE $\hat{\boldsymbol{\beta}}_n$ and $\hat{\sigma}_n^2$.
(2) Generate $\varepsilon_1^*,\ldots,\varepsilon_n^*$ that follow $F_0(\cdot/\hat{\sigma}_n)$ and construct $X_1^*,\ldots X_n^*$ obtained through Eq. (3.1) with $\boldsymbol{\beta}$ replaced by its LSE by letting $X_i^* = 0$ for all $i \le 0$. Then, calculate $\hat{T}_n$ with a preassigned m in (3.5) based on these r.v.s.
(3) Repeat the above procedure B times and calculate the $100(1-\alpha)\%$ percentile of the obtained B number of $\hat{T}_n$ values.
(4) Reject H_0 if the $\hat{T}_n$ value based on the original observations is larger than the obtained $100(1-\alpha)\%$ percentile in (3).

The above bootstrap method is easy to implement and gives satisfactory results as seen in the next section.

4 Simulation Results

In this section, we evaluate the performance of the maximum entropy test in Section 3 through a simulation study. In this simulation study, we consider the AR(1) model where ε_t are assumed to be iid $N(0,1)$ for the null hypothesis H_0. For alternative H_1, we consider that ε_t are assumed to follow a normal mixture model:

$$\varepsilon_t \sim p\,N(0,1) + (1-p)\,N(0,\sigma_0^2), \qquad p = 0.9, \sigma_0^2 = 10, 25.$$

We examine the empirical sizes and powers of the test with sample sizes n and m as $(n,m) = (100,3), (300,5)$, and $(500,7)$ at the nominal levels 0.01, 0.05, and 0.1. The empirical sizes and powers are calculated as the number of rejections of the null hypothesis H_0 out of 1000 repetitions. In performing a test, we consider the bootstrap test discussed in Remark 3.2 with B=500 and explore the sizes and powers for the AR(1) coefficient β_1= 0.3, 0.5, 0.7, and 1.0 in (3.1). Table 1 shows that the empirical sizes are close to the nominal levels and Table 2 shows that the bootstrap test produces reasonably good powers. Overall, our result strongly suggests that the maximum entropy test performs appropriately.

5 Real Data Analysis

In this section, we apply our test to the real data set in [14], Chapter 2, Example 2.2. The data is a daily log series of the S&P 500 index from January 1990 to December 2003: the total number of observations is 3532. Figure 1 suggests that the series exhibits a non-stationary phenomenon and has high sample serial correlations. Based on the results in Tsay (2005) and using the AIC method, the AR(14) model is selected as an underlying model. To test for a unit root in the AR(14) model, the augmented Dickey-Fuller (ADF) test is conducted and the ADF test value -0.9648 is obtained with p-value 0.9444, which supports the existence of a unit root in the AR(14) model. Table 3 summarizes the LSE of the autoregressive coefficients β_i, $i = 1,\ldots,14$. To test H_0: $\varepsilon_t \sim_{iid} N(0,\sigma^2)$ vs. H_1 : not H_0, the bootstrap test in Remark 3.2 with B=500 is implemented. The maximum entropy test statistic with m=10 turns out to have value 1.162, so the null hypothesis is rejected at the nominal levels 0.01 0.05 and 0.1. In fact, one can check that only β_i, $i = 1,2,4,5,8,12,14$, are nontrivial and conduct the same normality test for the restricted AR(14) model only with these coefficients. In this case, though, the normality assumption is rejected as before. Further, it can be seen that the choice of different values of $m = 7,9,12,15$ does not alter this conclusion. Our findings confirm that the error terms do not follow a normal distribution.

6 Concluding Remarks

In this study, we extended the maximum entropy test developed for iid samples to unstable autoregressive models. Its limiting null distribution is derived under regularity conditions. To cope with the small samples and the difficulty regarding the choice of the number of cells, say m, in the test (the stability of the test may depend upon m), we suggested to use a bootstrap method and showed through a simulation study that the bootstrap test has no size distortions and produce good powers. A real data analysis is also conducted for the S&P 500 index data set. It is revealed that the error terms in the fitted AR model are not normally distributed. Overall, our findings demonstrate the validity of our test.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0010936) grant number: 2011-0010936. We would like to thank Jiyun Lee for her help with the real data analysis. We also thank the referee for his/her careful reading and helpful suggestion.

Appendix

Table 1 Empirical sizes based on the bootstrap method with $B = 500$

β	$m=3, n=100$ 0.01	0.05	0.10	$m=5, n=300$ 0.01	0.05	0.10	$m=7, n=500$ 0.01	0.05	0.10
0.3	0.009	0.049	0.111	0.007	0.056	0.106	0.010	0.053	0.110
0.5	0.009	0.046	0.100	0.016	0.043	0.085	0.013	0.057	0.117
0.7	0.016	0.043	0.085	0.013	0.051	0.112	0.013	0.051	0.112
1.0	0.011	0.051	0.096	0.015	0.048	0.096	0.007	0.04	0.098

Table 2 Empirical powers based on bootstrap method with $B = 500$

β	σ_0^2	$m=3, n=100$ 0.01	0.05	0.10	$m=5, n=300$ 0.01	0.05	0.10	$m=7, n=500$ 0.01	0.05	0.10
0.3	10	0.251	0.386	0.466	0.567	0.771	0.858	0.824	0.942	0.973
0.5	10	0.245	0.368	0.461	0.569	0.767	0.862	0.837	0.928	0.954
0.7	10	0.257	0.394	0.487	0.592	0.747	0.838	0.820	0.933	0.964
1.0	10	0.296	0.410	0.488	0.624	0.787	0.866	0.853	0.954	0.980
0.3	25	0.740	0.823	0.857	0.992	0.997	1.000	1.000	1.000	1.000
0.5	25	0.769	0.828	0.864	0.990	0.998	0.998	0.999	1.000	1.000
0.7	25	0.770	0.852	0.885	0.992	0.998	0.999	1.000	1.000	1.000
1.0	25	0.810	0.859	0.888	0.995	1.000	1.000	1.000	1.000	1.000

Table 3 The least squares estimates of the AR(14) model

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
0.9964	-0.0211	-0.0125	0.0421	-0.0422	0.0169	-0.0244

$\hat{\beta}_8$	$\hat{\beta}_8$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$
0.0456	0.0085	0.0052	-0.0287	0.0655	-0.0196	-0.0317

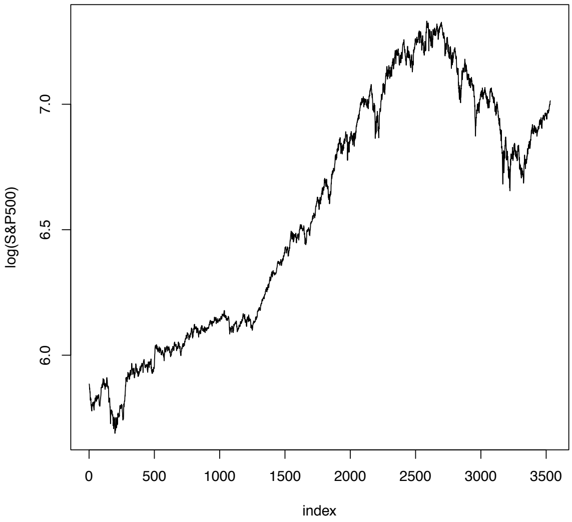


Fig. 1 Plot of the logarithm of daily S&P 500

References

1. Bera, A., Jarque, C.: Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals: monte carlo evidence. Econometric Letters 7, 313–318 (1981)
2. Chan, N.H., Wei, C.Z.: Limiting distributions of least squares estimates of unstable autoregressive processes. Ann. Statist. 16, 367–401 (1988)
3. D'Agostino, R.B., Stephens, M.A.: Goodness-of-Fit Techniques. Marcel Dekker, New York (1986)
4. Durbin, J.: Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. 1, 279–290 (1973)
5. Forte, B.: Entropies with and without probabilities: to Applications to questionnaires. Information Processing and Management 20, 397–405 (1984)
6. Forte, B., Hughes, W.: The maximum entropy principle: a tool to define new entropies. Reports on Mathematical Physics 26, 227–235 (1988)
7. Jaynes, E.T.: Information Theory and Statistical Mechanics. In: Ford, K. (ed.) Statistical Physics, pp. 181–218. Benjamin, New York (1963)
8. Lee, S., Na, S.: On the Bickel-Rosenblatt test for first-order autoregressive models. Statist. Probab. Lett. 56, 23–35 (2001)
9. Lee, S., Na, O., Na, S.: On the cusum of squares test for variance change in nonstationary and nonparametric time series models. Ann. Inst. Statist. Math. 55, 467–485 (2003)
10. Lee, S., Park, S., Lee, T.: A note on the Jarque-Bera normality test for GARCH innovations. J. Korean Statist. Soc. 39, 93–102 (2010)
11. Lee, S., Vonta, I., Karagrigoriou, A.: A maximum entropy type test of fit. Computational Statist. Data Anal. 55, 2635–2643 (2011)

12. Lee, S., Wei, C.Z.: On residual empirical processes of stochastic regression models with applications to time series. Ann. Statist. 27, 237–261 (1999)
13. Stute, W., Manteiga, W.G., Quindimil, M.P.: Bootstrap based goodness-of-Fit tests. Metrika 40, 243–256 (1993)
14. Tsay, R.S.: Analysis of Financial Time Series. Wiley Series in Probability and Statistics (2005)

Choice of Copulas in Explaining Stock Market Contagion

Kian-Guan Lim

Abstract. We provide in this paper an assessment of how well the Archimedean class of copulas can explain equity market contagion across regions. In particular we examine the Clayton, the Gumbel, and the Frank copulas. Three representative large equity markets across the globe in U.S., in U.K., and in Japan are studied. The S&P 500, FTSE 100, and the Nikkei 225 stock market indices of the three countries are used to compute proxy large portfolio returns. The joint daily return vectors of these three equity markets are tracked over the period from the beginning of January 1990 till end of April 2012. The Kullback-Leibler distances (divergences) or relative entropy of the copulas with respect to the empirical distribution are compared with a benchmark t-copula relative entropy. We then narrow the focus on the conditional joint tail losses of the multivariate return distribution using the Pareto Type II distribution to model the tails. The maximum likelihood approach is used for estimating the parameters of the marginal conditional tail distributions and the copulas. The observed joint returns in the loss region of at least one standard deviation away from the mean are then matched in frequencies across 27 three by three cells with the theoretical probabilities based on the estimated parameters under the competing copulas. A goodness-of-fit test together with the relative entropy results show that the Clayton copula is statistically the most appropriate copula in explaining contagion during this sampling period.

Keywords: Archimedean copulas, stock market contagion, conditional tail losses, Kullback-Leibler distance.

1 Introduction

The global financial crisis that started in 2008 with the sub-prime mortgage market collapse in U.S. led to immense loss of confidence in the world banking system.

Kian-Guan Lim
Lee Kong Chian Business School,
Singapore Management University, 50 Stamford Road, Singapore 178899
e-mail: kgl@smu.edu.sg

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 129–140.
DOI: 10.1007/978-3-642-35443-4_9

This problem continued in the form of the sovereign debt crisis in the Eurozone since 2010. The continuing weakness in world consumer demand, the high unemployment problems in Europe and the potentially slower growth and high inflation in China and Asia have led to high volatility and extreme movements in equity and other asset prices in recent years. There are two significant observations in the above development. First, country stock markets have since evidenced more frequent sudden sharp losses on a daily basis. This has heightened the attention of investment funds to risk management and the careful monitoring of portfolio positions to avoid being stuck in heavy losses. Second, such drops appear to be global in a contagion effect across countries around the globe within a matter of several trading hours or within the next trading day. Since the U.S. market is the largest capital market in the world and has been at the epicenter of the global financial earth-quake of 2008, it is also noticed that the contagion typically, though not always, starts with the U.S. market movement as the lead.

Choice of Copulas

We provide in this paper a study of the effectiveness of employing copulas to model contagion, and in particular focus on the choice of which copulas to use in the modeling. The contribution is to provide market analysts and researchers the approach to a better choice of copulas in such modeling, and also to provide empirical evidence of the choice. We consider the Archimedean class of copulas, specifically the Clayton, the Gumbel, and the Frank copulas, in explaining equity market contagion across regions. These Archimedean copulas are considered because they represent a parsimonious class of copulas employing only one copula parameter in the function of marginal distributions. In terms of obtaining maximum likelihood functions, this class of copulas also provides analytical derivations.

Copulas are a useful statistical tool because they constitute multivariate distribution function in a straight forward manner using marginal densities of the individual random variables, and the functional form of the copula with its parameter(s) characterizes the dependence structure of the individual random variables. The co-movements in the individual random variables are invariant to strictly monotone transforms and the dependence structure is more general than the linear Pearsonian correlation measure. Its drawbacks may be also the general difficulty of relating the dependence to the popular and more familiar Pearsonian correlations. The latter, however, is a further advantage of using the Archimedean class as the Clayton, the Gumbel, and the Frank copula parameters are associated directly with a Kendall's tau measure which is a non-parametric rank measure of correlations. See [11].

Large Equity Markets

Three representative large equity markets across the globe in U.S., in U.K., and in Japan are studied. The S&P 500, FTSE 100, and the Nikkei 225 stock market indices of the three countries are used to compute proxy large portfolio returns. The joint daily return vectors of these three equity markets are tracked over the period from the beginning of January 1990 till end of April 2012. Several procedures of aligning the U.S. S&P 500, the U.K. FTSE100 and the Japan NK225 daily index returns were tried, and it was found that the highest correlations were obtained by aligning reported day t returns of U.S. and U.K. with reported day $t+1$ return of the N225. Daily returns correlations between S&P 500 and FTSE100, and between S&P 500 and N225 were 0.486 and 0.401 respectively. That between FTSE100 and N225 was 0.258. The implication, as also found in many market information transmission research, is that both the U.S. and U.K. markets affect the Japanese market the next day after the U.S. and U.K. markets close for the day. The New York market and the London market overlap in opening hours though the U.K. market would close earlier. However, while New York market is trading, it would also affect the closing prices of the London market on the same reported trading day.

We first show the empirical distribution of the index returns for the entire sampling period (January 1990-April 2012) in Fig. 1. The set of daily returns from S&P 500 index, FTSE index, and N225 index in our sample comprised 6820 data points per index. Figure 1 shows that during the entire sampling period, the corresponding cdf uniform variables of the returns were mostly clustered around the centre of the unit cube and toward both the upper and lower ends. The cdf uniform variable of a return is obtained by first sorting the N returns from the smallest to the largest value, and then assigning k/N if a return is the k^{th} value from the smallest value. In terms of the implications to financial investments and bank portfolios, the lower end closest to the viewer is the simultaneous loss region that is of particular importance and special interest in our study.

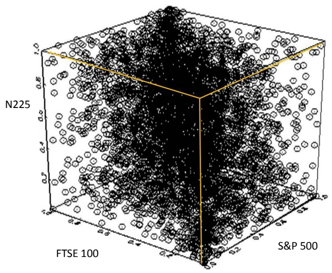


Fig. 1 CDF's of the S&P500, FTSE100, and N225 Index daily returns January 1990- April 2012

Modeling Contagion

To study contagion, we consider only the conditional joint tail losses of the joint multivariate return distributions. The conditional tail losses are modeled using Pareto Type II distribution. The latter type of distribution will be explained in detail in a later section. The maximum likelihood approach is used for estimating the distributional parameters.

Our initial assessment of the copulas was based on a semi-parametric approach by firstly computing the maximum likelihood estimates of the copula parameters, and secondly, computing estimates of the relative entropy or Kullback-Leibler (KL) distance or information criterion $E_0\left(\ln\frac{L_0}{L_1(\theta)}\right)$ where subscript "0" to the expectation operator denotes probability measure with respect to the null hypothesis distribution, and L_0 is also the log-likelihood function under the null hypothesis distribution. $L_1(\theta)$ is the log-likelihood function under the alternative probability or copula where its parameter θ is also estimated at the same time by maximizing the likelihood function. In the likelihood estimation, we use the empirical distribution of the uniform variates corresponding to the returns of S&P 500, FTSE100, and NK225, as null. Since the actual null distribution is not specified parametrically, the resulting relative entropy measure cannot be used to approximate an asymptotic χ^2 likelihood ratio test-statistic. The null distribution is unclear largely because the marginal distributions entering the copula functions are not specified. The empirical results are reported in Section 2 where we also provide further discussion on the specific copulas. In Section 3 we perform a modeling of the loss tails. In Section 4 we complete the study by using the maximum likelihood approach to estimate both the marginal distribution and the copula parameters and also perform a goodness-of-fit test on the various copulas. Section 5 contains the conclusions.

2 Maximum Likelihood Estimation

For the Clayton copula function,

$$C_C(u_1,u_2,\ldots,u_N)=\left(\sum_{i=1}^{N}u_i^{-\beta}-N+1\right)^{-\frac{1}{\beta}},\ \ \beta>0\ .$$

For the Gumbel copula,

$$C_G(u_1,u_2,\ldots,u_N)=\exp\left\{-\left[\sum_{i=1}^{N}(-\ln u_i)^{\beta}\right]^{\frac{1}{\beta}}\right\},\ \ \beta>1$$

For the Frank copula,

$$C_F(u_1,u_2,\ldots,u_N)=-\frac{1}{\beta}\ln\left\{1+\frac{\Pi_{i=1}^{N}\left(e^{-\beta u_i}-1\right)}{\left(e^{-\beta}-1\right)^{N-1}}\right\},\ \ \beta>0$$

for $N\geq 3$.

In all cases, $u_i(z_i)$ denotes the cumulative distribution function of underlying random variable z_i, for $i=1,2,\ldots,N$. Each copula is parameterized by a single β that of course carries different values in the different copulas. We also employ the t-copula $C_t(u_1,u_2,\ldots,u_N)$ as follows, as a benchmark for comparison:

$$\int_{-\infty}^{t_v^{-1}(u_1)}\cdots\int_{-\infty}^{t_v^{-1}(u_N)}\frac{\Gamma\left((v+N)/2\right)}{\Gamma(v/2)\sqrt{(\pi v)^N|\Sigma|}}\left(1+\frac{x'\Sigma^{-1}x}{v}\right)^{-(v+N)/2}dx$$

where t_v is a standard univariate student-t distribution with v degrees of freedom, and Σ is the correlation matrix of the $t_v^{-1}(u_i)$ variables implied by the uniform u_i's.

The pdf's of the multi-dimensional distributions corresponding to the various copula functions are found as follows. For the Clayton pdf,

$$f_C(u_1,u_2,\ldots,u_N) \equiv \frac{\partial^N C}{\partial u_1 \partial u_2 \ldots \partial u_N} \prod_{i=1}^{N} \frac{\partial u_i}{\partial z_i} = ABD \prod_{i=1}^{N} f_i(z_i) \tag{1}$$

where $f_i(z_i)$ is the pdf of r.v. z_i, and

$$A = \left(\prod_{j=1}^{N} u_j \right)^{-(\beta+1)}, \; B = \prod_{j=1}^{N-1} (1+j\beta), \; D = \left[\sum_{j=1}^{N} u_j^{-\beta} - N + 1 \right]^{-\left(\frac{1}{\beta}+N\right)}$$

For Gumbel pdf,

$$f_G(u_1,u_2,\ldots,u_N) = C_G BP \prod_{i=1}^{N} f_i(z_i) \tag{2}$$

where

$$B = \prod_{i=1}^{N} \left[\frac{1}{u_i} (-\ln u_i)^{\beta-1} \right], \quad P = \sum_{i=1}^{N} a_i^N A^{\frac{i}{\beta}-N} \; ,$$

$A = \sum_{i=1}^{N} (-\ln u_i)^{\beta}$, and a_i^N is defined as follows: $a_1^{k+1} = a_1^k(k\beta - 1)$, $a_2^{k+1} = a_1^k + a_2^k(k\beta - 2)$, $a_3^{k+1} = a_2^k + a_3^k(k\beta - 3), \ldots, a_k^{k+1} = a_{k-1}^k + a_k^k(k\beta - k)$, and $a_{k+1}^{k+1} = 1$, for $1 \le i \le N$.

For the Frank pdf,

$$f_F(u_1,u_2,\ldots,u_N) = \left(\sum_{j=1}^{N} c_j A^{-j} P_j \right) \prod_{i=1}^{N} f_i(z_i) \tag{3}$$

where

$$\begin{aligned} A &= 1 + \frac{\prod_{i=1}^{N} \left(e^{-\beta u_i} - 1\right)}{\left(e^{-\beta} - 1\right)^{N-1}} \\ P_j &= \frac{(-\beta)^{N-1} \prod_{k=1}^{N} e^{-\beta u_k} \left[\prod_{k=1}^{N} \left(e^{-\beta u_k} - 1\right) \right]^{j-1}}{\left(e^{-\beta} - 1\right)^{j(N-1)}} \\ c_j &= (-1)^{j-1}(j-1)! a_j \end{aligned}$$

and $a_j \equiv S(N,j)$ is the Stirling number of the second kind, where

$$S(N,j) = \frac{1}{j!} \sum_{k=0}^{j} (-1)^{j-k} \binom{j}{k} k^N$$

Using the above pdf's, and based on a null of the empirical distribution, maximum log-likelihood (ML) functions and parameter estimates of the different copulas are computed and reported as follows. The empirical pdf is simply one divided by the total number of sample points. To find the standard deviation of the ML estimates, we employ the jackknife method. This is useful for small samples as well as for the cases where simulation is not possible due to the unknown null distribution. The "z-values" are computed by dividing the $\hat{\beta}$ estimates by their standard deviations, and provide an indication of whether the estimates are significantly different from zeros. Sometimes, ML based on a null empirical distribution, or the semi-parametric approach, is also called a pseudo-ML where part or one of the parameters is estimated using an alternative consistent, asymptotic method instead of full ML due to reasons such as non-convergence or having an empirical distribution as marginal.

Table 1 Maximum likelihood estimates and Kullback-Leibler distances

Copula	$\hat{\beta}$	Jackknife Std. Dev.	"z-value"	KL distance	ML
Student-t	6	0	∞	5.68	-10,251
Clayton	1.15	0.017	67.6	3.72	-11.92
Gumbel	2.12	0.034	62.3	4.10	-1,982
Frank	4.29	0.052	82.4	4.29	-2,998

In Table 1, clearly the Clayton copula has the least KL distance or relative entropy to the empirical distribution. The t-copula is also found to be more distant and less consistent with the empirical distribution relative to all the other Archimedean copulas. The Clayton copula also has the highest maximum likelihood score. Using the estimated (pseudo)-ML parameter values, we provide the probability density function graphs on the three Archimedean copulas. The graphs depict only a 2-dimensional slice of the N-dimensional copula, but are representational in the characterization of the joint density effects. This is shown in Figure 2. From the fitted copulas as seen in Fig. 2, Clayton takes care of both loss and gain tails. Gumbel for the parameter loads on the positive tail with a bulk at the middle. Frank overloads and plateaus on the long side with much less density on the loss tail.

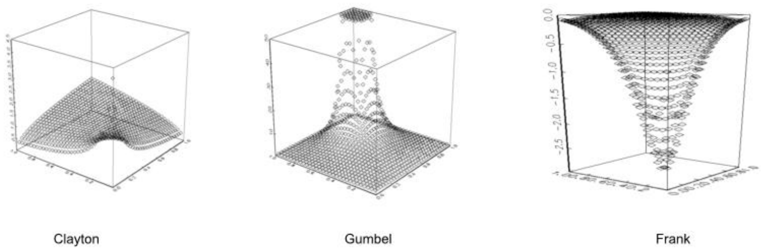


Fig. 2 PDF based on estimated copula parameter

3 Loss Tail Model

We study the loss tails of the marginal return distributions in more detail. The Pareto Type II distribution is used to model the conditional returns falling below one standard deviation away from the mean. The use of the Pareto class of distributions to model fat and long tails of distributions have been documented in many statistical studies such as [12], [15], [10], [13], [4], [1], and so on.

In a series of articles, viz. [7], [8], [9], [5], and [16], the tail dependence of multivariate distributions using copula method has been characterized via methods of regular variations and mixtures of scales. However, there were no empirical studies to validate the models or study situations of structural changes in the parameters in a time series setting. There have been several recent applications of the multivariate Pareto distribution. [14] provided an application of the multivariate Pareto distribution to computing the tail conditional expectation, or sometimes called conditional Value-at-Risk in risk measurement literature. They provided formulas for the conditional expectation, but there is no study of their statistical properties. [2] provided one of the earliest study of country index return correlations using the copula method, but used the non-central t-distribution as univariate density instead, and it did not provide statistical tests of the parametric estimators.

We define any end-of-day t return to a country market as the natural logarithm of the index level price relative over a day: $r_t = \ln(P_t/P_{t-1})$ where P_t is the index level at day t. The loss tail for each index is the set of negative daily returns below a threshold u that is fixed as the sample estimate of $E(\ln(P_t/P_{t-1}) - \sqrt{\operatorname{var}(\ln(P_t/P_{t-1}))}$ where $u < 0$ in all cases. We define the conditional negative returns in the loss region as $r_t < u < 0$. Let $z_t = |r_t - u| > 0$, the exceedance of loss beyond threshold or bound u.

The Pareto Type II (sometimes also called the Lomax distribution) decumulative distribution function (ddf) of z_t is $\bar{F} = \left(1 + \frac{z_t}{\gamma}\right)^{-\alpha}$, where $\alpha, \gamma > 0$, and $z_t > 0$. The (cumulative) distribution function of z_t is thus $F(z_t) = 1 - \bar{F} = 1 - \left(1 + \frac{z_t}{\gamma}\right)^{-\alpha}$. It is in the class of Pearson system Type IV distribution. It can also be expressed as $F(z_t) = 1 - \left(1 + \frac{\xi z_t}{\sigma}\right)^{-1/\xi}$, where $\xi = 1/\alpha$ and $\sigma = \gamma/\alpha$. The constants u, σ, and ξ are also called location, scale, and shape parameters for this distribution. The shape parameter affects the curvature and thinness of the tail. The scale as well as the shape parameters affect the fatness of the tail. Within the class of Pareto distributions, the Pareto II distribution is well suited and more general than Pareto Type I in modeling tail losses of returns. Figure 3 shows the pdf and the cdf of a Pareto Type II distribution. The tail of the pdf thickens or increases with higher γ and lower α values. To study contagion, we consider only the conditional joint tail losses of the joint multivariate return distributions. Using the copula to model multivariate distribution, we therefore model each loss tail as a marginal distribution. In particular we perform an empirical distributional fit using the Pareto Type II distribution conditional

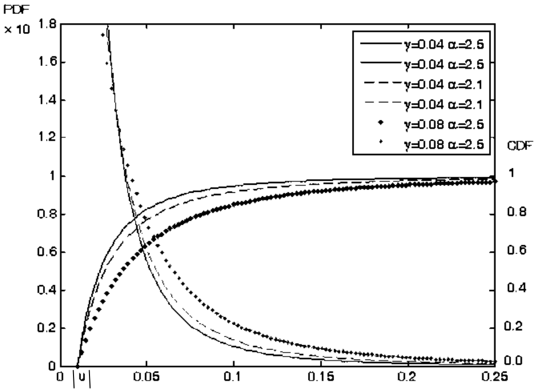


Fig. 3 PDF and CDF of Pareto Type II distributions of tail loss $z_t = |r_t - u|$ with parameters γ and α

on loss exceeding a bound that has been taken to be one standard deviation into the loss region away from the mean of the sample of returns for each index. Experimenting with slightly different bounds or location parameters does not produce any material qualitative differences.

The marginal Pareto II distribution parameters of $\hat{\gamma}_i$ and $\hat{\alpha}_i$ for $i = 1,2,3$, could be estimated using a full ML approach. However, the first order conditions are sensitive to small variations in data and do not converge at times to a solution. To avoid computational errors, we perform a pseudo-ML estimation by first estimating the $\hat{\gamma}$ via minimizing the Kolmogorov-Smirnov distance D between the theoretical CDF curve and the empirical CDF curve. We obtain the following estimates of $\hat{\gamma}$ for the shapes of the loss tails of the U.S., U.K., and Japanese stock index returns respectively: 0.0439, 0.0940, 0.0229. The fitted CDF curves theoretically converge in distribution, and the $\hat{\gamma}$ estimates are asymptotically consistent.

For each stock index i, the maximum likelihood estimates of α can be obtained as follows, given the estimates of γ obtained from the empirical fit above. The log-likelihood of the sample is expressed as

$$\begin{aligned} L &\equiv \ln f(z_{i,1}, z_{i,2}, \ldots, z_{i,t}, \ldots, z_{i,T}) \\ &= \ln \prod_{t=1}^{T} f(z_{i,t}) = \sum_{t=1}^{T} \ln f(z_{i,t}) \\ &= \sum_{t=1}^{T} \ln \left(\frac{\alpha_i}{\gamma_i} \left[1 + \frac{z_{i,t}}{\gamma_i} \right]^{-(\alpha_i+1)} \right) \quad , z_{i,t} > 0 \ , \ \forall i,t. \end{aligned} \tag{4}$$

where $f(z_{i,1}, z_{i,2}, \ldots, z_{i,t}, \ldots, z_{i,T})$ indicates the time series multivariate pdf of the sample $\{z_{i,1}, z_{i,2}, \ldots, z_{i,T}\}$. We assume independence of $z_{i,t}$ across time.

We maximize the likelihood function L in equation (4) and obtain the following first order optimality condition in equation (5). The second order condition is also satisfied.

$$\frac{\partial}{\partial \alpha_i}: \quad \frac{T}{\alpha_i} = \sum_{t=1}^{T} \ln\left(1 + \frac{z_{i,t}}{\hat{\gamma}_i}\right) \tag{5}$$

The empirical results for the estimates of $\hat{\gamma}_i$ and $\hat{\alpha}_i$ are reported in Table 2 below.

Table 2 Estimates of tail parameters under Pareto II distributions

Parameters	U.S. S&P 500	U.K. FTSE 100	Japan N225
Lower bound	0.0099	0.0112	0.0150
K-S D-statistic	0.0459	0.0663	0.0722
$\hat{\gamma}$	0.0439	0.0940	0.0229
$\hat{\alpha}$	3.4826	7.5707	2.1180
Mean $\frac{\hat{\gamma}}{\hat{\alpha}-1}$	0.0177	0.0143	0.0205
Std.Dev. $\frac{\hat{\gamma}}{\hat{\alpha}-1}\sqrt{\frac{\hat{\alpha}}{\hat{\alpha}-2}}$	0.0271	0.0167	0.0868

Table 2 shows that the Japanese stock index returns exhibited higher mean but more volatility relative to the other two indices. The $\hat{\alpha}$ estimates that are all greater than 2 also indicate that all the conditional loss tails possessed means and variances. The mean and standard deviation estimates were computed based on the Pareto Type II distribution. Actual sampling estimates are all approximately the same except for a lower sampling standard deviation in the N225 case.

4 Estimation Results on the Copulas

The conditional multivariate joint distributions are next developed using the Clayton, Gumbel and the Frank copulas. The maximum likelihood approach is again employed for estimating the copula parameters. After removing all dates whereby the joint returns did not exist due to holidays in the countries, and conditioning on all the countries' returns in the loss region beyond the thresholds u_1, u_2, and u_3 respectively for S&P500, FTSE100, and N225, a smaller sample of size 89 is left. Thus contagion effect occurred in less than 1.4% of the number of trading days.

Let $u_{i,t} = F(z_{i,t})$ where $i = 1,2,3$. The values of i being 1,2, or 3 represent U.S. S&P 500, U.K. FTSE100, and Japan N225 respectively. Using the Clayton, Gumbel, and Frank pdf's exposited in equations (1), (2), and (3), we can construct the respective log-likelihood functions of the different copulas. The maximum log-likelihoods of the respective copulas with the respective maximum likelihood estimates, $\hat{\beta}_i$ can be obtained. Due to the highly nonlinear copula function and its related first and second-order derivatives, a full maximum likelihood computation is highly sensitive to small sample errors and in our case produces rather unstable results. The $\hat{\beta}_i$'s are therefore obtained based on a two-stage procedure or inference for margin approach (see [6]) whereby $\hat{u}_i$, for $i = 1,2,\ldots,N$, is first obtained in stage one, based on estimates $\hat{\gamma}_i$ and $\hat{\alpha}_i$ as obtained earlier. The two-stage process under correct marginal distribution specification should produce consistency and asymptotically efficient estimates when sample size $T \to \infty$. The two-stage procedure is also discussed in [3].

The maximum likelihood estimates of the different copula parameter β_i's and the associated Kendall's τ statistics for the different copulas are shown below in Table 3. The results indicate that the Gumbel and Frank copulas tend to produce very high association estimates relative to the Clayton copula. This does not mean that they are more appropriate copulas. It means that if they are true, then they tend to capture higher associations. We therefore have to evaluate and compare them as in a horse race how well the copulas performed relative to the actual empirical data.

Table 3 Maximum likelihood estimates of the copulas

Estimates	Clayton	Gumbel	Frank
$\hat{\beta}_i$	0.4001	2.4782	4.3295
Kendall's τ	0.1667	0.5964	0.4121

In comparing the empirical performance of the copulas, a natural intuitive test would be the goodness-of-fit test with an asymptotic χ^2 test statistic. The observed joint returns beyond the loss boundary of at least one standard deviation away from the mean are matched in frequencies across 27 three by three cells with the theoretical probabilities under the competing copulas. Each domain of u_i for $i = 1,2,3$ is divided into 3 equal regions, $\hat{u}_i \in (0,\frac{1}{3}]$, $\hat{u}_i \in (\frac{1}{3},\frac{2}{3}]$, and $\hat{u}_i \in (\frac{2}{3},1]$, for $i = 1,2,3$. Table 4 shows the goodness-of-fit test statistic that is χ^2 with 26 degrees of freedom. At 1% significance level, the test cannot reject the null of Clayton copula while it clearly rejects the Gumbel copula, the Frank copula, and the joint multivariate uniform distribution. The goodness-of-fit test shows that the Clayton copula seems to be the most appropriate copula in explaining contagion during this sampling period.

Table 4 Goodness-of-fit tests of the theoretical copulae

Hypothesized Distribution	Asymptotic χ^2 d.f. 26	p-Value
Clayton copula	41.0	0.0310
Gumbel copula	292.7	0.0000
Frank copula	76.2	0.0000
Uniform	53.3	0.0012

5 Conclusions

There is a relative scarcity of rigorous studies using empirical data and advanced econometric methods to verify structural changes and systemic risk in the global financial markets. In this paper, we first employ (pseudo)-maximum likelihood method and a measure of relative entropy considering copulas and their estimated Kullback-Leibler distances from the empirical cdf of index return distributions. It is found that for the entire index return distributions of the S&P 500, the FTSE100, and N225 during the 1990 to early 2012 sample period, the Archimedean class of copulas capture the joint distribution of returns better than a benchmark t-copula based on univariate t-distributions. The Clayton copula shows the least relative entropy with respect to the empirical distribution.

We then employ the Pareto Type II distribution to model the conditional tail loss distributions of the respective index returns. Unlike most other empirical studies, we concentrate on the tail loss as it is the area most critical to the risk-taking and risk management decisions made by banks and financial institutions. It is also the area most concerned by regulatory bodies worldwide and in each of advanced economies with a mature financial market subject to sharp losses and contagion risk. One practical, not theoretical, disadvantage of using conditional distribution in modeling rather than the unconditional distribution is that when it comes to estimation, the sample size will be smaller due to the condition or restriction, and thus small sample bias may become more of an issue.

To study contagion based on the conditional tail losses, the joint multivariate joint distributions using the Clayton, Gumbel and the Frank copulas are estimated using the pseudo-maximum likelihood approach. Due to the sensitivity of full-scale maximum likelihood computations in its outcomes based on small samples, we used a two-stage approach in first estimating the parameters of the individual conditional loss tails and then estimating the copula parameter given the estimated univariate tail parameters. The observed joint returns beyond the loss boundary of at least one standard deviation away from the mean are then matched in frequencies across 27 three by three cells with the theoretical probabilities under the competing copulas. A goodness-of-fit test shows that the Clayton copula seems to be the most appropriate copula in explaining contagion during this sampling period. Using the copula and univariate tail distributional models via maximum likelihood estimation is evidenced as a reasonable approach to modeling and estimating contagion risk.

Acknowledgments. This work was supported by the OUB Chair professorship budget at the Singapore Management University. The author wishes to thank participants in earlier presentations of this work at the Sim Kee Boon Institute Princeton/QUT/SMU Tripartite Conference on Financial Econometrics, and the 9th International Conference on Applied Financial Economics, Samos Island, for valuable comments.

References

1. Asimit, A.V., Furman, E., Vernic, R.: On a multivariate Pareto distribution. Insurance: Mathematics and Economics 46, 308–316 (2010)
2. Cazzulani, L., Meneguzzo, D., Vecchiato, W.: Copulas: A Statistical Perspective with Empirical Applications. Working Paper, Intesa BCI Risk Management Department (2001)
3. Cherubini, U., Luciano, E., Vecchiato, W.: Copula methods in finance. Wiley Finance, John Wiley & Sons (2004)
4. Holger, R., Tajvidi, N.: Multivariate generalized Pareto distributions. Bernoulli 12, 917–930 (2006)
5. Joe, H., Li, H.: Tail Risk of Multivariate Regular Variation. Methodology and Computing in Applied Probability 13, 671–693 (2011)
6. Joe, H., Xu, J.J.: The Estimation Method of Inference Functions for Margins for Multivariate Models. Dept. of Statistics, University of British Columbia. Technical Report 166 (1996)
7. Li, H.: Tail Dependence of Multivariate Pareto Distribution. Dept. of Mathematics and Dept. of Statistics, Washington State University at Pullman, U.S.A. Technical Report 6 (2006)
8. Li, H.: Orthant tail dependence of multivariate extreme value distributions. Journal of Multivariate Analysis 100, 243–256 (2009)
9. Li, H., Sun, Y.: Tail dependence for heavy-tailed scale mixtures of multivariate distributions. Journal of Applied Probability 46, 925–937 (2009)
10. Mardia, K.V.: Multivariate Pareto distributions. Annals of Mathematical Statistics 33, 1008–1015 (1962)
11. Nelsen, R.B.: Copulas and Association. In: Dall'Aglio, G., Kotz, S., Salinetti, G. (eds.) Advances in Probability Distributions with Given Marginals. Kluwer Academic Publishers, Dordrecht (1991)
12. Pareto, V.: Cours d'Economie Politique. Rouge and Cie, Lausanne and Paris (1897)
13. Resnick, S.I.: Extreme values, Regular Variation and Point Processes. Springer, Berlin (1987)
14. Raluca, V.: Tail Conditional Expectation for the Multivariate Pareto Distribution of the Second Kind: Another Approach. Methodology and Computing in Applied Probability 13, 121–137 (2011)
15. Yeh, H.-C.: Some properties and characterizations for generalized multivariate Pareto distributions. Journal of Multivariate Analysis 88, 47–60 (2004)
16. Yin, C., Li, H.: Tail dependence for multivariate t -copulas and its monotonicity. Insurance: Mathematics and Economics 42, 763–770 (2008)

A Bayesian Perspective on Mixed GARCH Models with Jumps

Cathy W.S. Chen, Edward M.H. Lin, and Yi-Ru Lin

Abstract. In this paper, generalized autoregressive conditionally heteroskedastic (GARCH) models with jumps are investigated, where jump arrivals are time inhomogeneous and state-dependent. These models permit the conditional jump intensity to be time-varying and clustering, and allow volatility effects in the jump component. A Bayesian approach is taken and an efficient adaptive sampling scheme is employed for inference. A Bayesian posterior model comparison procedure is used to compare the proposed model with the standard GARCH model. The proposed methods are illustrated using both simulated and international stock market return series. Our results indicate that the mixed GARCH-Jump models provide a better fit for the dynamics of the daily returns in the US and two Asian markets.

1 Introduction

Due to advances in technology, investors have more efficient ways to obtain new information. With the opening of the international financial markets, the lifting of investment restrictions, the occurrence of capital fast flow, and the occurrence of unexpected economic events of world economic powers, the entire world market facilitates large volatility. When unanticipated or anticipated events occur, the latent news can all cause stock market volatility, resulting in price volatility. The biggest factor is the impact of potential news. Latent news can be divided into two parts: usual news and unusual news. Usual news assumes the conditional variance of return, resulting in smooth changes. Unusual news causes a drastic change in return. These impact the market. Unusual information is defined as a jump.

The family of autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) models, initiated by [10] and [2], respectively, have permitted conditional variance in time series to be modeled so that the magnitude of

Cathy W.S. Chen
Department of Statistics, Feng Chia University, Taiwan
e-mail: chenws@mail.fcu.edu.tw

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 141–154.
DOI: 10.1007/978-3-642-35443-4_10

volatility can be predicted from past news and lagged conditional variance. However, GARCH models do not presume the existence of jumps for financial asset prices. Nowadays, some stylized facts have emerged about the statistical behavior of many financial asset prices. They possess jumps, therefore, and the importance of considering jumps into models has been widely discussed in finance literature. In volatility dynamics, it is conventional wisdom that GARCH and stochastic volatility (SV) models provide a good approximation of these stylized facts by modeling the autoregressive structure in the conditional variance. GARCH and SV models can be devised to capture smooth and persistent changes in volatility. Most returns are measured in discrete time which suggests that jumps provide a natural framework to model price moves. [12] develop a mixed GARCH-Jump model incorporating the autoregressive conditional jump intensity parametrization proposed by [4].

Even though it shows inhomogeneous jump dynamics and volatility structures for the model, it does not explicitly specify how the jump dynamics depend on the state of asset prices. [9] propose several asymmetric GARCH-Jump models that mix time varying autoregressive jump intensities and volatility feedback in the jump component. They also extend the [12] specification by allowing the jump intensity to be both autoregressive and dependent on the return volatility or its proxy. In this paper, we make inference of the mixed GARCH-Jump model via Bayesian framework, that extends the [12] structure to consider not only time-varying jump intensities but also volatility effects in the jump dynamics component.

For a mixed GARCH-Jump model with large numbers of parameters, the Bayesian and Markov chain Monte Carlo (MCMC) methods can provide a more efficient way to estimate numbers of parameters in this model. [13] employ Bayesian approach modelling foreign exchange rates with jumps. We adopt the Bayesian approach using MCMC algorithms and apply the random walk and independent kernel Metropolis-Hastings (MH, [14], [11]) algorithms to estimate the unknown parameters. In the past, Bayesian methods have proven successful inference for many heteroscedastic models, such as: the GARCH models [1]; EGARCH models [16]; the asymmetric GARCH models [5]; and the Markov switching GARCH models [7] etc. Bayesian MCMC methods have the further advantage of being valid under the stationarity and positivity parameter constraints usually required for such models, and the ability to do joint finite sample inference on all model parameters.

The mixed GARCH-Jump model takes into account the jump size. Most return on capital has discrete jumps: modeling the GARCH-Jump model better captures statistical characteristics of the jump discontinuity. We also employ a formal Bayesian posterior model comparison procedure to compare the proposed model with competing models. In this paper, we adapt two approaches suggested by [15] and [8] to calculate marginal likelihood.

The rest of this study is structured as follows. Section 2 describes a mixed GARCH-Jump model with autoregressive jump intensity and volatility effects in the jump structure; Section 3 discusses model estimation via the Bayesian inference. Model selection is described in Section 4. A simulation study and three daily stock indices are conducted for illustration in Sections 5 and 6. The concluding remarks are provided in Section 7.

2 The Mixed GARCH-Jump Model

[2] generalizes [10]'s ARCH model by allowing the conditional variance to be a function of lagged errors and conditional variance itself. We follow [3] by employing a GARCH(1,1) process:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta_{t-1}, \tag{1}$$

where h_t is the conditional variance; ω, α, and β are unknown parameters; and ε_t is a random error term. [12] specify the components of returns in which two stochastic innovations, $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$, drive returns.

$$R_t = \mu + \phi_1 R_{t-1} + \varepsilon_{t,1} + \varepsilon_{t,2}, \tag{2}$$

where $\varepsilon_{t,1}$ is a mean-zero innovation with a normal stochastic process, $\varepsilon_{t,2}$ is a jump innovation specified so that it is also conditionally mean-zero, and $\varepsilon_{t,1}$ is contemporaneously independent of $\varepsilon_{t,2}$. The component $\varepsilon_{t,1}$ is:

$$\varepsilon_{t,1} = \sqrt{h_t} Z_t\ ,\ Z_t \sim \mathrm{N}(0,1), \quad h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{3}$$

where h_t is assumed to follow a GARCH(1,1) process and $\varepsilon_t = R_t - \phi_1 R_{t-1} - \mu$, $t = 1,\ldots,T$. The specification of ε_t contains the expected jump component and thus allows it to propagate and affect future volatility through the GARCH variance factor. In order to ensure positive variance and stationarity, we set the following standard restrictions on the variance parameters:

$$\omega, \alpha, \beta > 0 \text{ and } \alpha + \beta < 1. \tag{4}$$

Let N_t denotes the number of jumps that arrive between $t-1$ and t, among which the conditional jump size of the k-th jump is $Y_{t,k}$ given the history of returns. The jump component affecting returns period t is $J_t = \sum_{k=1}^{N_t} Y_{t,k}$. Hence, the jump innovation associated with period t is expressed as

$$\varepsilon_{t,2} = J_t - E(J_t | \mathscr{F}_{t-1}) = \sum_{k=1}^{N_t} Y_{t,k} - \theta \lambda_t,$$

where $Y_{t,k} \sim \mathrm{N}(\theta, \delta^2)$, $P(N_t = j | \mathscr{F}_{t-1}) = \lambda_t^j \exp(-\lambda_t)/j!$, $j = 0,1,2,\cdots$. The conditional jump intensity λ_t, which controls the jump dynamics, is assumed to follow the autoregressive process:$\lambda_t = \nu_0 + \nu_1 \lambda_{t-1} + \tau \varepsilon_{t-1}^2$, where ν_0 controls the expected number of jumps per day, ν_1 controls the persistence of jump-clustering effect, and τ controls the time-varying effects in jump dynamics. A sufficient condition given in [12] for $\lambda_t > 0$ is $\nu_0 > 0$, $\nu_1 \geq \tau$, and $\tau \geq 0$. For the conditional jump intensity to be well defined, the following parameters are set:

$$\nu_0, \tau \geq 0 \text{ and } 0 \leq \nu_1 < 1. \tag{5}$$

Under the stationarity specification in (5), the unconditional jump intensity is equal to $E(\lambda_t) = (\nu_0 + \tau E(h_t))/(1-\nu_1)$, where $E(h_t) = \omega/(1-\alpha-\beta)$ is the unconditional volatility of R_t. In summary, the mixed GARCH-Jump model in this paper is as follows:

$$\begin{aligned} R_t &= \mu + \phi_1 R_{t-1} + \varepsilon_t, \quad \varepsilon_t = \varepsilon_{t,1} + \varepsilon_{t,2}, \\ \varepsilon_{t,1} &= \sqrt{h_t} Z_t, \quad h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \\ \varepsilon_{t,2} &= \sum_{k=1}^{N_t} Y_{t,k} - \theta\lambda_t, \quad \lambda_t = \nu_0 + \nu_1\lambda_{t-1} + \tau\varepsilon_{t-1}^2, \\ Z_t &\sim \mathrm{N}(0,1), \quad N_t|\mathscr{F}_{t-1} \sim \mathrm{Poisson}(\lambda_t), \quad Y_{t,k} \sim \mathrm{N}(\theta, \delta^2). \end{aligned} \tag{6}$$

Bayesian MCMC methods have a number of advantages in estimation and inference. They account for parameter uncertainty in both probabilistic and point forecasting; provide exact inference for finite samples; efficiently and flexibly handle complex models and non-standard parameters; and supply efficient and valid inference under parameter constraints. Therefore, MCMC methods were generally used to estimate the mixed GARCH-Jump model in this paper.

3 Bayesian Inference

Bayesian methods usually require the specification of a likelihood function and prior distributions on model parameters. Define $\phi = (\mu, \phi_1), \alpha = (\omega, \alpha, \beta)$ and $\nu = (\nu_0, \nu_1, \tau)$ as the vectors of the mean, variance, and jump size parameters, respectively, and denote the vector of all parameters $\Theta = (\phi, \alpha, \nu, \theta, \delta^2)$. The ε_t given that j jumps occur and the information set $\mathscr{F}_{t-1}$ follows a Gaussian distribution:

$$\begin{aligned} &\varepsilon_t = R_t - \phi_1 R_{t-1} - \mu, \quad t = 1, \ldots, T, \\ &\varepsilon_t|(N_t = j, \mathscr{F}_{t-1}) \sim \mathrm{N}(\mu_{\varepsilon_t}, \sigma_{\varepsilon_t}^2), \quad N_t|\mathscr{F}_{t-1} \sim \mathrm{Poisson}(\lambda_t). \end{aligned}$$

Let $R = (R_1, \ldots, R_T)$, when jump is introduced the likelihood function become complicated, it may be non-differentiable in the parameters. By some conditioning parameters, the likelihood function can be written as follows:

$$\begin{aligned} \mathscr{L}(R|\Theta) &= \prod_{t=1}^{T} \sum_{j=0}^{\infty} f(\varepsilon_t|N_t = j, \mathscr{F}_{t-1}) p(N_t|\mathscr{F}_{t-1}) \\ &= \prod_{t=1}^{T} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_t}^2}} \exp\left\{ -\frac{(\varepsilon_t - \mu_{\varepsilon_t})^2}{2\sigma_{\varepsilon_t}^2} \right\} \cdot \frac{\lambda_t^j e^{-\lambda_t}}{j!}, \end{aligned} \tag{7}$$

where $\mu_{\varepsilon_t} = j\theta - \theta\lambda_t$, and $\sigma_{\varepsilon_t}^2 = \sigma_{t,1}^2 + j\delta^2$.

Prior distributions on the parameters must be assumed in Bayesian inference, in this paper, the prior distribution is chosen to be reasonably uninformative over the possible region. For ϕ in (2), the parameters of the mean equation of the GARCH-Jump model, we assume a bivariate normal prior for ϕ:

$$\pi(\phi) \sim \mathrm{N}\left(\begin{bmatrix} m_\mu \\ m_{\phi_1} \end{bmatrix}, \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_{\phi_1}^2 \end{bmatrix} \right),$$

where $\sigma_\mu^2 > 0$ and $\sigma_{\phi_1}^2 > 0$. For θ, the mean of the jump size, we assume a normal distribution prior, $\theta \sim \mathrm{N}(m_\theta, \sigma_\theta^2)$. We assume that the variance of the jump size, δ^2, is an inverse gamma distribution, $\delta^2 \sim \mathrm{IG}(\mathrm{a}, \mathrm{b})$. In order to satisfy the necessary constraints on parameters and ensure a proper posterior, for the parameters α and ν follow constrained uniform priors, defined by indicators $\mathrm{I}(S_1)$ and $\mathrm{I}(S_2)$, respectively, where S_1 is the set α that satisfies (4), and S_2 is the set ν that satisfies (5).

Let γ be one of the parameters in Θ and $\pi(\gamma|R, \Theta_{-\gamma})$ denote the posterior distribution, where $\Theta_{-\gamma}$ is the vector of all parameters without γ. MCMC methods require conditional posterior distribution for each choice of γ. In each case, the target posterior is:

$$\pi(\gamma|\Theta_{-\gamma}) \propto \mathscr{L}(R|\Theta)\pi(\gamma). \tag{8}$$

By using Bayes rule, the posterior distribution in (8), for each choice of γ does not have known or standard forms in the parameters. Such as the parameters ϕ of the mean equation, the full conditional posterior is written by

$$\pi(\phi|\Theta_{-\phi}) = \prod_{t=1}^{T} \sum_{j=0}^{\infty} f(\varepsilon_t|N_t = j, \mathscr{F}_{t-1}) p(N_t|\mathscr{F}_{t-1}) \pi(\phi). \tag{9}$$

Note that Equation (9) looks like a bivariate normal distribution but it is a non-standard from actually. For θ, the mean of normal distribution of the jump size, the full conditional posterior can be written by

$$\begin{aligned} \pi(\theta|\Theta_{-\theta}) &= \prod_{t=1}^{T} \sum_{j=0}^{\infty} f(\varepsilon_t|N_t = j, \mathscr{F}_{t-1}) p(N_t|\mathscr{F}_{t-1}) \pi(\theta) \\ &= \prod_{t=1}^{T} \sum_{j=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_t}^2}} \exp\left\{ -\frac{(\varepsilon_t - \mu_{\varepsilon_t})^2}{2\sigma_{\varepsilon_t}^2} \right\} \frac{\lambda_t^j e^{-\lambda_t}}{j!} \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left\{ -\frac{(\theta - \mu_\theta)^2}{2\sigma_\theta^2} \right\} \end{aligned} \tag{10}$$

For δ^2, the variance of normal distribution of the jump size, the full conditional posterior for δ^2 with a prior IG(a,b) is written by

$$\pi(\delta^2|\Theta_{-\delta^2}) \propto \prod_{t=1}^{T} \sum_{j=0}^{\infty} (\delta^2)^{(-\mathrm{a}-\frac{1}{2})-1} \exp\left\{ -\frac{1}{2} \left[\frac{2\mathrm{b}}{\delta^2} + \frac{\varepsilon_t^2 - 2j\theta\varepsilon_t + 2\theta\lambda_t\varepsilon_t + j^2\theta^2 - 2j\theta^2\lambda_t + \theta^2\lambda_t^2}{\sigma_{t,1}^2 + j\delta^2} \right] \right\}. \tag{11}$$

The conditional posteriors for each parameter group are non-standard. We therefore incorporate the Metropolis-Hastings (MH) methods to draw the MCMC iterates for the parameter groups. To speed convergence and allow optimal mixing, we employ an adaptive MH-MCMC algorithm that combines a random walk Metropolis and an independent kernel MH algorithm. The MCMC algorithm for the mixed GARCH-Jump models is described as follows:

Step1: Set $l = 0$ and specify an initial value for $(\phi, \alpha, \nu, \theta, \delta^2)$.
Step2: For $j = 1, \ldots, g$,

Step 2a: Sample $\phi|\Theta_{-\phi}$ from (9).
Step 2b: Sample α from $\mathscr{L}(R|\Theta)I(S_1)$.
Step 2c: Sample ν from $\mathscr{L}(R|\Theta)I(S_2)$.
Step 2d: Sample $\theta|\Theta_{-\theta}$, from (10).
Step 2e: Sample $\delta^2|\Theta_{-\delta^2}$, from (11).

Step 3: Set $l = l + 1$ and go to Step 2.

The Gibbs sampler is repeated for N+M iterations from Step 2 to Step 3, which contain M burn-in iterations and N iterations to estimate posterior parameters. We first use the random walk algorithm from the first M iterations and then use the independent kernel MH algorithm algorithm from iteration $M + 1$ onwards. A detailed description of MH algorithm is provided by [6].

4 Model Selection

Model comparison and selection are important issues in practice. A common approach to compare the k models in the Bayesian setting is to use the marginal likelihood, which, for model M_k, is $f(R|M_k)$. In this paper, we use the approaches of [15] (NR hereafter) and [8] to calculate the marginal likelihoods. NR uses the posterior distribution, $\pi(\theta|R, M_k)$, as the important sampling distribution to numerically evaluate the marginal likelihood,

$$M_{lik}(M_k) = \int f(R|\theta, M_k)\pi(\theta|M_k)d\theta.$$

The logarithmic marginal likelihood can then be approximated by the following importance sampling estimator:

$$\log\widehat{M_{lik}}(M_k) = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\log f(R|\theta^{(i)}, M_k)\right)^{-1}\right]^{-1},$$

where $\theta^{(i)}$ is an MCMC iteration which draws from the posterior distribution $\pi(\theta|R, M_k)$. When the harmonic mean of the marginal log likelihood value is large, its estimated result is better.

Another method we adopt is an approach by [8] for computing the model to obtain the logarithm of marginal likelihood:

$$\log M_{lik}(M_k) = \log f(R|\theta, M_k) + \log\pi(\theta, M_k) - \log\pi(\theta|R, M_k), \tag{12}$$

where θ is the parameter set in the M_k model, $f(R|\theta, M_k)$ is the likelihood function, $\pi(\theta, M_k)$ is the prior probability density, and $\pi(\theta|R, M_k)$ is the posterior density.

Let θ^* denote a set of posterior means of unknown parameters. Using Chib's method, the logarithm of marginal likelihood in Eq. (12) can be estimated by substituting θ^*. Therefore, we are able to calculate the log likelihood estimate from the likelihood function of the GARCH-Jump model, and obtain the value of $\log\pi(\theta^*, M_k)$ from each prior setting of unknown parameters. Since the full conditional posterior density for the GARCH-Jump model is unknown, we apply the kernel density estimate of a non-parametric approach to approximate $\log\pi(\theta^*|R, M_k)$:

$$\log \hat{\pi}(\theta^*|R, M_k) \cong \log \left[(nh^d)^{-1} \sum K\left((\theta^* - \theta_i)/h\right)\right],$$

where $h > 0$ is a bandwidth, d is the dimension of space, and $K(\cdot)$ is a Gaussian kernel function. A common method of choosing the h is to minimize the asymptotic mean integrated squared error (AMISE). Here, we implemented an empirical rule for choosing the bandwidth of a Gaussian kernel density estimator.

5 Simulation Study

To examine the effectiveness of the MCMC sampling scheme in the Bayesian inference, we conduct a simulation study from a GARCH-Jump model and considering the sample size $n = 2,000$. For the model in Eq. (6), 100 data sets were generated and then used for analysis.

For unknown model parameters, α, ν, we consider $\pi(\alpha, \nu) \propto 1$. The prior for θ is chosen as a normal distribution, and that for δ^2 is chosen as a inverse gamma distribution; these priors are described in Section 3. For the hyper-parameters, we choose $(m_\theta, \sigma_\theta^2) = (0, 0.1)$ and (a, b) $= (4, 0.06)$. For the initial values, we choose (μ, ϕ_1)=(0, 0), (ω, α, β)=(0.15, 0.1, 0.1), (ν_0, ν_1, τ)=(0.01, 0.01, 0.01), $\theta = 0$, and $\delta^2 = 0.01$.

The Gibbs sampler is run for a total of $N + M = 20,000$ MCMC iterations for each data set, using the random walk MH for the first $M = 10,000$ burn-in iterations, and the independent kernel MH algorithm for the next $N = 10,000$ iterations. Convergence is monitored heavily using trace and ACF plots, through which, we observed that the adaptive method can indeed lead to reduced auto-correlations for MCMC iterates. These plots are not provided in this paper in order to save space, but are available upon request. For the setting of the jump number, we add a total of four in our computational task. It is easy to increase the jump number in our FORTRAN codes.

Table 1 provides summary statistics of parameter estimates averaged over 100 replications, including true values, posterior means, medians, standard deviations, and 2.5 and 97.5 percentiles. All of the means of the estimates are close to their true values. However, the average of standard deviation ($\cong 0.20$) is slightly large for ν_1. The histograms of 100 posterior mean estimates and true values in vertical lines are given in Figure 1. The simulation results indicate that the Bayesian method gives sound inference except for ν_1.

Table 1 Parameter estimates from 100 replications for the GARCH-Jump model

	True	Mean	Median	Std	2.5%	97.5%
μ	0.05	0.0508	0.0509	0.0243	0.0034	0.0984
ϕ_1	0.10	0.1226	0.1225	0.0233	0.0769	0.1683
ω	0.05	0.0760	0.0721	0.0271	0.0342	0.1413
α	0.07	0.0799	0.0788	0.0161	0.0517	0.1154
β	0.88	0.8553	0.8591	0.0339	0.7775	0.9111
ν_0	0.10	0.1038	0.0843	0.0824	0.0036	0.3061
ν_1	0.30	0.2773	0.2374	0.2038	0.0107	0.7447
τ	0.03	0.0337	0.0285	0.0256	0.0013	0.0949
θ	0.00	0.0001	0.0001	0.0995	-0.1950	0.1948
δ^2	0.02	0.0195	0.0163	0.0116	0.0069	0.0522

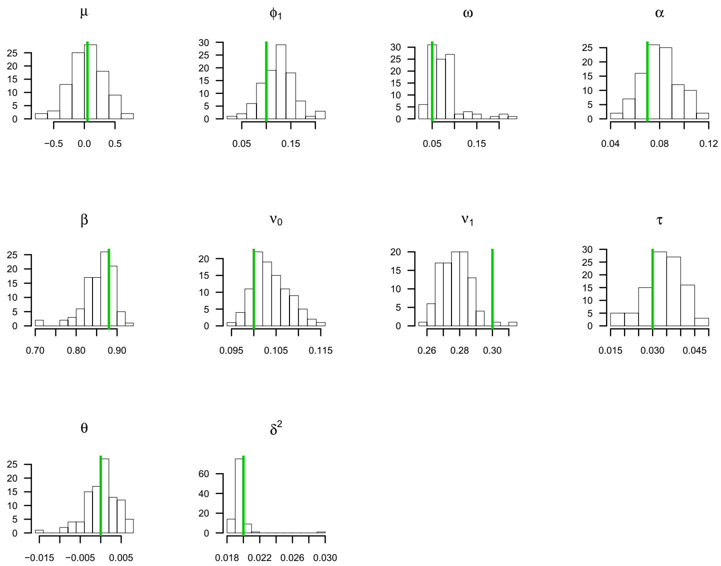


Fig. 1 Histograms with 100 replications. The vertical line in each plot is the true value of the parameter.

Table 2 Summary statistics of daily stock returns

Data	No obs.	Min	Q1	Median	Mean	Q3	Max	Std	Skewness	Excess Kurtosis
S&P500	2437	-9.4695	-0.5643	0.0776	0.0079	0.6104	10.9572	1.9407	-0.1990	8.2978
Nikkei 225	2375	-12.1110	-0.7891	0.0363	-0.0130	0.8701	13.2346	2.5250	-0.5173	7.9249
HSI	2395	-13.5820	-0.6769	0.0572	0.0235	0.7946	13.4068	2.6804	0.0519	8.7794

6 Empirical Study

We illustrate our methods using three daily stock market indices obtained from Datastream International for the period from May 24, 2002, to January 31,2012: the Nikkei 225 of Japan, the HSI (HANG SENG Index) of Hong Kong, and the S&P500 of the US. Daily log returns are calculated as $R_t = (\log(p_t) - \log(p_{t-1})) \times 100$, where p_t is the price index at time t. To provide a general understanding of the nature of each market return, summary statistics of daily returns are presented in Table 2. The returns range from -13.58% to 13.41%, but their mean returns are all close to zero. In terms of variance, the S&P500 has the smallest value, 1.94, which implies that it is the most stable stock market. The HSI indicates that Hong Kong's stock market is a high-risk market. For the S&P500 and Nikkei 225, the negative skewness, which indicates that the extreme values happen at low quantile, thus causing a longer left tail and high excess kurtosis value, shows that the data contains extreme negative values. Figure 2 shows the time series plots of the three markets. It is clear that all series of returns were more volatile during the global financial crisis in 2008-2009.

The mixed GARCH-Jump model with the first-order autoregressive process in Eq. (6) is used to investigate the structure of dynamic jumps under various market

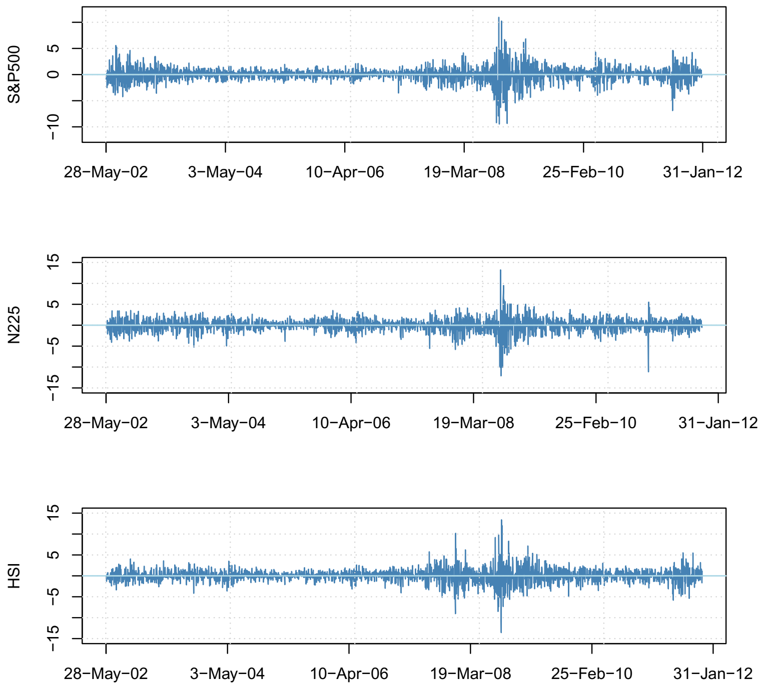


Fig. 2 The time series plots for daily stock returns

Table 3 Empirical results for the GARCH-Jump model

		Mean	Median	Std.	2.5 %	97.5 %
S&P 500	μ	0.0593	0.0593	0.0167	0.0272	0.0910
	ϕ_1	-0.0807	-0.0809	0.0209	-0.1203	-0.0373
	ω	0.0018	0.0014	0.0015	0.0001	0.0057
	α	0.0857	0.0851	0.0104	0.0661	0.1079
	β	0.9020	0.9026	0.0106	0.8807	0.9212
	ν_0	0.1860	0.1858	0.0647	0.0650	0.2963
	ν_1	0.3910	0.3768	0.1629	0.1617	0.6837
	τ	0.0057	0.0052	0.0040	0.0003	0.0151
	θ	-0.2435	-0.2440	0.0701	-0.3814	-0.1091
	δ^2	0.3595	0.3494	0.1071	0.1808	0.6041
Nikkei 225	μ	0.0482	0.0479	0.0245	0.0014	0.0996
	ϕ_1	-0.0180	-0.0176	0.0221	-0.0628	0.0245
	ω	0.0419	0.0407	0.0113	0.0226	0.0673
	α	0.1195	0.1187	0.0139	0.0941	0.1485
	β	0.8655	0.8662	0.0149	0.8350	0.8929
	ν_0	0.0793	0.0728	0.0492	0.0034	0.1857
	ν_1	0.2993	0.2869	0.1783	0.0209	0.6741
	τ	0.0062	0.0054	0.0045	0.0003	0.0172
	θ	-0.0223	-0.0203	0.1010	-0.2253	0.1664
	δ^2	0.0216	0.0167	0.0186	0.0068	0.0671
HSI	μ	0.0579	0.0586	0.0227	0.0133	0.1023
	ϕ_1	0.0195	0.0190	0.0206	-0.0208	0.0599
	ω	0.0158	0.0154	0.0048	0.0076	0.0264
	α	0.0716	0.0711	0.0089	0.0558	0.0906
	β	0.9216	0.9221	0.0094	0.9015	0.9389
	ν_0	0.1153	0.0984	0.0843	0.0048	0.3053
	ν_1	0.2715	0.2267	0.2018	0.0052	0.7324
	τ	0.0054	0.0045	0.0043	0.0002	0.0162
	θ	-0.0096	-0.0069	0.1029	-0.2217	0.1823
	δ^2	0.0207	0.0168	0.0158	0.0069	0.0605

conditions. The parameter estimates for the model in each market are summarized in Table 3. Posterior summaries - mean, median, standard error, and the 95% Bayesian intervals for the model parameters - are given. The three parameters ν_0, ν_1, and τ are important in characterizing the different structures of the jump dynamics among the various financial assets, and the parameters α and ν are satisfied in equations (4) and (5), respectively, and by $\delta^2 > 0$. The jump-size mean θ is negative for all three markets. The estimated mean of θ is only significantly different from zero for the S&P 500. Note that the impact of jumps on the conditional mean of returns tends to be centered around zero on average, which does not imply that jumps do not affect the distribution of returns. The estimated time-varying volatilities of GARCH-Jump models are given in Figure 3. The plots reveal the presence of exceedingly large

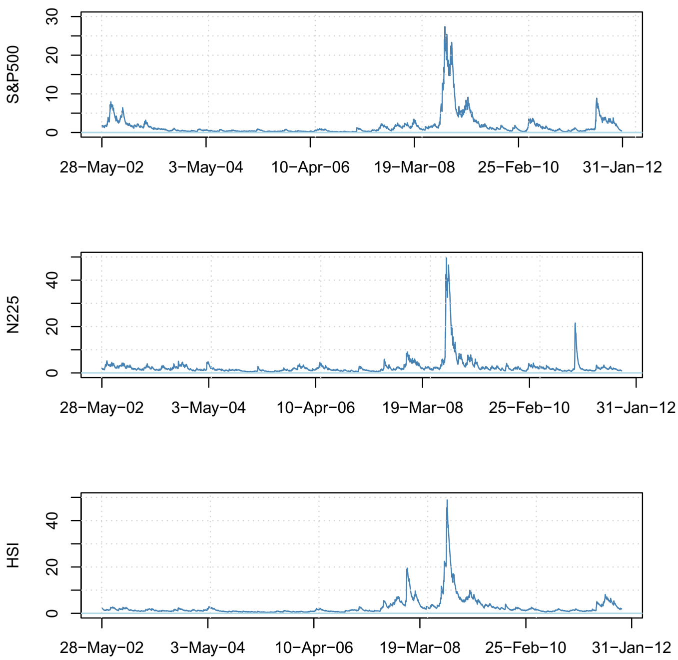


Fig. 3 The time-series plots of estimated conditional volatilities from GARCH-Jump models

movements in the volatilities variation. The unconditional jump intensity is 0.1874, 0.1040, and 0.1325 for the S&P500, Nikkei 225, and HSI, respectively. It seems that there are more frequent jumps to returns for the US market during this period. This implies that for Nikkei 225 and HSI, jumps to returns arrive on the average less than once per two weeks (around 8-10 days). But for S&P500 jumps to returns could arrive as often as once every five business days.

We attempt to compare the proposed model with the GARCH(1,1) model in Eq. (1) by utilizing NR's and Chib's methods. We report the results of the model selection in Table 4 for the GARCH and GARCH-Jump models. The likelihood values of NR's and Chib's methods are obtained from the average of the five replications, and standard deviations are given in parentheses. It is clear that the mixed GARCH-Jump model outperforms the GARCH model based on both methods. The estimated conditional volatilities of the GARCH models look similar to those of the GARCH-Jump model in Figure 3. Actually, there is a large discrepancy between the two the estimated conditional volatilities of th two models, as shown in Figure 4.

Table 4 Model selection for GARCH and GARCH model with jumps

	Method	NR		Chib	
	Model	ML	Std.	ML	Std.
S&P500	GARCH-n	-3566.2536	(0.0334)	-3579.6220	(0.2525)
	GARCH-Jump	-3533.9728	(0.8184)	-3556.1998	(0.6767)
Nikkei 225	GARCH-n	-4075.0416	(0.0149)	-4085.8530	(0.4793)
	GARCH-Jump	-4049.0128	(0.9106)	-4070.2372	(1.2633)
HSI	GARCH-n	-4014.9520	(0.0597)	-4027.9106	(0.3782)
	GARCH-Jump	-4002.8998	(0.5430)	-4021.7002	(1.3826)

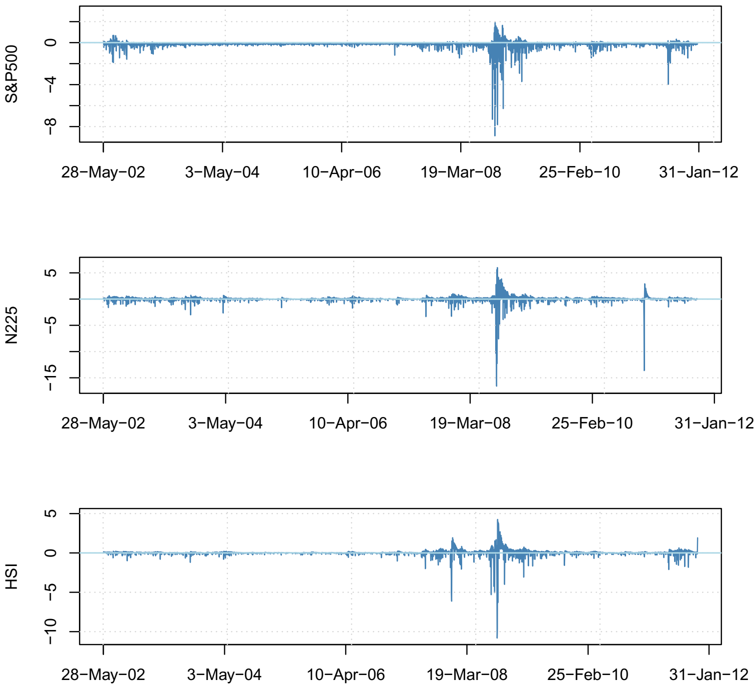


Fig. 4 The estimated conditional volatility differences between GARCH-Jump and GARCH models

7 Conclusion and Future Research

In this paper, we propose a Bayesian inference for a mixed GARCH-Jump model to capture the dynamic volatility including the time-varying volatility effects in the jump dynamics component. A simulation study shows that the Bayesian MCMC gives sound parameter estimates. For model comparison, we calculate the marginal log likelihood using NR's and Chib's methods for the mixed GARCH-Jump and

GARCH model. The results show that the mixed GARCH-Jump model outperforms a standard GARCH model and show the existence of jumps in the empirical evidence from all of the three financial markets.

In the future, the mixed GARCH-Jump model can be extended in various directions to obtain more flexibility and applicability. If we have extra information such as opening prices, minimum/maximum prices, or traded volumes, we can also input information into the jump intensity, improving the behavior of the mixed GARCH-Jump model. Moreover, we can consider Bayesian forecasting for GARCH-Jump model to carry out Value-at-Risk and Expected Shortfall forecasts, both of which are popular risk measures in financial risk management.

Acknowledgement. We thank the editor and referee for their insightful and helpful comments, which improved this paper. Cathy W.S. Chen is supported by the grant (NSC 101-2118-M-035-006-MY2) from the National Science Council (NSC) of Taiwan. The post-doctoral fellowship of Edward Lin is supported by the grant (NSC 99-2118-M-035-001-MY2) from NSC of Taiwan.

References

1. Bauwens, L., Lubrano, M.: Bayesian inference on GARCH models using Gibbs sampler. Econometrics Journal 1, 23–46 (1998)
2. Bollerslev, T.: Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327 (1986)
3. Bollerslev, T., Chou, R.Y., Kroner, K.P.: ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5–59 (1992)
4. Chan, W.H., Maheu, J.M.: Conditional Jump Dynamics in Stock Market Return. Journal of Business & Economic Statistics 20, 377–389 (2002)
5. Chen, C.W.S., Gerlach, R., So, M.K.P.: Comparison of non-nested asymmetric heteroscedastic models. Computational Statistics & Data Analysis. Special Issue on Nonlinear Modelling & Financial Econometrics 51, 2164–2178 (2006)
6. Chen, C.W.S., So, M.K.P.: On a threshold heteroscedastic model. International Journal of Forecasting 22, 73–89 (2006)
7. Chen, C.W.S., So, M.K.P., Lin, E.M.H.: Volatility forecasting with double Markov switching GARCH models. Journal of Forecasting 28, 681–697 (2009)
8. Chib, S.: Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321 (1995)
9. Daal, E., Naka, A., Yu, J.S.: Volatility clustering, leverage effects, and jump dynamics in the US and emerging Asian equity markets. Journal of Banking & Finance 31, 2751–2769 (2007)
10. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica 5, 1–50 (1982)
11. Hastings, W.K.: Monte-Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109 (1970)
12. Maheu, J.M., McCurdy, T.H.: News arrival, jump dynamics, and volatility components for individual stock returns. Journal of Finance 64, 755–793 (2004)
13. Maheu, J.M., McCurdy, T.: Modeling foreign exchange rates with jumps. In: Forecasting in the Presence of Structural Breaks and Model Uncertainty. Eds David Rapach and Mark Wohar, Emerald (2008)

14. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, E.: Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1091 (1953)
15. Newton, M.A., Raftery, A.E.: Approximate Bayesian inference with the weighted likelihood Bootstrap (with discussion). Journal of the Royal Statistical Society, Ser. B. 56, 3–48 (1994)
16. Vrontos, I.D., Dellaportas, P., Politis, D.N.: Full Bayesian inference for GARCH and EGARCH models. Journal of Business & Economic Statistics 18, 187–198 (2000)

Risk Measures and Asset Pricing Models with New Versions of Wang Transform

Baokun Li, Tonghui Wang, and Weizhong Tian

Abstract. To provide incentive for active risk managements, tail-preserving and coherent distortion risk measures are needed in the actuarial and financial fields. In this paper we propose new versions of Wang transform using two different forms of skew-normal distribution functions, and prove that the related risk measures in Choquet integral form are coherent and degree-two tail-preserving for usual bi-atomic risk distributions. Also under some plausible conditions, the portfolio optimization is explored for the capital asset pricing model where the pricing strategy uses the new Wang transforms as the distortion functions.

1 Introduction

Risk measures are used to decide insurance premiums and required capital for a given risk portfolio by examining its downside risk potential. A widely used risk measure for the risk of loss on a specific portfolio of financial assets is the value at risk (VaR), a threshold value in which the probability that the mark-to-market loss on the portfolio exceeds this value is the given probability level. Mathematically, the VaR is simply a percentile on the distribution of losses.

Although the VaR is widely accepted and used in financial control, risk management, and governance of trusts and pension plans, several authors have pointed out

Baokun Li
School of Statistics,
Southwestern University of Finance and Economics, China
e-mail: bali@swufe.edu.cn

Tonghui Wang
College of Science, Northwest A and F University, China
and New Mexico State University, USA
e-mail: twang@nmsu.edu

Weizhong Tian
New Mexico State University, USA

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 155–167.
DOI: 10.1007/978-3-642-35443-4_11

the two deficiencies of the VaR, lack of subadditivity and difficulty to optimize as it may have multiple local minima (Artzner et al.[2]; Basak and Shapiro [4]). Also the VaR is often criticized for not taking into account the magnitude of losses when the VaR is exceeded. Thus the conditional VaR (CVaR), known as expected shortfall, is proposed as an alternative to the VaR. The CVaR is defined as the conditional expectation of the losses exceeding VaR, which provides additional information of the losses in the tail exceeding the VaR.

Distortion risk measures based on Wang transform overcome the drawbacks of both the VaR and the CVaR, but they do not always provide incentive for risk management since they do not give capital relief in some simple risk distributions. Because they, like the CVaR, preserve only degree 1 and 0 tail-preserving order (Hürlimann [7]). In this paper we propose new versions of Wang transform using two forms of skew-normal distribution functions.

This paper is organized as follows. Two distortion risk measures and their properties are discussed in Section 2 and two forms of skew normal distributions are introduced in Section 3. New versions of Wang transform under skew normal settings, together with their distortion functions, are studied in Section 4. Results on properties of risk measures with new distortion functions are obtained in Section 5. In Section 6, we discuss the skew normal distortion functions for the capital asset pricing model(CAPM) and an example is given for illustrating our results.

2 Two Types of Distortion Risk Measures

2.1 Coherent Distortion Risk Measures

Let $(\Omega, \mathscr{A}, P)$ be a probability space such that Ω is the space of outcomes or states of the world, $\mathscr{A}$ is the σ-algebra of events, and P is the probability measure. For a measurable real-valued random variable X on this probability space, that is, a map $X : \Omega \to R$, the probability distribution of X is defined and denoted by $F_X(x) = P(X \leq x)$.

In the present paper, the random variable X represents a financial loss such that, for $\omega \in \Omega$, the real number $X(\omega)$ is the realization of a loss or profit function, with $X(\omega) \geq 0$ for a loss and $X(\omega) < 0$ for a profit. Let $\mathscr{X}$ be the set of financial losses. A *risk measure*, $\rho(X)$, is a functional from the set of losses to the extended nonnegative real numbers described by a map $\rho : \mathscr{X} \to R^+$. Note that $\rho(X)$ can be considered as an amount that a company must reserve to face financial loss X, that is $\rho(X)$ is a minimum amount that the company insurance must reserve to pay the damage made by risk X so that $\rho(0) = 0$.

Definition 3. *A risk measure is said to be* **coherent** *if it satisfies the following desirable properties (see, e.g., Arztner et al.[1] and [2]):*

(i) Monotonicity: $\rho(X) \leq \rho(Y)$ *provided that* $P(X \leq Y)$.
(ii) Homogeneity: for any $c > 0$ *and* $X \in \mathscr{X}$, $\rho(cX) = c\rho(X)$.

(iii) Subadditivity: for any $X,Y \in \mathscr{X}$, $\rho(X+Y) \leq \rho(X) + \rho(Y)$.
(iv) Translation invariance: $\rho(X+c) = \rho(X) + c$ *for any* $X \in \mathscr{X}$ *and* $c \in \Re$.

A *distortion function* is a continuous non-decreasing function $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$. Let $F_X(x)$ be the distribution function of the risk X, the transform $F_X^g(x) = g(F_X(x))$ defines a distribution function, which is called the *distorted distribution function*. In this paper, we will focus on the distortion risk measure defined by Choquet integral, given by

$$\rho_g(X) = -\int_{-\infty}^{0} [1 - g(S_X(x))]dx + \int_{0}^{\infty} g(S_X(x))dx, \tag{1}$$

where $S_X(x) = 1 - F_X(x)$ is the *survival function* of X. Wang [9] proved that the risk measure given in (1) is coherent if the distortion function $g(x)$ is a concave function. We will show in Section 5 that the risk measure defined in (1) is coherent for the skew normal distortion functions.

2.2 Tail-Preserving Distortion Risk Measures

For any random variable $X \in \mathscr{X}$ with probability distribution function $F_X(x)$, the higher order partial moments $\pi_X^n(x) = E[(X-x)_+^n]$, $n = 0,1,2,...$, are said to be the *degree n stop-loss transforms*. Note that $\pi_X^0(x) = S_X(x) = 1 - F(x)$ is the survival function of X. For two random variables X and Y, if $\pi_X^n(x) \leq \pi_Y^n(x)$ for all x, the X is said to precede Y in *degree n stop-loss transform order*, denoted by $X \leq_{slt}^{(n)} Y$. Also, we say that the X precedes Y in *(n+1)-convex order*, denoted by $X \leq_{(n+1)-cx} Y$ if $X \leq_{slt}^{(n)} Y$ and $E[X^k] = E[Y^k]$, for all $k = 0,1,\ldots,n$.

A desirable property for a risk measure is that increased risk should be penalized with an increased measure. A distortion measure $\rho_g(X)$ with concave distortion function preserves the stop-loss order(see, e.g., Hürlimann [7]). With equal means and variances, a stop-loss order relation between different random variables cannot exist. In this situation, one is interested in distortion measures that preserve higher degree convex orders. Such measures are called *tail-preserving distortion measures*.

Definition 4. *A risk measure* $\rho : \mathscr{X} \to R$ *is said to be a* ***degree*** n ***tail-preserving risk measure*** *if it is preserved under the* $(n+1)$*-convex order. That is, if two random variables* X *and* $Y \in \mathscr{X}$ *satisfy* $X \leq_{(n+1)-cx} Y$*, then* $\rho(X) \leq \rho(Y)$.

It is known that a distortion measure $\rho_g(X)$ with concave distortion function $g(x)$ preserves $(n+1) - cx$ order for $n = 0,1$ so that it is a tail-preserving risk measure of degree zero and one. From Examples 3.1 and 3.2 of Hürlimann [7], both the CVaR and Wang distortion risk measure are not the degree-two tail-preserving coherent risk measures.

3 Two Forms of Skew-Normal Distributions

The popular skew-normal distribution was introduced by Azzalini [3] and has been studied by many authors in last two decades.

Definition 5. *The random variable X is said to have the skew normal distribution with location parameter* $\mu \in \Re$*, scale parameter* $\sigma > 0$*, and skewness parameter* $\alpha \in R$*, denoted by* $X \sim SN(\mu, \sigma^2, \alpha)$*, if its probability density function is of the form*

$$f_X(x|\mu,\sigma,\alpha) = 2\phi(x|\mu,\sigma)\Phi\left[\alpha\left(\frac{x-\mu}{\sigma}\right)\right],$$

where $\phi(.)$ *and* $\Phi(.)$ *are probability density function and cumulative distribution function of standard normal distribution, respectively.*

For simplicity, we call $SN(\mu, \sigma^2, \alpha)$ the α-skew normal distribution. Note that the distribution is skewed to the right if $\alpha > 0$ and to the left if $\alpha < 0$. When $\alpha = 0$, the distribution is reduced to $N(\mu, \sigma^2)$. the normal distribution with mean μ and variance σ^2.

The Inverse scale factor skew distribution was introduced by Fernandez and Steel [8]. Suppose that g is the probability density function symmetric about 0 and $\gamma \in (0, \infty)$ is a scalar. The inverse scale factor skew distribution of a random variable Y is defined by means of the probability density function

$$s(y|\gamma) = \frac{2}{\gamma + 1/\gamma} g(y\gamma^{-sgn(y)}),$$

where $sgn(t)$ is the sign function which is equal to 1 if $t > 0$, -1 if $t < 0$, and 0 if $t = 0$. Using this inverse scaled factor mechanism, we have the following definition.

Definition 6. *The random variable Y is said to have the* **inverse scale factor skew-normal distribution** *with location parameter* μ *and scale parameter* σ^2 *and skewness parameter* γ*, denoted by* $Y \sim ISN(\mu, \sigma^2, \gamma)$*, if its probability density function is given by*

$$f_Y(y|\mu,\sigma,\gamma) = \frac{2}{\gamma + 1/\gamma} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\gamma^{-sgn(y-\mu)}\right)^2\right].$$

For simplicity, we call $Y \sim \mathrm{ISN}(\mu, \sigma^2, \gamma)$, the γ-skew normal distribution. The idea of the skewing mechanism is to scale Y differently for negative and positive values. The parameter γ clearly determines the direction and the intensity of the skewness.

For both the α-skew normal and the γ-skew normal distributions, by straightforward calculation, we obtain the following results.

Lemma 1. *(i) The mean and variance of* $X \sim SN(\mu, \sigma^2, \alpha)$ *are*

$$E(X) = \mu + \sigma\delta\sqrt{\frac{2}{\pi}}, \quad Var(X) = \sigma^2\left(1 - \frac{2\delta^2}{\pi}\right) \quad with \quad \delta = \alpha/\sqrt{1+\alpha^2}.$$

(ii) The mean and variance of $Y \sim ISN(\mu,\sigma,\gamma)$ are

$$E(Y) = \mu + \sigma\Delta, \quad Var(Y) = \sigma^2\left(\frac{\gamma^4-\gamma^2+1}{\gamma^2} - \Delta^2\right) \quad with \quad \Delta = \sqrt{\frac{2}{\pi}}\frac{\gamma^2-1}{\gamma}.$$

Since the effects of skewness parameter γ on mean and variance are not limited, the γ-skew normal distribution brings more flexibility to construction of distortion function in risk measures than the α-skew normal distribution.

4 New Versions of Wang Transform

Motivated by directly extending the Sharpe ratio, used to characterize how well the return of an asset compensates the investor for the risk taken, to risks with skewed distributions, Wang [9] proposed the following Wang transform:

$$g_\lambda(x) = \Phi(\Phi^{-1}(x) + \lambda), \tag{2}$$

where $\Phi(z)$ is the cumulative distribution function (cdf) of the standard normal random variable $Z \sim N(0,1)$. The Wang transform given in (2) is a pricing formula that recovers CAPM and Black-Scholes formula under normal asset-return distributions.

Wang transforms are made of distribution functions with symmetric densities. In many real world applications, the data sets collected are not symmetrically distributed so that normal may not be good fit. Specifically, data sets are skewed to the left or skewed to the right in most applications so that the skew normal distributions maybe the better choices. In this paper, we will construct two new versions of Wang transform under skew normal settings. Denote

$$S\Phi(x|\mu,\gamma) = \int_{-\infty}^{x} f(t|\mu,1,\gamma)dt \tag{3}$$

as the cdf of a skew-normal random variable X, we have

Definition 7. *For $x > 0$, the α-skew normal distortion function and the γ-skew normal distortion function are defined by*

$$g_\alpha(x) = S\Phi_\alpha(\Phi^{-1}(x)|\mu,\alpha) \qquad for \quad \mu < 0, \quad \alpha \le 0 \tag{4}$$

and

$$g_\gamma(y) = S\Phi_\gamma(\Phi^{-1}(y)|\mu,\gamma) \qquad for \quad \mu < 0, \quad 0 < \gamma \le 1, \tag{5}$$

respectively, where $S\Phi_\alpha(\cdot)$ is the cdf of the α-skew normal distribution $SN(\mu,1,\alpha)$ and $S\Phi_\gamma(\cdot)$ is the cdf of γ-skew normal distribution $ISN(\mu,1,\gamma)$.

When $\alpha = 0$ and $\gamma = 1$, both distortion functions are to reduced to the Wang transform with $\lambda = \mu$. The graphs of the α-skew normal and the γ-skew normal distortion functions, compared with the CVaR and Wang transform are listed in Figure 1 and Figure 2, respectively.

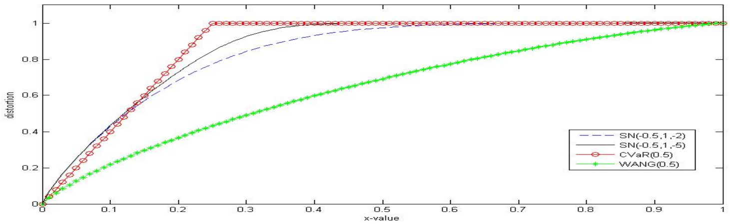


Fig. 1 The α-skew normal distortion functions with different transforms

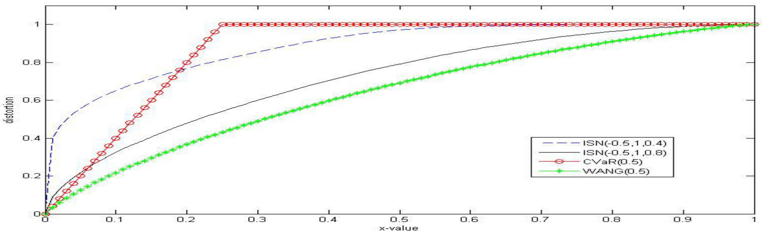


Fig. 2 The γ-skew normal distortion functions with different transforms

5 Risk Measures with New Versions of Wang Distortion Functions

A distortion risk measure can be defined as the distorted expectation of any non-negative loss random variable X. It is accomplished by using a utility or the distortion function g as follows:

$$\rho_g(X) = \int_0^{\infty} g(S_X(x))dx = \int_0^{\infty} g(1 - F_X(x))dx, \quad (6)$$

where $S_X(x)$ denotes the survival function of X, while $g(S_X(x))$ is referred to as a distorted survival function.

For the gain/loss-distributions, we know that the loss random variable X can take any real value so that the distortion risk measure is defined as in (1). The VaR is not a distortion risk measure because its distortion function is discontinuous in this case. Except the fact that properties of the distortion risk measures come from standard results about the Choquet integral (see Denneberg [5]), Wirch and Hardy [10] proved that distortion risk measures are sub-additive, i.e.,

$$\rho_g(X+Y) \le \rho_g(X) + \rho_g(Y),$$

if and only if the distortion function $g(x)$ is concave. Thus the concave distortion risk measures are coherent risk measures. In order to show that each of our extended versions of distortion risk measures is coherent, one only need to prove that each of our extended distortion functions in (4) and (5) are concave.

Theorem 1. *(i) The distortion risk measure corresponding to the α-skew normal distortion function $g_\alpha(x)$ given in (4) is coherent and*
(ii) The distortion risk measure corresponding to the γ-skew normal distortion function $g_\gamma(x)$ in (5) is coherent.

Proof. To show that g_α is coherent, we need to show that g_α has negative second derivative everywhere for $x \in \mathscr{X}$. Thus for (i), we obtain that the first order derivative

$$g'_\alpha(x) = \frac{f_X(\Phi^{-1}(x)|\mu, 1, \alpha)}{\phi(\Phi^{-1}(x))} = 2\Phi\left[\alpha\left(\Phi^{-1}(x) - \mu\right)\right],$$

which is exists and positive for all $x \in \mathscr{X}$. The second derivative of g_α also exists and is obtained as

$$g''_\alpha(x) = \frac{2\alpha}{\phi[\Phi^{-1}(x)]}\phi\left[\alpha\left(\Phi^{-1}(x) - \mu\right)\right].$$

Given $\alpha \leq 0$ and $\mu < 0$, it is easy to check that $g''_\alpha(x)$ is always negative. So the distortion risk measure, corresponding to g_α in (4), defined by Choquet integral is coherent.

Similarly, for (ii), we have

$$g'_\gamma(y) = \frac{2}{\gamma + 1/\gamma}\exp\left\{\frac{1}{2}\left((\Phi^{-1}(y))^2 - (\Phi^{-1}(y) - \mu)^2\gamma^{-2sgn(\Phi^{-1}(y)-\mu)}\right)\right\},$$

which is continuous except at the point $x = \Phi(\mu)$. Note that when $x = \Phi(\mu)$,

$$g'_\gamma(y) = \frac{2}{\gamma + 1/\gamma}\exp\left\{\frac{1}{2}(\Phi^{-1}(y))^2\right\}$$

because the term $(\Phi^{-1}(x) - \mu)^2 = 0$. Thus $g'_\gamma(y)$ exists for all $y \in \mathscr{X}$. Also we have

$$g''_\gamma(y) = \frac{g'_\gamma(y)}{\phi(\Phi^{-1}(y))}\left[\Phi^{-1}(y) - (\Phi^{-1}(y) - \mu)\gamma^{-2sgn(\Phi^{-1}(y)-\mu)}\right],$$

which exists because of the same reason discussed above. Given $\gamma \leq 1$ and $\mu < 0$, it is easy to see that $g''_2(x)$ is always negative. So the distortion risk measure, corresponding to g_γ in (5), defined by Choquet integral is coherent. □

For a loss random variable $X \sim N(\mu_0, \sigma^2)$,

$$S_X(x) = 1 - F_X(x) = 1 - \Phi\left(\frac{x - \mu_0}{\sigma}\right) = \Phi\left(-\frac{x - \mu_0}{\sigma}\right).$$

We can prove the following result:

Theorem 2. *Assume that $X \sim N(\mu_0, \sigma^2)$, the normal loss distribution. Let $F_X(x)$ be the cdf of X and $S_X(x) = 1 - F_X(x)$.*

(i) For the α-skew normal distortion function given in (4),

$$g_\alpha(S_X(x)) = 1 - S\Phi_\alpha(x|(\mu_0 - \sigma\mu), \sigma, -\alpha).$$

(ii) For the γ-skew normal distortion function given in (5),

$$g_\alpha(S_X(x)) = 1 - S\Phi_\gamma\left(x|(\mu_0 - \sigma\mu), \sigma, \frac{1}{\gamma}\right).$$

Proof. Note that

$$g_\alpha(S_X(x)) = S\Phi_\alpha\left(-\frac{x-\mu_0}{\sigma}|\mu, \sigma, \alpha\right) \quad \text{and} \quad g_\gamma(S_X(x)) = S\Phi_\gamma\left(-\frac{x-\mu_0}{\sigma}|\mu, \sigma, \gamma\right).$$

The desired results follow from the facts given by

$$S\Phi_\alpha(y|\mu, \sigma, \alpha) = 1 - S\Phi_\alpha(-y| - \mu, \sigma, -\alpha)$$

and

$$S\Phi_\gamma(y|\mu, \sigma, \gamma) = 1 - S\Phi_\gamma\left(-y| - \mu, \sigma, \frac{1}{\gamma}\right). \qquad \square$$

From Theorem 2, we know that for a normally distributed loss random variable, its distorted loss distribution is skew-normal. Denote the distorted distribution as the distribution of the loss random variable X_*, then the risk measures in Equation (1) for both g_α and g_γ are

$$\rho_{g_\alpha}(X) = E(X_*) = \mu_0 - \mu\sigma - \sigma\left(\frac{\alpha}{\sqrt{1+\alpha^2}}\right)\sqrt{\frac{2}{\pi}} \tag{7}$$

and

$$\rho_{g_\gamma}(X) = E(X_*) = \mu_0 - \mu\sigma + \sigma\left(\frac{1}{\gamma} - \gamma\right)\sqrt{\frac{2}{\pi}}, \tag{8}$$

respectively. Note that when $\alpha = 0$ or $\gamma = 1$, the risk measures are reduced to that of Wang distortion function.

Remark. From the results given above, we can conclude that both g_α and g_γ affect the risk measures by the location parameter μ, scale parameter σ and skew parameter α or γ. Note that the range of $\frac{\alpha}{\sqrt{1+\alpha^2}}$ is $(-1, 1)$ and the range of $\frac{1}{\gamma} - \gamma$ is $(-\infty, \infty)$ so that the effect of the skew parameter α to the risk measure $\rho_{g_\alpha}(X)$ is limited. But the γ-skew normal distortion function brings more flexibility in distorting the loss distribution compared to Wang transform. $\square$

In order to show that the distortion risk measure is also degree-two tail-preserving, we need the following lemma.

Lemma 2. (Hürlimann [7]). *Let $g(x)$ be a continuous and differentiable increasing concave distortion function. The coherent distortion risk measure $\rho_g(X)$ is a degree-two tail-preserving risk measure for the subset of bi-atomic losses if and only if the following condition holds:*

$$\frac{t}{1-t}(1-g(t))+g(t)-2tg'(t)\geq 0, \quad \textit{for} \quad t\in(0,1). \tag{9}$$

Theorem 3. *(i) The distortion risk measure corresponding to α-skew normal distortion function given in (4) is the degree-two tail-preserving risk measure for the bi-atomic losses.*
(ii) The distortion risk measure corresponding to α-skew normal distortion function given in (5) is the degree-two tail-preserving risk measure for the bi-atomic losses.

Proof We prove the part (i) only and the part (ii) can be done similarly. Note that the first two terms in the inequality (9) are positive because of the properties of distribution function, and

$$g_2'(x)=2\Phi[\alpha(\Phi^{-1}-\mu)]\to 0 \quad \text{as} \quad \alpha\to-\infty.$$

So the inequality (9) holds and the desired result follows from Lemma 2. □

6 Skew Normal Distortion Function for CAPM

Both versions of skew normal distortion functions, $S\Phi_\alpha(\Phi^{-1}(x)|\mu,\alpha)$ given in (4) and $S\Phi_\gamma(\Phi^{-1}(x)|\mu,\gamma)$ given in (5), could also be used in capital asset pricing model (CAPM), where the parameters are assumed to be $\mu>0$ and $\alpha>0$ (or $\gamma>1$).

Specifically, we denote the return rate of the i-th stock at day t as

$$r_{i,t}=\frac{p_{i,t}}{p_{i,t-1}}-1, \qquad i,t=1,2,\cdots,$$

where $p_{i,t}$ is the closing price for the the i-th stock at day t. For simplicity, we brief $r_{i,t}$ to be r_i. Given a portfolio containing k stocks with corresponding daily return rate r_i, its total return rate for the specific day t is

$$r=\sum_{i=1}^{k}w_ir_i, \qquad \text{with} \quad w_i\geq 0, \quad \sum_{i=1}^{k}w_i=1,$$

where w_i's are the weights of the i-th stock in your portfolio.

Let the Choquet integral be the distorted value, the expected value of the portfolio under the distortion function, the optimization problem for the portfolio is set as follows. The objective function is $h(\mathbf{w})$ defined by

$$h(\mathbf{w}) = \rho_g(\mathbf{r}) = -\int_{-\infty}^{0} [1 - g(S_X(x))]dx + \int_{0}^{\infty} g(S_X(x))dx$$

with restrictions

$$\mathbf{e}_k^{\mathrm{T}}\mathbf{w} = \sum_{i=1}^{k} w_i = 1 \quad \text{and} \quad w_i \geq 0, \quad i = 1, 2, \cdots, k,$$

where the distortion function can be either the g_α defined in (4) or the g_γ defined in (5), $\mathbf{w} = (w_1, w_2, \cdots, w_k)^{\mathrm{T}}$, a column vector in $(0,1]^k$, $\mathbf{e}_k = (1, 1, \cdots, 1)^{\mathrm{T}}$, and $\mathbf{r} = (r_1, r_2, \cdots, r_k)^{\mathrm{T}}$, the column vector of k daily return rates. Note that from Theorem 1, we know that both risk measures are convex with $\mu > 0$ and $\alpha > 0$ (or $\gamma > 1$).

We will use the $\rho_{g_\gamma}(\mathbf{r})$ to illustrate the procedure for solving this optimization problem and the use of $\rho_{g_\alpha}(\mathbf{r})$ can be treated similarly.

Suppose that the return rates of k stock prices $\mathbf{R} = (R_1, R_2, \cdots, R_k)^{\mathrm{T}}$ follows a multivariate normal distribution with mean vector $\mu_0\mathbf{e}_k$ and covariance matrix Σ. Then the total return rate $R = \sum_{i=1}^{k} w_i R_i \sim N(\mu_0, \sigma^2)$, the normal distribution with mean value μ_0 and variance $\sigma^2 = \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}$. Note that a realization of $\mathbf{R}$ is $\mathbf{r}$, which is the vector of the daily return rates of stocks on a specific day. From Theorem 2, we know that its distortion distribution is the γ-skew normal with location parameter $\mu_0 - \sigma\mu$, scale parameter σ and skewness parameter $\gamma > 1$. Therefore by (8), the optimization problem is simplified as

$$h(\mathbf{w}) = \mu_0 - \mu\sigma + \sqrt{\frac{2}{\pi}}\left(\frac{1}{\gamma} - \gamma\right)\sigma. \tag{10}$$

with restrictions

$$\mathbf{w}^{\mathrm{T}}\mathbf{e}_k = \sum_{i=1}^{k} w_i = 1 \qquad \text{and} \qquad w_i \geq 0, \quad i = 1, 2, \cdots, k,$$

where $\mu > 0$ and $\gamma > 1$ are known parameters.

Let $t = \mu - \sqrt{2/\pi}(1/\gamma - \gamma)$. Then Equation (10) can be rewritten as $H(\mathbf{w}) = \mu_0 - t\sigma$. Denote $\mathbf{y} = \Sigma^{1/2}\mathbf{w}$, $\nu = \Sigma^{-1/2}E(\mathbf{R})$, and $\theta = \Sigma^{-1/2}\mathbf{e}_k$. The objective function with restrictions is reduced to

$$H(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}\nu - t\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}} \qquad \text{with} \qquad \mathbf{y}^{\mathrm{T}}\theta = 1. \tag{11}$$

It is well known that the unique solution of this mean-standard deviation optimization problem may not exist. In the following we will provide a solution for the optimization problem by adding an extra condition.

Theorem 4. *The solution of the optimization problem given in (11) is unique if* $t > \sqrt{\nu^T\nu}$.

Proof. Using Lagrange multiplier method, the objective function is

$$L(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}\nu - t\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}} + \lambda(\mathbf{y}^{\mathrm{T}}\theta - 1). \tag{12}$$

Taking partial derivatives with respect to $\mathbf{y}$ and λ, respectively, and set them to 0, we obtain

$$\frac{\partial L}{\partial \mathbf{y}} = \nu - \frac{t\mathbf{y}}{\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}}} + \lambda\theta = 0 \qquad \text{and} \qquad \frac{\partial L}{\partial \lambda} = \mathbf{y}^{\mathrm{T}}\theta - 1 = 0. \tag{13}$$

From the first equation of (13), we obtain

$$\frac{t\mathbf{y}}{\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}}} = \nu + \lambda\theta.$$

Multiplying to the left the transpose of itself in both sides of this equation, we obtain

$$(\theta^{\mathrm{T}}\theta)\lambda^2 + 2(\theta^{\mathrm{T}}\nu)\lambda + \nu^{\mathrm{T}}\nu - t^2 = 0.$$

Under the assumption that $t > \sqrt{\nu^T\nu}$, this quadratic form of λ has a unique positive root. Let $\hat{\lambda}$ be the root, the solution $\hat{\mathbf{y}}$ of $\mathbf{y}$ can be uniquely determined by

$$\frac{\mathbf{y}}{\sqrt{\mathbf{y}^{\mathrm{T}}\mathbf{y}}} = \frac{\nu + \hat{\lambda}\theta}{t} \qquad \text{and} \qquad \mathbf{y}^{\mathrm{T}}\theta = 1.$$

Now, we will show that if the solution $\hat{\mathbf{y}} \in (R^+)^k = \{(y_1, y_2, ..., y_k)^{\mathrm{T}} | y_i > 0, i = 1, 2, ..., k\}$, then the objective function given in (13) reaches its maximum. Since the domain $(R^+)^k$ is open, we only need to show that the Hession matrix of $L(\mathbf{y})$ is negative semi-definite and is given by

$$\begin{aligned}
H = \frac{\partial^2 L(\mathbf{y})}{\partial \mathbf{y}\partial \mathbf{y}^{\mathrm{T}}} &= -t \begin{bmatrix} \frac{\partial^2 L}{\partial y_1^2} & \frac{\partial^2 L}{\partial y_1 \partial y_2} & \cdots & \frac{\partial^2 L}{\partial y_1 \partial y_k} \\ \frac{\partial^2 L}{\partial y_2 \partial y_1} & \frac{\partial^2 L}{\partial y_2^2} & \cdots & \frac{\partial^2 L}{\partial y_2 \partial y_k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 L}{\partial y_k \partial y_1} & \frac{\partial^2 L}{\partial y_k \partial y_2} & \cdots & \frac{\partial^2 L}{\partial y_k^2} \end{bmatrix} \\
&= -t \left(\sum_{s=1}^{k} y_s^2 \right)^{-3/2} \begin{bmatrix} \sum_{s=1}^{k} y_s^2 - y_1^2 & -y_1 y_2 & \cdots & -y_1 y_k \\ -y_2 y_1 & \sum_{s=1}^{k} y_s^2 - y_2^2 & \cdots & -y_2 y_k \\ \vdots & \vdots & \vdots & \vdots \\ -y_k y_1 & -y_k y_2 & \cdots & \sum_{s=1}^{k} y_s^2 - y_k^2 \end{bmatrix} \\
&= -t \left(\sum_{s=1}^{k} y_s^2 \right)^{-3/2} B.
\end{aligned}$$

Let $\mathbf{u} = (u_1, u_2, \cdots, u_k)^{\mathrm{T}}$. The for all $\mathbf{u} \in R^k$, we have

$$\mathbf{u}^{\mathrm{T}} B \mathbf{u} = \sum_{i \neq j} (y_i y_j - u_i u_j)^2 \geq 0,$$

so B is non-negative definite matrix and hence H is negative semi-definite. Because this is true for any $Y \in R^{k+}$, the optimization problem solution $\hat{Y}$ is the unique maximum. □.

In the following, an empirical example is given for this allocation problem.

Example 2. *The average daily return rates for 5 stocks over sixty days in China's stock markets are given as follows:*

Code	*Company Name*	*Daily Rate*
600019	*Baoshan Steel Ltd*	*1.0022*
600086	*Oriental Gold*	*1.0012*
600036	*China Merchants Bank*	*1.0025*
600068	*Gezhouba Dam*	*1.0013*
000822	*Shandong Ocean Chemical Group*	*1.0056*

Using the price time series data of the these 5 stocks, their variance-covariance matrix is obtained using Maple as:

$$\Sigma = \begin{bmatrix} .000296 & .000177 & .000139 & .000096 & .000112 \\ .000177 & .000992 & .000076 & .000266 & .000265 \\ .000139 & .000076 & .000238 & .000055 & .000149 \\ .000096 & .000266 & .000055 & .001028 & .000195 \\ .000112 & .000265 & .000149 & .000195 & .000817 \end{bmatrix}$$

For $t = 1$*, the solution of the optimization problem is obtained as:*

$$\mathbf{w} = (0.283, 0.021, 0.505, 0.075, 0.116)^T$$

which is the weights for these 5 stocks to be considered when you want to allocate them. □

Acknowledgements. We wish to thank referees for their valuable suggestions which led to the improvement of this paper. Also we wish to thank Dr. Rendao Ye of Hangzhou Dianzi University and New Mexico State University for his comments during the preparation of this paper.

References

1. Artzner, P., Delbaen, F., Eber, J., Heath, D.: Thinking Coherently. RISK 10(11), 68–71 (1997)
2. Artzner, P., Delbaen, F., Eber, J., Heath, D.: Coherent Measures of Risk. Mathematical Finance 9(3), 203–228 (1999)
3. Azzalini, A.: A class of distributions which includes the normal ones. Sc. and. J. Statistic. 12, 171–178 (1985)
4. Basal, S., Shapiro, A.: Value-at-risk-based risk management: optimal policies and asset prices. Rev. Finance. Stud. 14(2), 371–405 (2001)
5. Denneberg, D.: Representation of the Choquet integral with the sub-additive Möbius transform. Fuzzy Sets and Systems, 139–156 (1994)

6. Hürlimann, W.: On Stop-Loss Order and the Distortion Pricing Principle. ASTIN Bulletin 28(1), 119–134 (1998)
7. Hürlimann, W.: Distortion risk measures and economic capital. North American Actuarial Journal 8(1), 86–95 (2004)
8. Fernandez, C., Steel, M.F.J.: On Bayesian Modelling of Fat Tails and Skewness. Journal of the American Statistical Association 93, 359–371 (1998)
9. Wang, S.S.: A Class of Distortion Operators for Pricing Financial and Insurance Risks. Journal of Risk and Insurance 67(1), 15–36 (2000)
10. Wirch, J., Hardy, M.R.: A synthesis of risk measures for capital adequacy. Insurance: Mathematics and Economics 25, 337–347 (1999)

Part III
Applications

Purchasing Power Parity Puzzle and the Australian Dollar Real Exchange Rate

Khorshed Chowdhury

Abstract. This paper examines mean reversion in the real exchange rate (RER) index of Australia in the presence of structural breaks from 1984 quarter 1 till 2011 quarter 1. Testing for mean reversion in RER is one way of testing the purchasing power parity (PPP) theory of international trade and finance. Mean reversion is examined by using a minimum Lagrange Multiplier unit-root test that allows for breaks in level and trend. We were able to reject the unit-root null hypothesis and find evidence of mean reversion and hence purchasing power parity (PPP). Our finding reverses the results of past studies that failed to prove convergence to PPP in the long-run. The corresponding structural break dates are 1988 quarter 2 and 2002 quarter 4 respectively and these breaks are statistically significant. The break dates mostly correspond to the period of RER instability (1986-1989) and the recovery of the Australian dollar driven by the resources boom (2001-2002).

Keywords: Real exchange rate, purchasing power parity, unit-root, structural breaks.

JEL Classification: F13, F31, F41.

1 Introduction

Real exchange rate (RER)–the ratio of price of tradables to price of nontradables –measures the cost of foreign goods relative to domestic goods. [14] defines it as "...the product of the nominal exchange rate, expressed as the number of foreign currency units per home currency unit, and the relative price level, expressed as the ratio of the price level in the home country to the price level in the foreign country." RER measures the external competitiveness of an economy and is useful

Khorshed Chowdhury
School of Economics, University of Wollongong,
NSW2522, Australia
e-mail: khorshed@uow.edu.au

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 171–184.
DOI: 10.1007/978-3-642-35443-4_12

in explaining trade behaviour and national income. The policy issue of 'overvaluation/undervaluation' and the resultant existence and magnitude of distortions is discussed in terms of RER movements. Since RER is a price that ensures internal and external equilibrium simultaneously, it plays a pivotal role in macroeconomic adjustment. RER misalignment has adverse welfare and efficiency costs on small, open economies like Australia.

Testing for mean reversion in RER is one way of testing the purchasing power parity (PPP) theory. The basis for PPP is the "law of one price" derived from international trade theory. Short-run deviations from PPP are significant, while the deviations from PPP dissipate in the long-run. The absence of unit-root in RER will indicate that long-run PPP holds. To highlight this point, let us consider the logarithms of the Australian dollar price of a unit of foreign currency (s_t), the logarithms of the Australian price level (p_t), the logarithms of foreign price level (p_t^*) and the logarithms of RER (q_t). Thus, q_t can be expressed as follows:

$$q_t = s_t + p_t^* - p_t \quad (1)$$

The absolute version[1] of PPP theory implies that nominal exchange rate (st) is proportional to the relative price ratio (p_t/p_t^*) thus rendering qt to remain constant over time. If qt changes over time and follows a stationary autoregressive moving average (ARMA) process, then deviations from PPP are transient. Short-run deviations from PPP are perfectly consistent with efficiently functioning financial markets. However, if qt is non-stationary, then the deviations will not be eliminated resulting in the failure of PPP in the long-run.

Empirical examinations in the 1960s lend some support of PPP over long periods of time. Since then empirical evidence on the validity of PPP has been mixed so that the validity of PPP remain doubtful. It was generally assumed that the exchange rate would move quickly in line with changes in relative price levels after the collapse of the Bretton Woods system. [11] 'overshooting' hypothesis provided some theoretical justification for the transient deviations from PPP. Empirical tests of the mid-1980s tended to reject PPP except in countries with high inflation [15]. This view was criticised because the time series properties of exchange rates and relative prices were ignored. Since 1973 increasing evidence of mean reversion of RERs in industrialised countries has been found in studies employing the panel unit-root test ([29]; [34]; [36]; [35] *inter alia*). Critics are sceptical of the evidence given the low power and size distortions of these tests. Some studies [16]; [30]; [41]; [45]; [2]; [26]) show that the behaviour of the exchange rate can be non-linear where the exchange rate adjustment can be characterised as a smooth transition autoregressive (STAR) process[2].

[1] Mean reversion is a tendency for a stochastic process to remain near, or tend to return over time to a long-run average value. Mean reversion also implies stationarity of a stochastic process

[2] We do not pursue this strand of research as it is beyond the scope of this paper.

Given the conundrum of results, the objective of this paper is to test for mean reversion of RER of Australia in the presence of structural breaks[3] since December 1983[4]. The traditional unit-root tests (like Dickey-Fuller (DF), Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP)) and tests accounting for a single structural break have low power when multiple structural breaks are ignored[5]. To the best of my knowledge, it is the first study that employs Australian RER data and tests for unit-root in the presence of structural breaks. Allowing for structural breaks is particularly important considering the nature of the post-float experience for Australia.

The structure of the paper is as follows: In Section II we provide a critique of the previous studies on testing for unit-roots of RER of Australia. In Section III, we conduct a bevy of unit-root tests that ignores structural breaks in the data generation process (DGP). Next we conduct the powerful [24], henceforth LS, minimum Lagrange Multiplier (LM) unit-root test with structural breaks. The LS test with two structural breaks endogenously determines the location of two breaks in level and trend and tests the null of a unit-root. The LS test with two structural breaks is invariant to the magnitude of the breaks. The alternative hypothesis of the LS test unambiguously implies trend stationarity. The results are discussed in Section IV. Section V concludes with a summary of the findings.

2 Past Studies of Unit-Root of RER of Australia

Past studies on testing for unit-root of RER of Australia are sparse. A majority of these studies have used the traditional tests (DF, ADF, KPSS and others) which suffer from power deficiency when structural breaks are ignored. A few studies ([6]; [10]; [20] have incorporated a single endogenous structural break while testing for unit-root with opposing results. So far empirical results ([7]; [10]; [20]) are overwhelming in favour of rejection of the mean reversion hypothesis[6].

In earlier studies, the Australian RER was characterised as a unit-root process ([4]; [3] and [17]). [18] "estimate the real exchange rate models over the post-float period; a sample so short that tests of non-stationarity generates ambiguous results. Tests on a longer sample of Australia's trade-weighted RER suggest it is stationary, possibly around a trend [19]". [44], using RBA quarterly data from 1973 quarter 4 to 1995 quarter 2, found the trade-weighted RER to be stationary around a trend by using the ADF test and [22] (KPSS) test. A notable feature of [44] is that RER was

[3] The examples of policies with break consequences include frequent devaluations, deregulation of both real and financial sectors and policy regime shifts, abrupt exogenous changes like the H1N1, SARS pandemic etc. This can lead to huge forecasting errors and unreliability of the model in general.

[4] After the collapse of the Bretton Woods system in February 1973, the Reserve Bank of Australia (RBA) pegged the Australian dollar with a basket of currencies of its trading partners. The Australian dollar was completely floated from December 1983, allowing its value to fluctuate dependent on supply and demand on international money markets.

[5] A succinct review of the unit-root tests are given in the Appendix.

[6] [33] found shocks to RER have finite life and interpret their results as evidence in favour of PPP.

found to be stationary on the basis of ADF and KPSS unit-root test for the entire sample period while for the post-float period RER was non-stationary which was contradicted by the KPSS test.

[5] used RBA quarterly data from 1981 quarter 3 to 2000 quarter 4 to quantify the extent to which the Australian trade-weighted RER was misaligned relative to its long-run equilibrium value. [5] wrote "The time series properties of the data were examined. The Dickey-Fuller test was unable to reject the null hypothesis of stationarity for all of the variables[7]." Results reported in Table 1 page 19 are erroneous.

By employing the ADF test and data from 1973 quarter 1 to 1995 quarter 3, [1] finds the RER of Australia to be non-stationary. [1] defined the bilateral RER(q) $= eCPI^{US}/CPI^{AUS}$, where, $e =$ nominal exchange rate and CPI^{US}, CPI^{AUS} represent the consumer price indices of the US and Australia respectively. This definition of RER is restrictive and does not capture the overarching influence of relative prices and bilateral exchange rates of the trading partners[8].

These unit-root tests were carried out while modelling the fundamental determinants of the Australian RER. It seems that the result is sensitive to the test method and the size of the sample. Further, these studies ignored structural breaks and the profound influence it can have on the DGP. Some researchers ([20] and [6]) enter this debate by including the influence of structural change.

[20] used [47] and [38] unit-root tests failed to find evidence of mean reversion in RER of Australia over the period 1973 quarter 1 till 1999 quarter 1. It is worth noting that trade-weighted RER has been calculated from [23] index of RER without reference to various trade-weights being used and the number of trading partners. Thus, the RER measure on page 653 of Henry and Olekalns (2002) may not be an accurate and comprehensive measure of RER.

The data accuracy problem was addressed by [6] who used the RER indices of RBA. [6] comprehensively examined the unit-roots of four RER indices by taking into account one structural break from 1970 quarter 4 to 1995 quarter 2. [6] estimated a bevy of unit-root tests which include: [47], [39] Innovational Outlier (IO) and Additive Outlier (AO) models, and [38] AO model and IO models I and II.

Using the [43] general-to-specific search procedure, [6] found [38] AO model was the optimal model. His findings show that Trade-weighted index (TWI), Export-weighted index (EWI) and Import-weighted index (IWI) are*stationary* while G7-GDP weighted index is non-stationary. The structural break dates for these variables are 1990 quarter 3 for TWI; 1991 quarter 3 for EWI; 1989 quarter 2 for IWI and 1982 quarter 4 for G7-GDP respectively. [6] result reverses the result obtained by [20]. In addition, [6] and [20] report the break date without reporting the statistical significance.

Importantly, unit-root tests in the above studies, which either do not allow for a break under the null hypothesis such as [47] or model the break as an Innovational Outlier (IO) as [38], suffer from severe spurious rejections in finite samples when a

[7] The null hypothesis of DF test is non-stationary. It is only in the KPSS test that the null hypothesis is stationary.

[8] The conceptually correct method for calculating an RER index has been described by [14]. Hence, the result obtained by [1] is suspect.

break is present under the null hypothesis ([25]; [24]). Because the spurious rejections are not present in the case of a known break point, [25] identify the inaccurate estimation of the break date as source of the incorrect rejections. Furthermore, [25] found that the asymptotic null distributions of the DF-type endogenous break test statistics are affected by nuisance parameters.

This shallow evidence in the Australian literature highlights the difficulties of detecting robust evidence in favour of, or against, the PPP theory. A summary of past results is given in Table 1 for a ready reference. Therefore, further research is warranted to determine if PPP provides a valid representation of the long-run equilibrium relation between the exchange rate and relative prices in Australia by exploring the possibility of including multiple structural breaks. The next section is devoted to this particular aspect.

Table 1 Summary of Previous Results of Unit-root in the Australian RER

Author(s)	Finding	Data Source	Sample Period	Test Method
Blundell-Wignall & Gregory (1990)	NS	Authors calculation with OECD data	1970:1 to 1988:4	ADF
Blundell-Wignall & Fahrer & Heath (1993)	NS	RBA data.	1973:2 to 1992:3	ADF
Gruen & Wilkinson (1994)	NS	RBA data.	1969:4 to 1990:4	ADF
Gruen & Shuetrim (1994)	S ard.a trend	RBA data.	1970:1 to 1993:4	ADF
Gruen & Kortian (1996)	Ambiguous	RBA data.	1984:1 to 1993:4	ADF & others
Tarditi (1996)	S ard. a trend	RBA data.	1973:4 to 1995:2	ADF & others
Chand (2001)	S	RBA data.	1981:3 to 2000:4	ADF
Bagchi et al. (2004)	NS	Authors' calculation with IFS data.	1973:1 to 1995:3	ADF
Henry & Olekalns (2002)	NS	Authors' calculation. Data source unknown.	1973:1 to 1999:1 Single break date @: 1984:1	Zivot & Andrews (1992), Vogelsang (1997)
Chowdhury (2007)	S	RBA data.	1970:4 to 1995:2 Single break date@: 1990:3	Perron (1997) AO Model & 4 other unit-root tests

Note: S = Stationary; NS = Non-stationary; @=Assume no break under the null hypothesis of unit root.

3 Time-Series Properties of RER in the Presence of Structural Breaks

3.1 Data and Data Source

We performed the LS minimum Lagrange Multiplier (LM) unit-root tests to determine structural breaks endogenously. The LS unit-root test with two structural breaks endogenously determines the location of two breaks in level and trend and tests the null of a unit-root. The LS unit-root test with two structural breaks is invariant to the magnitude of the breaks. LS noted that the alternative of the minimum LM unit-root test with two structural breaks unambiguously implies trend stationarity; however, it could be true that the series can possess unit-root with structural breaks.

Unit-root tests for one (LS1) and two breaks (LS2) were conducted with RATS 7.2. We estimated two models: LS-Break Model and LS-Crash Model. The LS-Break Model captures the change that is gradual whereas LS-Crash Model picks up the change that is rapid. We have reported the results of both models in Table 2 which are contradictory to each other. The result of the unit-root test is contingent upon the way the breaks are modelled. The choice of the "model" should be based on economic theory and reality. Based on our judgement, we think the LS Trend Break model is the optimal model to discuss.

On the basis of LS1unit-root test we find LnRER to be stationary. By applying the LS2 unit-root test we found that LnRER is also stationary. Rejection of the unit-root null provides evidence of mean reversion and hence PPP.

Table 2 Unit-Root Tests in the Absence and Presence of Structural Breaks

Variable:LnRER		**Traditional Unit Root Tests**			
Test	τ	Time of Break1	Time of Break2	k	Decision
ADF	-2.425	NC	NC	2	NS
Elliot et al.	399.551	NC	NC	2	S
Ng-Perron5	30.418	NC	NC	4	S
KPSS	0.184	NC	NC	5	NS
Variable:LnRER		**LS-Break Model Result**			
Test	τ	Time of Break1	Time of Break2	k	Decision
LS1	-3.568*	2003:2***	NC	5	S
LS2	-3.877**	1998:2**	2002:4***	5	S
Variable:LnRER		**LS-Crash Model Result**			
Test	τ	Time of Break1	Time of Break2	k	Decision
LS1	-2.334	1989:1	NC	5	NS
LS2	-2.714	1998:1*	1995:1**	5	NS

Note:

1. NC = Not calculated; S = Stationary, NS = Nonstationary.
2. t-statistic for the null hypothesis =0.
3. DF Test critical values at 1, 5 and 10 per cent level are -4.054; -3.456 and -3.153 respectively.
4. Critical values of the endogenous two-break LM unit-root test at 10%, 5% and 1% level of significance are -3.504, -3.842 and -4.545 respectively from Table 2 Lee and Strazicich (2003:1084).
5. We report the first unit root test statistic developed by Ng and Perron which is the Elliot, Rothenberg, and Stock (1996) point optimal statistic for GLS de-trended data. The other three statistics, MZ_{α}^{d} ,MZ_{t}^{d} and MSM^{d} are the enhancements of the Phillips-Peron (PP) test statistics, which are not reported here.
6. (*), (**) and (***) refer to significant at 10, 5 and 1 per cent level of significance respectively.

ADF test fails to reject the null hypothesis for LnRER (refer to Table 2). The GLS test proposed by [13] and M-test suggested by [31] reject the null of a unit-root for LnRER. However, based on the KPSS test we reject the null hypothesis of stationarity for LnRER. On balance, the evidence in Table 2 is inconclusive.

3.2 Lee and Strazicich (2003) (LS) Unit-Root Test

We performed the LS minimum Lagrange Multiplier (LM) unit-root tests to determine structural breaks endogenously. The LS unit-root test with two structural breaks endogenously determines the location of two breaks in level and trend and tests the null of a unit-root. The LS unit-root test with two structural breaks is invariant to the magnitude of the breaks. LS noted that the alternative of the minimum LM unit-root test with two structural breaks unambiguously implies trend stationarity; however, it could be true that the series can possess unit-root with structural breaks.

Unit-root tests for one (LS1) and two breaks (LS2) were conducted with RATS 7.2. We estimated two models: LS-Break Model and LS-Crash Model. The LS-Break Model captures the change that is gradual whereas LS-Crash Model picks up the change that is rapid. We have reported the results of both models in Table 2 which are contradictory to each other. The result of the unit-root test is contingent upon the way the breaks are modelled. The choice of the "best model" should be based on economic theory and reality. Based on our judgement, we think the LS Trend Break model is the optimal model to discuss.

On the basis of LS1unit-root test we find LnRER to be stationary. By applying the LS2 unit-root test we found that LnRER is also stationary. Rejection of the unit-root null provides evidence of mean reversion and hence PPP.

3.3 Endogenously Determined Structural Break Dates

The estimated single structural break date determined by the LS1 Break Model corresponds to 2003 quarter 2 for LnRER. The break date is statistically significant at the 5 per cent level. By considering the two breaks LS2 Trend Break Model, the corresponding break dates for LnRER are 1988:2 and 2002:4. The structural break dates are all statistically significant. The first break date of LnRER coincides with the abandonment of the "check-list" approach in favour of "discretionary" approach to monetary policy by RBA in 1988 quarter 2. This structural break may also be capturing the effect of the stock market crash of October 1987, and the onset of recession at the end of the 1980s culminating into the recession in 1990. The behaviour of the Australian RER shows periods of instability. One such period was centred around June 1986, the other between March 1998 and June 1999. After a sustained period of depreciation, appreciations of the RER occurred during 1986-1989 so that the break date for the RER is picked up in 1988 quarter 2 followed by the meltdown in 2001 and again a recovery in early 2002. The second break date is found to be in 2002 quarter 4 which is due to the sudden appreciation of the Australian dollar. Between January 2002 and July 2008, the Australian dollar appreciated sharply from 51 US cents to 97 US cents which was largely driven by increased demand for Australian exports.

4 Summary and Conclusion

We investigate evidence of mean reversion in the Australian dollar RER. Conventional unit-root tests fail to provide evidence of stationarity of RER. If RER is non-stationary, then PPP is no longer valid as a representation of the long-run equilibrium relation between the exchange rate and relative prices. The conventional unit-root tests may suffer from severe size distortions and results might be erroneous since they do not account for structural breaks in the data. To overcome the loss of power in conventional unit-root tests, we performed the [24] minimum Lagrange Multiplier unit-root tests in the presence of structural breaks.

Based on our result, we were able to reject the unit-root null hypothesis and find evidence of mean reversion and hence PPP. This result is consistent with [6] finding although the break dates are different. This finding reverses the findings of past works that failed to reject non-stationarity. The corresponding break dates for RER are 1988 quarter 2 and 2002 quarter 4 respectively; and the break dates are all statistically significant. The estimated break dates mostly correspond to the period of RER instability (1986-1989) and the recovery of the Australian dollar driven by the resources boom (2001-2002).

Appendix

A Brief Review of Unit-root Tests[9]

Traditional (First Generation Models) tests for unit-roots (such as Dickey-Fuller, Augmented Dickey-Fuller and Phillips-Perron) have low power in the presence of structural break. [37] demonstrated that, in the presence of a structural break in time-series, many perceived non-stationary series were in fact stationary. [37] re-examined [32] data and found that 11 of the 14 important US macroeconomic variables were stationary when known exogenous structural break is included[10]. [37] allows for a one time structural change occurring at a time $T_B(1 < T_B < T)$, where T is the number of observations.

The following models were developed by [37] for three different cases. Notations used in equations 2–19 are the same as in the papers quoted. Null Hypothesis:

$$Model(A)\ y_t = \mu + dD(TB)_t + y_{t-1} + e_t \tag{2}$$

$$Model(B)\ y_t = \mu_t + y_{t-1} + (\mu_2 - \mu_1)DU_t + e_t \tag{3}$$

$$Model(C)\ y_t = \mu_t + y_{t-1} + dD(TB)_t + (\mu_2 - \mu_1)DU_t + e_t \tag{4}$$

Where $D(TB)_t = 1$ if $t = T_B + 1,\ 0$ otherwise, and $DU_t = 1$ if $t > T_B, 0$ otherwise.

[9] The discussion that follows is for reference only and may be omitted.

[10] However, subsequent studies using endogenous breaks have countered this finding with [47] concluding that 7 of these 11 variables are in fact non-stationary.

Alternative Hypothesis:

$$Model(A)\ y_t = \mu_t + \beta t + (\mu_2 - \mu_1)DU_t + e_t \tag{5}$$

$$Model(B)\ y_t = \mu + \beta_t t + (\beta_2 - \beta_1)DT_t^* + e_t \tag{6}$$

$$Model(C)\ y_t = \mu + \beta_1 t + (\mu_2 - \mu_1)DU_t + (\beta_2 - \beta_1)DT_t + e_t \tag{7}$$

Where $DT_t^* = t - T_B,\ if\ t > T_B$, and 0 otherwise.

Model A permits an exogenous change in the level of the series whereas Model B permits an exogenous change in the rate of growth. Model C allows change in both. [37] models include one known structural break. These models cannot be applied where such breaks are unknown. Therefore, this procedure is criticised for assuming known break date which raises the problem of pre-testing and data mining regarding the choice of the break date [28]. Further, the choice of the break date can be viewed as being correlated with the data.

Second Generation Models

Unit-Root Tests in the Presence of a Single Endogenous Structural Break

Despite the limitations of [37] models, they form the foundation of subsequent studies that we are going to discuss hereafter. [47], [39], and [38] among others have developed unit-root test methods which include one endogenously determined structural break. Here we review these models briefly and detailed discussions are found in the cited works.

[47] models are as follows:

Model with Intercept

$$y_t = \hat{\mu}^A + \hat{\theta}^A DU_t(\hat{\lambda}) + \hat{\beta}^A t + \hat{\alpha}^A y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^A \triangle y_{t-j} + \hat{e}_t \tag{8}$$

Model with Trend

$$y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\gamma}^B DT_t^*(\hat{\lambda}) + \hat{\alpha}^B y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^B \triangle y_{t-j} + \hat{e}_t \tag{9}$$

Model with Both Intercept and Trend

$$y_t = \hat{\mu}^C + \hat{\theta}^C DU_t(\hat{\lambda}) + \hat{\beta}^C t + \hat{\gamma}^C DT_t^*(\hat{\lambda}) + \hat{\alpha}^C y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^C \triangle y_{t-j} + \hat{e}_t \tag{10}$$

where,

$$DU_t(\alpha) = 1 \text{ if } t > T\alpha, 0 \text{ otherwise}; DT_t^*(\lambda) = t - T\lambda \text{ if } t > T\lambda, 0 \text{ otherwise}.$$

The above models are based on [37]'s models. However, these modified models do not include DT_b.

On the other hand, [39] include DTb but exclude t in their models. [39] models are given below:

Innovational Outlier Model (IOM)

$$y_t = \mu + \delta DU_t + \theta D(T_b)_t + \alpha y_{t-1} + \sum_{j=1}^{k} c_i \triangle y_{t-i} + e_t \tag{11}$$

Additive Outlier Model (AOM) - Two Steps

$$y_t = \mu + \delta DU_t + \tilde{y}_t \tag{12}$$

and

$$\tilde{y}_t = \sum_{j=0}^{k} w_t D(T_b)_{t-1} + \alpha \tilde{y}_{t-1} + \sum_{j=1}^{k} c_j \triangle \tilde{y}_{t-j} + e_t \tag{13}$$

$\tilde{y}$ in the above equations represents a detrended series y.

[38] includes both t (time trend) and DT_b (time at which structural change occurs) in his Innovational Outlier (IO1 and IO2) and Additive Outlier (AO) models.

Innovational Outlier Model allowing one time change in intercept only (IO1):

$$y_t = \mu + \theta DU_t + \beta t + \gamma D(T_b)_t + \alpha y_{t-1} + \sum_{j=1}^{k} c_i \triangle y_{t-j} + e_t \tag{14}$$

Innovational Outlier Model allowing one time change in both intercept and slope (IO2):

$$y_t = \mu + \theta DU_t + \beta t + \gamma D(T_b)_t + \gamma D(T_b)_t + \alpha y_{t-1} + \sum_{j=1}^{k} c_i \triangle y_{t-j} + e_t \tag{15}$$

Additive Outlier Model Allowing One Time Change in Slope (AO)

$$y_t = \mu + \beta t + \gamma DT_t^* + \tilde{y}_t \tag{16}$$

where $DT_t^* = 1(t > T_b)(t - T_b)$

$$\tilde{y}_t = \alpha \tilde{y}_{t-1} + \sum_{j=1}^{k} \hat{c}_j^C \triangle \tilde{y}_{t-j} + e_t \tag{17}$$

The Innovational Outlier models represent the change that is gradual whereas Additive Outlier model represents the change that is rapid.

Regarding the power of tests, the [39] model is robust. The testing power of [38] and [47] models are almost the same. On the other hand, [38] model is more comprehensive than [47] model as the former includes both t and DTb while the latter includes t only.

Additional test methods have been proposed for unit-root test allowing for multiple structural breaks in the data ([27] (LP); [24] (LS)). One important issue common to the ZA and LP (and other similar) endogenous break tests is that they assume no break(s) under the unit-root null and derive their critical values accordingly. Thus, the alternative hypothesis would be "structural breaks are present" which includes the possibility of a unit-root with break(s). Thus, rejection of the null does not necessarily imply rejection of a unit-root per se, but would imply rejection of a unit-root without breaks.

Third Generation Models

Lee and Strazicich (LS) (2003) Minimum LM Unit-Root Test

LS propose a minimum Lagrange multiplier (LM) unit-root test in which the alternative hypothesis unambiguously implies trend stationarity. Consider the DGP as follows:

$$\triangle y_t = \delta' + \triangle Z_t + \phi \tilde{S}_{t-1} + u_t \tag{18}$$

where $\tilde{S}_t = y_t - \tilde{\psi}_x - Z_t \tilde{\delta}(t = 2, \ldots T$ and is a vector of exogenous variables defined by the data generating process; $\tilde{\delta}$ is the vector of coefficients in the regression of $\triangle y_t$ on $\triangle Z_t$ respectively with $\triangle$ the difference operator; and $\hat{\psi}_x = y_1 - Z_1 \tilde{\delta}$, with y_1 and Z_1 the first observations of y_t and Z_t respectively.

Model B of Perron (1989) is omitted by LS (2003), as it is commonly held that most economic time-series can be adequately described by model A or C. Equivalent to Perron's (1989) Model C, which allows for a shift in intercept and change in trend slope under the null hypothesis and is described as $Z_t = [1, t, D_t, DT_t]'$, where $DT_t = t - T_B$ for $t > T_B + 1$, for $t > TB + 1$, and zero otherwise. It is important to note here that testing regression (18) involves using $\triangle Z_t$ instead of Z_t. $\triangle Z_t$ is described by $[1, B_t D_t]'$ where $B_t = \triangle D_t$ and $D_t = \triangle DT_t$. Thus, B_t and D_t correspond to a change in the intercept and trend under the alternative and to a one period jump and (permanent) change in drift under the null hypothesis, respectively.

The unit-root null hypothesis is described in (18) by $\phi = 0$ and the LM t-test is $\tilde{\tau} = t$ given by; where $\tilde{\tau} = t -$ statistic for the null hypothesis $\phi = 0$.

The augmented terms $\triangle \tilde{S_{t-j}}, j = 1, \ldots k$, terms are included to correct for serial correlation. The value of k is determined by the general to specific search procedure. General to specific procedure begins with the maximum number of lagged first differenced terms max $k = 8$ and then examine the last term to see if it is significantly different from zero. If insignificant, the maximum lagged term is dropped and then estimated at $k = 7$ terms and so on, till the maximum is found or $k = 0$. To endogenously determine the location of the break (T_B), the LM unit-root searches for all possible break points for the minimum (the most negative) unit-root t-test statistic as follows:

$$\inf \tilde{\tau} = \inf_{\lambda} \tilde{\tau}(\lambda); \text{ where } \lambda = T_B/T. \tag{19}$$

The two-break LM unit-root test statistic can be estimated analogously. Critical values of the two-break LM unit-root test (T = 100) is reported in Table 3 by LS. LS (2003: 1087) conclude "summary, the two-break minimum LM unit-root test provides a remedy for a limitation of the two-break minimum LP test that includes the possibility of a unit-root with break(s) in the alternative hypothesis. Using the two-break minimum LM unit-root test, rejection of the null hypothesis unambiguously implies trend stationarity."

References

1. Bagchi, B., Chortareas, G.E., Miller, S.M.: The Real Exchange Rate in Small, Open Developed Economies: Evidence from Cointegration Analysis. Economic Record 80(248), 76–88 (2004)
2. Baum, C.F., Barkoulas, J.T., Caglayan, M.: Nonlinear Adjustment to Purchasing Power Parity in the Post-Bretton Woods Era. Journal of International Money and Finance 20, 379–399 (2001)
3. Blundell-Wignall, A., Fahrer, J., Heath, A.: Major Influences on the Australian Dollar Exchange Rate. In: Blundell-Wignall, A. (ed.) The Exchange Rate, International Trade and the Balance of Payments, Proceedings of a Conference, Reserve Bank of Australia, pp. 30–78 (1993)
4. Blundell-Wignall, A., Gregory, R.G.: Exchange Rate Policy in Advanced Commodity Exporting Countries: Australia and New Zealand, "Exchange Rate Policy in Advanced Commodity-Exporting Countries: The Case of Australia and New Zealand", OECD Economics Department Working Papers, No. 83 (1990)
5. Chand, S.: How Misaligned is the Australian Real Exchange Rate? Working Paper Number 012, Asia Pacific School of Economics and Government, Australian National University, Canberra (2001)
6. Chowdhury, K.: Are the Real Exchange Rate Indices of Australia Non-Stationary in the Presence of Structural Break? International Review of Business Research Papers 3(5), 161–181 (2007)
7. Corbae, D., Ouliaris, S.: A Test of Long-Run Purchasing Power Parity Allowing for Structural Breaks. The Economic Record 67, 26–33 (1990)
8. Cuestas, J.C., Gil-Alana, L.A.: Further Evidence on the PPP Analysis of the Australian Dollar: Non-Linearities. Fractional Integration and Structural Changes Economic Modelling 26, 1184–1192 (2009)
9. Cuestas, J.C., Regis, J.C.: Testing for PPP in Australia: Evidence from Unit Root Tests against Nonlinear Trend Stationarity Alternatives. Economics Bulletin 27, 1–8 (2008)
10. Darné, O., Hoarau, J.F.: The Purchasing Power Parity in Australia: Evidence from Unit Root Test with Structural Break. Applied Economics Letters 15, 203–206 (2008)
11. Dornbusch, R.: Expectations and Exchange Rate Dynamics. Journal of Political Economy 84, 1161–1176 (1976)
12. Edison, H., Cashin, P., Liang, H.: Foreign Exchange Intervention and the Australian Dollar: Has it Mattered? IMF Working Paper, WP/03/99 (1999)
13. Elliot, G., Rothenberg, T.J., Stock, J.H.: Efficient Tests for an Autoregressive Unit Root. Econometrica 64, 813–836 (1996)

14. Ellis, L.: Measuring the Real Exchange Rate: Pitfalls and Practicalities Research Discussion Paper 2001-04, Economic Research Department, Reserve Bank of Australia (2001)
15. Frenkel, J.A.: The Collapse of Purchasing Power Parities during the 1970s. European Economic Review 16, 145–165 (1981)
16. Granger, C.W.J., Terasvirta, T.: Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford (1993)
17. Gruen, D.W.R., Wilkinson, J.: Australia's Real Exchange Rate Is It Explained by the Terms of Trade or by Real Interest Rate Differential? The Economic Record 70, 204–219 (1994)
18. Gruen, D.W.R., Kortian, T.: Why does the Australian Dollar Move with the Terms of Trade? Research Discussion Paper 9601, Economic Group, Reserve Bank of Australia (1996)
19. Gruen, D.W.R., Shuetrim, G.: Internationalisation and the Macroeconomy. In: Lowe, P., Dwyer, J. (eds.) International Integration of the Australian Economy. Proceedings of a Conference, pp. 309–363. Reserve Bank of Australia, Sydney (1994)
20. Henry, O.T., Olekalns, N.: Does the Australian Dollar Real Exchange Rate Display Mean Reversion. Journal of International Money and Finance 21, 651–666 (2002)
21. Henry, O.T., Olekalns, N.: Summers PM Exchange Rate Instability: A Threshold Autoregressive Approach. The Economic Record 77, 160–166 (2001)
22. Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.: Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometrics 55, 159–178 (1992)
23. Jones, M.T., Wilkinson, J.: Real Exchange Rates and Australian Export Competitiveness, Discussion Paper No. 9005, Economics Research Department, Reserve Bank of Australia (1990)
24. Lee, J., Strazicich, M.: Minimum Lagrange Multiplier Unit Root Test with Two Structural Breaks. The Review of Economics and Statistics 85(4), 1082–1089 (2003)
25. Lee, J., Strazicich, M.: Break Point Estimation and Spurious Rejections with Endogenous Unit Root Tests. Oxford Bulletin of Economics and Statistics 63(5), 535–558 (2001)
26. Liew, K.S., Baharumshah, A.Z., Lau, E.: Nonlinear Adjustment Towards Purchasing Power Parity in ASEAN Exchange Rates. International Journal of Applied Economics 3(6), 7–18 (2004)
27. Lumsdaine, R., Papel, D.H.: Multiple Trend Breaks and the Unit Root Hypothesis. Review of Economics and Statistics 79, 212–218 (1997)
28. Maddala, G.S., Kim, I.M.: Unit Roots, Cointegration, and Structural Change. Cambridge University Press, Cambridge (2003)
29. MacDonald, R.: Panel Unit Root Tests and Real Exchange Rates. Economic Letters 50, 7–11 (1996)
30. Michael, P., Nobay, A.R., Peel, D.A.: Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. Journal of Political Economy 105, 862–879 (1997)
31. Ng, S., Perron, P.: Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. Econometrica 69, 1519–1554 (2001)
32. Nelson, C., Plosser, C.: Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. Journal of Monetary Economics 10, 139–162 (1982)
33. Olekalns, N., Wilkins, N.: Re-examining the Evidence for Long-Run Purchasing Power Parity. The Economic Record 74, 54–61 (1998)
34. Papell, D.: Searching For Stationarity: Purchasing Power Parity under the Current Float. Journal of International Economics 43, 313–332 (1997)

35. Papell, D., Theodoridis, H.: The Choice Of Numeraire Currency in Panel Tests of Purchasing Power Parity. Journal of Money, Credit, and Banking 33, 790–803 (2001)
36. Papell, D., Theodoridis, H.: Increasing Evidence of Purchasing Power Parity over the Current Float. Journal of International Money and Finance 17(1), 41–50 (1998)
37. Perron, P.: The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. Econometrica 57, 1361–1401 (1989)
38. Perron, P.: Further Evidence on Breaking Trend Functions in Macroeconomic Variables. Journal of Econometrics 80, 355–385 (1997)
39. Perron, P., Vogelsang, T.J.: Nonstationary and Level Shifts with an Application to Purchasing Power Parity. Journal of Business and Economic Statistics 10, 301–320 (1992)
40. Phillips, P.C.B., Perron, P.: Testing for a Unit Root in Time Series Regression. Biometrica 75, 335–346 (1988)
41. Sarno, L.: Real Exchange Rate Behaviour in High Inflation Countries: Empirical Evidence From Turkey, 1980-1997. Applied Economics Letters 7, 285–291 (2000a)
42. Sarno, L.: Real Exchange Rate Behaviour in the Middle East: A Re-Examination. Economics Letters 66, 127–136 (2000b)
43. Shrestha, M.B., Chowdhury, K.: A Sequential Procedure for Testing Unit Roots in the Presence of Structural Break in Time Series Data: An Application to Quarterly Data in Nepal, 19702003. International Journal of Applied Econometrics and Quantitative Studies 2(2), 1–16 (2005)
44. Tarditi, A.: Modelling the Australian Exchange Rate, Long Bond Yield and Inflationary Expectations, Research Discussion Paper 9608, Economic Analysis Department, Reserve Bank of Australia (1996)
45. Taylor, M.P., Peel, D.A.: Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals. Journal of International Money and Finance 19, 33–53 (2000)
46. Vogelsang, T.J.: Wald-Type Test for Detecting Breaks in the Trend Function of a Dynamic Time Series. Econometric Theory 13, 818–849 (1997)
47. Zivot, E., Andrews, D.W.: Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypotheses. Journal of Business and Economic Statistics 10, 251–270 (1992)

An Empirical Analysis of Price Behavior of Natural Rubber Latex: A Case of Central Rubber Market Hat Yai, Songkhla, Thailand

Hari Sharma Neupane and Peter Calkins

Abstract. Hat Yai City in Songkhla Province, Thailand has three unique advantages: its Central Rubber Market lies in the largest rubber growing region in the world, it can easily access the new (2004) deep-sea port in Songkhla, and it lies directly on the improved transport infrastructure of the Asia Highway and the North-South Economic Corridor linking it to other growing areas in Southeast Asia and Thailand. Despite these advantages, the rubber industry has always been susceptible to the price volatility of rubber latex, which destabilizes the benefits of rubber production to the local economy, particularly to small-holder producers. Since volatility may theoretically either decrease in the future with the integration of more numerous supplying regions or increase with the intensified co-dependence of supplying and demanding countries, careful modeling of rubber price volatility on the Hat Yai market could both inform development policy today and serve as a baseline for future studies.

This paper therefore attempts to identify the best econometric model to capture price volatility of latex type RSS3 in Thailand for the period 2004-2011. The daily price of latex type RSS3 was modeled by adopting and comparing conditional volatility models, GARCH, GARCH-GJR and EGARCH. The price volatility of natural rubber latex type RSS3 is strongly persistent, and the estimated results are statistically valid. If implemented, the findings of this paper with respect to economic, environmental, and transportation policy could lead to benefits to small holders and to price stabilization mechanisms on national and export.

Keywords: Natural rubber latex, price volatility, smallholders, livelihood.

JEL codes: C22, N50

Hari Sharma Neupane · Peter Calkins
Faculty of Economics, Chiang Mai University, Thailand
e-mail: hsneupane@yahoo.com

Peter Calkins
CREATE, Laval University, Canada
e-mail: peter_calkins2004@yahoo.com

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 185–201.
DOI: 10.1007/978-3-642-35443-4_13

1 Historical Background of Rubber Production and Marketing

Rubber plantations are established in many tropical nations as an indispensable resource for a prosperous life and other economic and environmental benefits (Dean 1987). Over the 21-year period 1990-2011, global rubber production has increased from 15 million to 26 million tons, while rubber consumption has rose from 14.8 million to 25.8 million tons. These figures account for the 2.8% average annual increment in both rubber production and consumption in the world (Table 1).

Table 1 Growth, shares, consumption and production of rubber in the world

Production and consumption of rubber by type	1990	1995	2000	2005	2011	Growth (%)
a. Rubber production (million tons)	15.01	15.55	17.63	20.98	26.09	2.80%
NR production (million tons)	5.12	6.07	6.76	8.90	10.97	3.88%
Share of NR (%)	34.11	39.04	38.35	42.45	42.06	
Share of SR (%)	65.89	60.96	61.65	57.55	57.94	
b. Rubber consumption (million tons)	14.87	15.22	18.17	20.99	25.85	2.80%
NR consumption (million tons)	5.12	5.95	7.34	9.21	10.92	3.86%
Share of NR (%)	35.04	39.09	40.40	43.21	43.35	
Share of SR (%)	64.96	60.91	59.60	56.79	56.65	
C. Stock(million tons)	0.14	0.33	-0.54	-0.01	0.24	2.71%
Percent of total Production	0.93	2.12	-3.06	-0.05	0.92	

Growth = Average annual growth rate

Table 2 Natural rubber consumption (000' mt.) by major economies (2002-2010)

Countries	2002	Percent of world	2006	Percent of world	2010	Percent of world	Growth (%)
China	1,395	18.47	2,769	28.58	3646	33.83	12.76
Thailand	278	3.69	321	3.31	459	4.26	6.44
South Korea	326	4.31	364	3.75	384	3.56	2.08
Germany	247	3.27	269	2.78	291	2.70	2.08
Japan	749	9.92	874	9.02	750	6.96	0.02
Malaysia	408	5.40	383	3.96	458	4.25	1.45
USA	1,111	14.71	1,003	10.35	926	8.59	-2.26
World	**7,552**	**100.00**	**9,690**	**100.00**	**10778**	**100.00**	**4.55**

Growth = Average annual growth rate

The production of NR has risen by an average annual rate of 3.9% and that of shares in total rubber production by 2.8% between 1990 and 2011. The consumption of NR has also followed the same trend. As demand for NR has grown, the shares of SR production and consumption have dropped by almost 10% (decline from 66% to reach 56%) and nearly 9% (drop from 64.96% to reach 56.65%), with an average annual growth rate of only 2.49%. Likewise, China consumed more than 33% of

NR produced in the world, the USA more than 10%, Japan 9% and Thailand 4%, respectively, in 2010 (Table 2)[1]. Conversely, the average annual growth trend of NR consumption was negative in the USA (-2.26%) over the period 2002-2010. The average annual growth of global NR consumption was 3.68% in the same period. The upward trend in NR demand was fueled by increased consumption by emerging economies, averaging 12.8% per year in China and 6.5% in Thailand.

2 Problems of Price Volatility and Market Concentration

The demand for both natural rubber (NR) and synthetic rubber (SR) is well secured and growing apace with improvements in living standards around the world (FAO 2002). However, the price of latex is quite volatile. Since the price of the natural input is a critical determinant of industrial demand for NR to produce rubber tires and tubes for motor vehicles; footwear, belts and hoses for wire cable industries (Grilli 1980), such volatility can wreak havoc with profitability, planning and worker incomes. FAO (2002) mentioned that SR is a purely petroleum-based industrial product made from monomers and subject to price volatility, whose production and consumption are dominated by large global enterprises. Therefore, economic planners need to know the relative volatility of NR and SR prices, as well as the best means of controlling excessive market price volatility in NR. Realizing this need, the first International Natural Rubber Agreement (INRA) was signed in 1979 under the support of the United Nations Conference on Trade and Development (UNCTAD) and agreement was updated periodically (UNCTAD 2007). Market regulation in the NR sector included the creation and management of international buffer stock schemes envisaged under the 1979, 1987 and 1995 INRAs. These schemes gave due consideration to increasing producer benefits, specifically for smallholders, and to stabilizing and increasing export earnings through expanded export volumes. The East Asian International Tripartite Rubber Organization (ITRO) established in 2001, aims to manage NR production through the maintenance of orderly market growth and guaranteed minimum price to domestic producers (UNCTAD 2007). Although NR latex or rubber smoked sheets (RSS) serve as both raw products for industry and in rubber items; they are perfect substitutes for the end user (Saidur and Mekhilef 2010). In fact, such items could be made either from NR, SR or blends of both in various proportions. Therefore, manufacturers determine the type of rubber used in the rubber items based on technological advantage, product availability, current market prices and, as noted, price volatility.

At present, vast land areas are already covered by fast-growing tree plantations where the dominant crop is *Hevea* in Africa, Asia, Latin America, Oceania, the southern USA and even some European countries (Spain and Portugal) (TERRA 2004). Perhaps, Thailand is the country that has most pioneered large-scale NR plantations, which were eventually extended to other countries of the region. Over 20 millions of families are dependent on NR plantation for their basic income in

[1] Source: `http://www.thainr.com/en`, and Office of Agricultural Economics, Thailand.

the world NR market (Khin et al. 2008). The production of NR has a major role in the socioeconomic development of the producing countries or regions and is the sources of livelihood for smallholders, especially in and around the west Malaysian peninsula. The NR industry has always been susceptible to crisis induced by fluctuating prices and manufacturers are most vulnerable to raw materials, production and export of rubber based articles. Hence, economic shocks to the rubber marketing channel have always had major impact upon not only the economy but also the social and political stability of the region (Allen 2004; Stubbs 1983). Fluctuations in both farm gate prices and macro exchange rates have exacerbated the food insecurity of millions of rural dwellers and low prices paid for NR contribute to rural poverty in many countries, especially smallholders in South East Asia where currency turmoil has greatly diminished the purchasing power for essentials like medicines (Khin et al. 2008).

3 Natural Rubber Production and Export in Thailand

Natural rubber plantation in Thailand has been promoted by Thai Government from 1961 onwards through special policies and programs. Since 1991, Thailand has been the world's largest producer and exporter in both NR and rubber articles (TRA, 2007). The Siam peninsula accounted for more than 31 and 34% of world production and exports, respectively, in 2010. Nevertheless, the total share of Thailand's production in the world rubber market has been reduced by 4.5% due to a relatively weak average annual growth rate of 2.76 % over the period 2002-2010. Meanwhile, consumption of NR grew by 5.22% and exports grew by 3.95%. The economic contribution of the NR industry to Thai export receipts increased significantly, from 74.6 billion in 2002 to 249.3 billion TB in 2010, with an average annual growth rate of 16.27 % (Table 3)[2].

4 The Study Site and Purpose of Study

Songkhla province is one of the leading NR growing provinces in the south. The monoculture production system has replaced a traditional system of rubber forests, where rubber used to be grown as an intercrop within fruit orchards and natural forests known as a 'suan somrom' meaning "integrated garden" (TRA 2007). The rubber plantation has been promoted through governments' welfare fund and land use pattern now depicts that some 61% of total land cover of the province is occupied by Para rubber plantations. Forests including shrubs and grasses constitute 14.5%; rice fields hold 12%; agriculture, pasture, mines and other activities comprise nearly 7%; and fisheries, wetlands, and water bodies occupy about 4% of the total land cover. The remaining 2% is devoted to urban and rural settlements and institutional land (Figure 1)[3].

[2] Source: `http://www.thainr.com/en` and `http://www.thainr.com/en`

[3] Source: GIS data file is obtained from Prince of Songhkla University, Hat Yai, Thailand

Table 3 Natural Rubber production and Exports in Thailand

Description	2002	2006	2010	Growth (%)*
1. Production of NR in world ('000 mt.)	7,326	9,827	10,401	4.48
1.1. Production of NR in Thailand ('000 mt.)	2,615	3,137	3,252	2.76
1.2. Thailand's share in world market (%)	35.70	31.92	31.27	-
2. Export of NR in world market ('000 mt.)	5,782	7,597	7,881	3.95
2.1. Export of NR from Thailand ('000 mt.)	2,354	2,772	2,683	1.65
2.2. Share in world market (%)	40.72	36.48	34.05	-
3. Thailand's export value of NR (billion Thai Bhat)	74.61	205.36	249.26	16.27

Average annual growth rate

The NR industry is one of the key sectors of the provincial economy of the Songkhla, with extensive impacts on economic, social and environmental systems. In the 112 years since NR was first introduced into Thailand from Malaysia through Trang province (Somboonsuke and Cherdchom 2000), it has become an integral part of the cultural and economic life for Southern Thai's (TRA 2007). The Siamese council of Ministers agreed to participate in an international rubber control scheme in 1934 without prior knowledge of dynamics of expansion process or the extent of the planting of small stand of rubber trees in peninsular Siam (Stifel 1973). The TRA argued that "the Para rubber tree has played [sic] a vital role related to or hosts a routine of many smallholders, laboring entrepreneurs and government officers: a great deal of Thai citizens (at least 6 million people or no less than 1 million families of the Country). The worthy botany has brought an immense income for our country that comes from the export of NR products, rubber productivities and rubber wood manufactures of which the values were estimated by over 400,000 million Baht per year".

Rubber production is the dominant source of household income for small rubber growers involved in integrated farming systems and small holding rubber grower has become increasingly prominent in both hectarage and production (Somboonsuke and Shivakoti 2001). A full 93-95% of total rubber farmers directly engaged in NR production are small holders, whereas only 5-7% are estate owners (Kittipol 2008; Somboonsuke and Cherdchom 2000). Thus, small producers are highly vulnerable to market uncertainties (Viswanathan 2006).

Conversely, the rubber sector itself is controlled by large processing plants that purchase the material via local dealers. There are two marketing chains, one for the large plantations owned by large firms; the other for the small producers, with lots of middlemen and collecting centers. Prices are set one month before in the Hat Yai market. There is also the emergence of a rubber futures market. The Agricultural Futures Exchange of Thailand (AFET) is involved in NR futures trading. AFET started trading in the specific product known as Ribbed Smoked Sheet No.3 (RSS3). The aims of the AFET are to provide an efficient system for trading, clearing, settlement

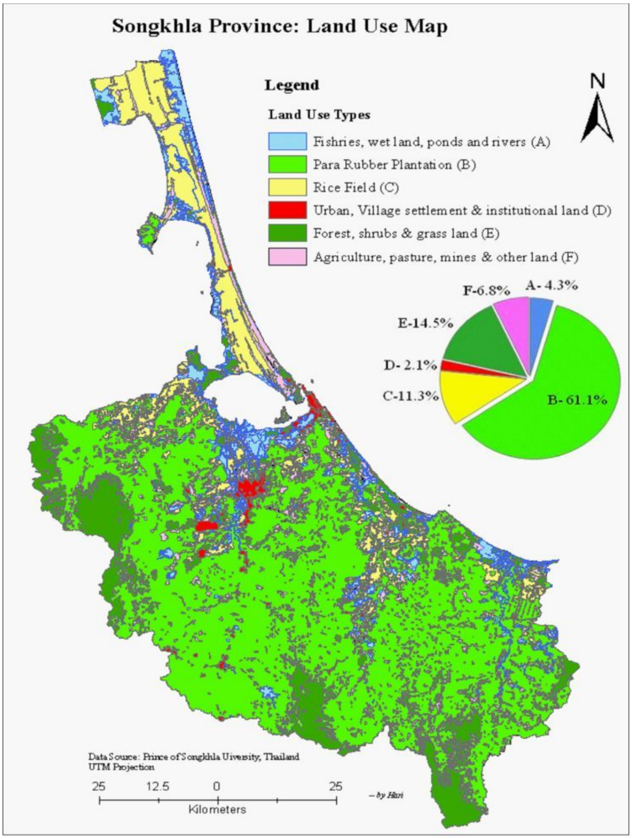


Fig. 1 Landuse use map of Songkhla showing coverage of Para rubber

and deliveries, which will serve as an instrument to hedge the price risks and for future price discovery (IRSG 2010).

Realizing the immense potential of the NR industry in the world market, the Thai government has reformed the country's rubber development system with the establishment of an international standard rubber research institute at Hat Yai. It has even declared that Hat Yai is to be known as "rubber city." The Central Rubber Market Hat Yai (CRMH) has played such a great role in the marketing of NR products that the city is considered as a production and marketing hub for Thai NR products. This role is likely to grow as a result of both the establishment of the Deep-Sea Port in Songkhla (Gov/Thai PRD 2004) and the great amelioration of road transport through the North-South Economic Corridor. The Asian Highway AH 2(No. 4) links the North-South Economic Corridor of the Greater Mekong Sub-region (GMS) at Bangkok to the Indonesia-Malaysia-Thailand growth Triangle, embracing several provinces of southern Thailand and ending in the Sadao Municipality of Songkhla at the Malaysian border. This Asian Highway or its sub-branches serve to directly or indirectly tie together the various agricultural production patches and agro-based industries of southern peninsular Siam and even northeastern Thailand, providing ample opportunities for national and international market through land shipment. Not only does this growing road network connect the Hat Yai market to rubber shipments coming up from Malaysia; it will be able to rapidly and efficiently channel rubber shipments from such Northeastern Thai regions as Mukdahan. Thus, Songkhla province, as an international marketing hub for NR and other agricultural products, has played a leading role in the socioeconomic development of southern Thailand. The province seems to be one of the driving forces or catalysts for economic development of the peninsular Siam.

Even though these infrastructural developments are "good," they add additional elements of uncertainty to rubber supply, demand, trade, and hence price volatility. Measuring such volatility becomes a pressing research priority, which the present paper will endeavor to meet. The price behavior of NR at the CRMH already plays a significant role in provincial economy. It affects the export status of NR extracts and rubber based products on the international market as well as the livelihood of the large number of small rubber holding farmers in the south. The obvious environmental advantages which NR possesses over the synthetic rubbers have never translated into financial advantage: both kinds have for decades suffered from poor prices. Despite this, producing NR remains the main and often the sole source of family income for millions of small farmers around the world (Allen 2004). Thus, we propose to use advanced econometric methods to observe the volatility of the rubber prices and its persistence over time by using the daily bid price of the rubber latex at the CRMH.

Fundamentally, the price of NR depends upon the global demand of NR and on the world petroleum price, which directly affects the cost of SR production. But changes in the prices of rubber are also altered by market fundamentals. Improvement in the world economy leads to an increase in rubber demand, while a decline in the price of NR relative to SR induces a falling share of SR in total rubber consumption. Meanwhile, a weak currency exchange in the producing countries encourages an increase in exports of NR and its products (Khin et al. 2008). Furthermore, the price of NR is highly influenced by global rubber stock controllers. The rubber market in Thailand is controlled by Singaporean, Malaysian, and to a certain extent Thai investors, leaving the rubber growers themselves no role in price setting. As mere suppliers of the rubber produce, Thai farmers are offered a lower per unit price for RSS grade and fresh latex, on average around Thai Bhat 35.00 per kilogram. In contrast, market prices of rubber have fluctuated around 50-100 baht per kilo in 2007 (Kaiyoorawong 2008).

The growing debate on how to enhance the green economy has recognized that the carrying capacity of ecosystems is vulnerable and that, in the face of climate change, there is an urgent need to reduce global carbon emissions. The NR industry is fortunate from this point of view in that it not only provides eco-friendly products; it is also capable of reabsorbing the carbon dioxide from the atmosphere. The natural rubber industry is based upon minimal environmental disturbance, far less than that required to produce typical food crops (Jones 1997). Hence, the findings of this paper about the nature of fluctuating rubber price in the CRMH may have implications for both economic and environmental policy. If implemented, they could lead to benefits of small holders and price stabilization mechanism in the Songkhla and its peripherals as well as NR and NR based products export market.

5 Data Sources and Methods

5.1 Source and Nature of the Data

The data for this research were gathered from various sources. The daily bid prices of NR latex of type RSS3 for the period June 2004 to February 2011 were obtained from the webpage of the Songkhla Provincial Agriculture Office[4]. World rubber consumption and production and the value of NR exports were downloaded from the webpage of Thai Rubber Association[5] and from the webpage on Indian NR[6]. The nature of the series/variables are plotted in the Figure 2 and 3.

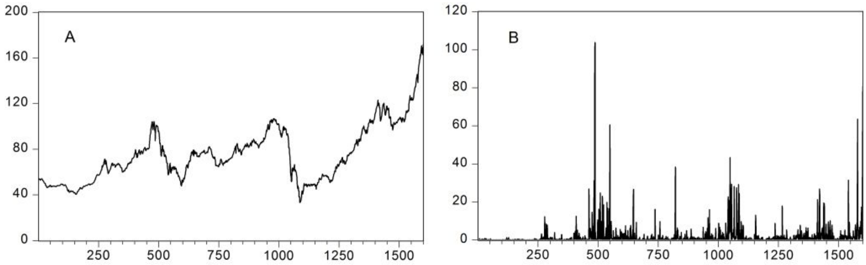


Fig. 2 A. Daily prices of RSS3 at CRHM (Bhat/Kg) and B. volatility of price

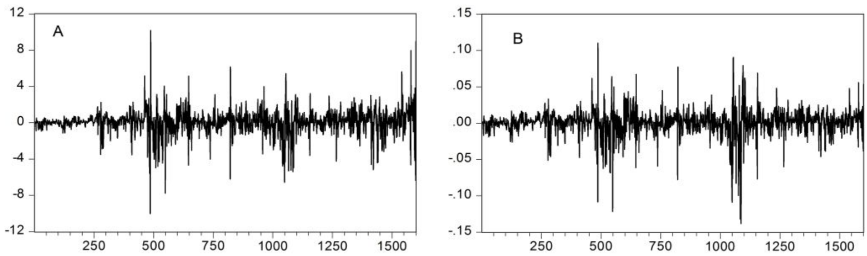


Fig. 3 First difference (A) and log difference of daily prices of RSS3 (B)

[4] `http://songkhla.doae.go.th/`
[5] `http://www.thainr.com/en`
[6] `http://rubber.wordpress.com/`

5.2 *Theoretical Setup and Models*

5.2.1 Unit Root Test

A standard time series model assumes linearity and symmetric adjustments and, if adjustment is approximately symmetric, the Dickey-Fuller test is more powerful than any other test (Enders and Granger 1998). A study variable series y_t (RSS3 in our case), is said to be stationary if the mean, variance and covariance of the series remain constant over time (Lim et al. 2009). If the y_t is correlated at higher order lags, the assumption of white noise disturbances, ε_t, is violated and the Augmented Dickey-Fuller (DF) test allow us to perform a parametric correction for higher-order serially correlated error processes. The formulation of an augmented DF test of dependent variable with deterministic trend is presented in equation (1) where Δy_t is the first difference of series y_t; 'p' is the lag-length terms for y_t, ε_t is the error term and α, δ, ψ and θ are the parameters. The lags Δy_t of capture any dynamic structure present in the dependent variables in order to insure that ε_t is not autocorrelated (Brooks 2008; Maddala 1992). Likewise, the Phillips-Perron method for unit root test estimates the non-augmented DF test, and modifies the t-ratio of the α coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic (QMS 2007)

$$\Delta y_t = \alpha + \delta trend + \psi y_{t-1} + \sum_{i=1}^{p} \theta_i \Delta y_{t-i} + \varepsilon_t \tag{1}$$

The classic methods of Dickey and Fuller (1979, 1981) and Phillips and Perron (1988) are suffered from low power and size distortions (Divino and McAleer 2010). However, such limitations are overcome by adopting other modified tests methods as suggested by Perron and Ng (1996), Elliott et al. (1996), and Ng and Perron (2001), (cited in Divino and McAleer 2009). These suggested modified unit root tests methods are also subject to low power and size distortions under the short run persistence implied by GARCH component. However, Divino and McAleer (2009) argued that such size distortions might be even greater for the traditional Dickey Fuller test, despite the sensitivity of the modified tests to the degree of volatility in the GARCH process. Hence, we have adopted the modified Augmented Dickey-Fuller generalized least square (MADFGLS) test and the MPP^{GLS} test, which both use generalized least square (GLS) de-trended data and the MAIC in order to choose the truncation lag MPP^{GLS}.The outcomes of unit root test are presented in Table 4 and the result reveals that the null hypothesis of a unit root is rejected for the transformed variables. The series RSS3 and logarithm of RSS3 may not be stationary. However, the unit root test clearly depicts that the first difference and log difference variables are stationary. These empirical results allow us to use the univariate conditional mean and conditional volatility models to estimate the bid price behavior of RSS3 at CRMH.

Table 4 Unit root test statistics for daily price of RSS3

Variables	$MADF^{GLS}$ statistics		MPP^{GLS} statistics	
	Z{1}	Z {1,t}	Z{1}	Z {1,t}
RSS3 (Y1)	1.22	-1.11	3.18	-5.68
dY1	-9.68***	-4.72***	-151.96***	-18.21**
log Y1	-0.078	-2.03	-0.42	-10.63
dlogY1	-9.59***	-7.52***	-138.08***	-59.52***
Critical values				
At 1% level	-2.57	-3.48	-13.8	-23.8
At 5% level	-1.94	-2.89	-8.1	17.3

These results are obtained on lag length 10.Values with (***) denotes null hypothesis of a unit root is rejected at the 1% level of significance.

5.2.2 Conditional Mean and Conditional Volatility Models

Recently, researchers have turned their attention to the risk associated with economic variables. For instance, international tourist arrivals, prices of agricultural products, stock and exchange rates are carefully monitored. Conditional volatility models then adopted to predict the risk of returns for these industries and to capture symmetric and asymmetric effects using daily, weekly and monthly data for example (Divino and McAleer 2009; Huang B.-W. et al. 2009; Lim and McAleer 2000; Shareef and McAleer 2007; Yang et al. 2010). Price volatility of RSS3 follows a similar pattern to that of financial volatility. Recently, theoretical developments and their results for univariate and multivariate time series models with conditional volatility errors and a wide range of univariate and multivariate, conditional and stochastic models for volatility have been extensively reviewed (Li et al. 2002; McAleer 2005). The residual series of such selected models should follow the white noise process. The Akaike and Schwarz information criteria can be practiced to determine the lag length, although it is very common to impose GARCH (1, 1) specification in advance (Coshall, 2009; Huang B.-W. et al., 2009). The general form of the GARCH (p, q) formulation is presented in equation (2) and (3). Likewise, the univariate stationary form of AR (1) - GARCH (1, 1) model for bid price of RSS3 and transformed variables (i.e. Δy_t; $\log y_t$; and $\Delta \log y_t$ as appropriate) at CRMH, is presented in equation (4)

$$y_t = \psi_1 + \psi_2 y_{t-i} + \varepsilon_t;\ \text{where}\ \ |\psi_2| < 1 \tag{2}$$

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-1}^2 + \sum_{j-1}^{q} \beta_j h_{t-j} \tag{3}$$

$$y_t = \psi_1 + \psi_2 y_{t-1} + \varepsilon_t,\ \text{where}\ \ |\psi_2| < 1 \tag{4}$$

Where y_t denotes the daily prices of RSS3 for $t = 1, \ldots, n$ and ε_t is an error process. The shock or movement in daily bid price of RSS3 is denoted by,

$$\varepsilon_t = v_t\sqrt{h_t}; \quad v_t \approx iid(0,1) \tag{5}$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{6}$$

Where, $\sigma_v^2 = 1$ and v_t is a white-noise process and the conditional and unconditional means of ε_t are equal to zero. The conditional variance of the error process is h_t. The AR (1) process in equation (4), can easily be extended to univariate or multivariate ARMA (p, q) processes (Ling and McAleer 2003) and it is expressed in equation (8), where $\omega > 0$, $\alpha \geq 0$, and $\beta \geq 0$; are the sufficient conditions to ensure $h_t > 0$; and α and β are the ARCH and GARCH terms, respectively. The ARCH effect designates the short run persistence of shocks while the GARCH term represents the long run persistence of the shock, that is (α + β). The larger the value of both α and β , the more plausible it becomes to increase the conditional volatility, but in different ways and a higher value of α indicates a more distinct shock in the subsequent period (Enders 2004; Ling and McAleer 2003). Since the GARCH process is a function of the unconditional shock, the quasi-maximum likelihood (QML) estimators for GARCH (p, q) are consistent if the second moment of ε_t is finite (Ling and McAleer 2003). Parameters in the conditional variance equation are generally estimated by the maximum likelihood method under the conditional normality assumption.Moreover, even if the conditional normality assumption does not hold, parameters can be estimated by applying the QML method.

The sufficient and necessary condition for the presence of the second moment of ε_t for a GARCH (1, 1) process is $\alpha + \beta < 1$ (Huang B.-W. et al., 2009). Under normality, the necessary and sufficient condition for the existence of the fourth moment is $(\alpha + \beta)^2 + 2\alpha^2 < 1$ (Divino and McAleer, 2009). The QML estimators are asymptotically normal locally and globally if the fourth and sixth moments of ε_t are finite for GARCH (p, q) (Ling and Li 1997; Ling and McAleer 2002; Ling and McAleer 2003). The previous literature suggests that the non-negativity condition of GARCH (p, q) model might be violated by the estimated model, and assumed that the positive and negative socks are same in the symmetric GARCH process, under the provision of conditional variance, h_t and it can able to capture thick tailed and volatility clustering (Enders 2004).Additionally, Glosten et al. (1993) proposed GJR (1, 1) model suited to capture the leverage effect and asymmetry behaviour which is defined as:

$$h_t = \omega + [\alpha + \gamma I(v_{t-1})]\varepsilon_{t-1}^2 + \beta \log h_{t-1} \tag{7}$$

where $\omega > 0$, $\alpha + \gamma \geq 0, \beta \geq 0$ and $\mathrm{I}(v_t) < 0$, is an indicator variable and defined as, $\mathrm{I}(v_t) = 1$, if $\varepsilon_t < 0$, otherwise, 0, if, $\varepsilon_t \geq 0$ and asymmetry of the series is captured by the coefficient, γ.The regularity condition for GJR (1, 1) is $\alpha + \beta + 1/2\gamma < 1$ and ε_t and v_t have the same sign. It is expected that the value of coefficient, $\gamma \geq 0$, particularly when handling financial data because negative shocks increase the risk (Huang et al. 2009), this fact also holds for prices of NR (for growers).

$$\log h_t = \omega + \alpha + |v_{t-1}| + \gamma v_{t-1} + \beta \log h_{t-1}; \quad \text{where} \quad |\beta| < 1 \tag{8}$$

The exponential GARCH model or EGARCH (equation 8) is suggested to capture the asymmetric effect in the data series; this model allows no restriction on parameters (α, β and γ) required to ensure $h_t > 0$, if, $|\beta| < 1$ is a sufficient condition for consistency of asymptotically normal values of Quasi-maximum likelihood estimators (QMLE). It was noted that GARCH and GJR models are dependent upon lagged unconditional shocks, while EGARCH depends upon lagged conditional shocks to the standardized residuals . Extensions of several of these results (ν_{t-1}) for asymmetric conditional volatility models are given in ref. (McAleer et al. 2007).

6 Empirical Results

The series daily price of the rubber latex type RSS3 at CRMH and its logarithm may not be stationary. However, the null hypothesis of unit root is statistically rejected and clearly depicts that the first difference and log difference variables are stationary (Table 4). Based on these empirical results of unit root test, the univariate conditional mean and conditional volatility models, namely, GARCH (1, 1), GARCH-GJR and EGARCH models are adopted by introducing AR (1) term in mean equation to estimate the price behavior of RSS3 at CRMH. The estimated parameters and their respective standard errors for transformed variables of price of RSS3 are presented in Table 5. The estimates of lagged dependent variables in the equations (4)–(9) are supported by the empirical findings and most of the estimated coefficients in mean and variance equation are statistically significant.

The estimates of the AR(1)-GARCH(1, 1) model for the logarithm of price of RSS3 are positive and statistically significant at the 1% level of significance. The estimated coefficients symbolically, $\alpha = 0.170$ (short run shock) and $\alpha+\beta = 0.992$ (long run shock) reveals that there is high long run persistence of shock or price volatility at CRMH for latex type RSS3. Moreover, when one takes account of asymmetric behavior by considering similar magnitudes of the positive or negative shock on the bid price of RSS3, the asymmetry coefficient, γ, for the GARCH-GJR model is found to be positive. The contribution of the shock to both short run $(\alpha + \frac{1}{2}\gamma)$ and long run persistence ($\alpha + \beta$ + $\frac{1}{2}$ γ) are 0.166 and 0.988, respectively. The positive value of the γ coefficient implies that decreases in the bid price of rubber latex type RSS3 increases its price volatility. The EGARCH model, based on the standardized residuals, yields statistically significant estimated coefficients, except that for the constant in the mean equation. The second moment condition i.e. $(\alpha + \beta < 1)$ for both GARCH (1, 1) and GARCH-GJR models are satisfied with value of 0.992 and 0.988 respectively. The regularity condition,$(|\beta| < 1)$, for EGARCH is satisfied with an estimated coefficient of 0.956. Therefore, the estimated QMLE are statistically consistent and asymptotically normal.

Uniformly, the estimated parameters for the GARCH, GARCH-GJR and EGARCH models for logarithm differences in the daily bid price of rubber latex type RSS3 are also statistically significant (Table 5). The estimates of GARCH (1, 1) and GARCH-GJR reveal that the short run persistence of shock or risk (α or α + γ) are 0.201 and 0.203 correspondingly, while the long run persistent of the

Table 5 Estimated parameters and their respective standard errors from conditional mean and volatility models

Parameters	Dependent Variable: Log (RSS3)			Dependent Variable: D Log (RSS3)		
	GARCH	GJR	EGARCH	GARCH	GJR	EGARCH
Constant	3.161*	3.263	13.418	0.0009**	0.0009**	0.001*
	(1.127)	(0.998)	(51.06)	(0.0004)	(0.0004)	(0.0004)
AR(1)	1.000*	1.000*	0.999*	0.280*	0.279*	0.301*
	(0.000)	(0.000)	(0.0008)	(0.031)	(0.028)	(0.026)
ω	0.000*	0.000*	-0.652*	0.000*	0.000*	-0.625
	(0.000)	(0.000)	(0.046)	(0.000)	(0.000)	(0.044)
GARCH/GJR α	0.170*	0.156*	-	0.201*	-	0.191*
	(0.013)	(0.016)		(0.015)		(0.071)
GJR γ	-	0.020	-	-	0.012	-
		(0.018)			(0.022)	
GARCH/GJR β	0.822*	0.822*	-	0.798*	0.801*	
	(0.012)	(0.012)		(0.012)	(0.031)	
EGARCH α	-	-	0.337*	-	-	0.363*
			(0.015)			(0.017)
EGARCH γ	-	-	-0.021**	-	-	-0.008
			(0.009)			(0.010)
EGARCH β	-	-	0.950*	-	-	0.956*
			(0.004)			(0.004)
Diagnostic						
Second moment	0.992	0.988	-	0.999	0.998	-
LM(1):nR^2 [Prob.]	2.11	1.97	2.35	1.85	1.64	2.27
	[0.14]	[0.15]	[0.12]	[0.17]	[0.20]	[0.13]
Jarque-Bera	2263.60	2299.64	3197.76	2812.33	2892.18	3416.76
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]

Notes: Values with (*, ** and ***) denotes the coefficients are significant at 1, 5 and 10 percent level of significance respectively. The values in parentheses are standard errors. The numbers in brackets are p-values for the Lagrange multiplier (LM) test and Jarque-Bera diagnostic tests for ARCH (1) residuals, respectively.

risk i.e. $(\alpha+\beta)$ and $(\alpha+\beta+\frac{1}{2}\gamma)$ are 0.798 and 0.813, respectively. Moreover, the GARCH-GJR (1, 1) estimates express the asymmetry behavior of the positive or negative shocks of the bid price of the rubber latex type RSS3 under the assumption of similar patterns or magnitudes of the positive or negative shocks. The positive estimated coefficients for GJR (1, 1) reveal that decreases in the bid price of RSS3 increase price volatility. However, the estimated parameter i.e. γ is not statistically significant. The second moment condition for regularity and asymptotic normality is also satisfied for both GARCH and GARCH-GJR models, i.e. $\alpha+\beta<1$ and α + β+ $\frac{1}{2}\gamma<1$, and the estimated coefficients are positive. Therefore, the estimated QMLE are asymptotically normal for both models.

The EGARCH estimates are treated as the logarithm of the volatility; and the coefficient α represents the magnitude (size effect). The estimated coefficient α for the EGARCH model is positive and statistically significant at the 1% level of significance. The estimated coefficient of the lagged dependent variable $|\beta|$ is 0.956 and statistically significant for a logarithm differenced variable. This suggests that the statistical properties of the QMLE for EGARCH(1,1) will be consistent and asymptotically normal.

7 Conclusion

The demand for natural rubber is growing rapidly around the world along with the economic advancement and improvements in standard of living. Synthetic rubber is a purely petroleum product and highly subject to price volatility. The price of natural latex is quite volatile as in SR and price volatility can cause havoc with profitability and workers income. Uncertainties in the prices of NR adversely affect financing in NR based industries and the livelihoods of rubber growers. Hence, the volatility in the price of NR products plays a decisive role in both the export market and in the livelihood of millions of small rubber growing households and tappers. Since 1991, Thailand is the one of largest NR growing area in the world, accounting for more than 31% of total production and 34% of exports of the world market in 2010. The economic contribution of the NR industry in Thailand in terms of export receipts has increased significantly, Thai Bhat 74.61 billion in 2002 to reach Thai Bhat 249.26 billion in 2010 with 16.27 % of average annual growth. The daily bid price of rubber latex type RSS3 at CRMH, Songkhla is modeled by adopting symmetric and asymmetric conditional volatility models. Most of the coefficients are statistically significant. The GARCH (1, 1) model for logarithm of latex price showed there is high long run persistence of shock or price volatility $\alpha = 0.172$ and $((\alpha+\beta) = 0.992)$. The estimated coefficient, γ for GARCH-GJR is positive which indicates decrease in bid price of rubber latex type RSS3 increases the price volatility. But, estimated parameter, γ is not statistically significant. The short and long run persistence of shock are, 0.166 and 0.988 respectively. Likewise, the GARCH (1, 1) estimates for logarithm difference variable are, $\beta = 0.201$ and $\alpha+\beta = 0.999$. These empirical results indicate that there is long run persistence of price volatility. Coefficient $|\beta|$ for EGARCH model for both variables (log Y and Δ log Y) is less than unity, i.e. $\beta = 0.950$ for logarithm of price and $\beta = 0.956$ for logarithm difference variable.

The second moment condition for GARCH (1, 1) and GJR (1, 1); and $|\beta| < 1$ in the case of EGARCH (1, 1) are satisfied for all cases. Therefore, the estimated QMLE are asymptotically normal. Hence, the volatility can be inferred as risk associated with the bid price of latex rubber type RSS3 in CRMH, Songkhla. This cannot be overlooked to ensure the benefits for smallholders and harness greater benefits from natural rubber export through the imposition of effective policy measures. Stainability of the natural rubber industry is crucial for the provincial economy of Songkhla to overcome rural poverty by generating jobs at the local level and foreign receipts earnings from the natural rubber trade. Thus, volatile rubber price in the CRMH may have implications for both economic and environmental policies. So that future policies could address the uncertainties behind the household level income of most of the small holding rubber grower due to price volatility in Songkhla province and towards the rubber price stabilization in the local market as well as in the international market to harness the huge export potential of natural rubber and rubber articles. This baseline study has modeled volatility for the recent historical period (2004-2011). Future research on volatility, as well as the extension of modeling for cointegration of price considering spatial destination, could further explore the long-run relationships among the underlying variables and refine the policy implications of this paper.

Acknowledgement. We are grateful to acknowledge Thailand Research Fund and Chaing Mai University, Thailand for providing financial and academic support for this study. We are also thankful to all the persons and Officials who had provided information during field study and anonymous reviewer for providing valuable comments and suggestions.

References

1. Allen, P.W.: Non-wood products-rubber trees. Encyclopedia of Forest Sciences, 627–633 (2004)
2. Brooks, C.: Introductory econometrics for finance. Cambridge University Press, New York (2008)
3. Coshall, J.T.: Combining volatility and smoothing forecasts of UK demand for international tourism. Tourism Management 30(4), 495–511 (2009)
4. Dean, W.: Brazil and the struggle for rubber: A study in environmental history. Cambridge University Press, UK (1987)
5. Divino, J.A., McAleer, M.: Modeling sustainable international tourism demand to the Brazilian Amazon. Environmental Modeling and Software 24(12), 1411–1419 (2009)
6. Divino, J.A., McAleer, M.: Modeling and forecasting daily international mass tourism to Peru. Tourism Management 31(6), 846–854 (2010)
7. Enders, W.: Applied econometric time series. Jhon Wiley and Son Inc., U.K. (2004)
8. Enders, W., Granger, C.W.: Unit-root test and asymmetric adjustment with an example using the term structure of interest rates. Journal of Business and Economic Statistics 16(3), 304–311 (1998)
9. Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50(4), 987–1007 (1982)
10. FAO, Recent developments in natural rubber prices. Consultation on agricultural commodity price problems. Food and Agriculture Organization (FAO), Roam (2002)
11. Gov.Thai PRD: Hat Yai of Songkhla Province designated as rubber city. The Government Public Relation Department (Gov/Thai PRD) (2004), `http://thailand.prd.go.th/view` (accessed September 5, 2010)
12. Grilli, E.: The world rubber economy. Johns Hopkins University Press, Baltimore (1980)
13. Huang, B.-W., Chen, M.-G., Chang, C.-L., McAleer, M.: Modeling risk in agriculture finance: Application to the poultry industry in Taiwan. Mathematics and Computers in Simulation 79(5), 1472–1487 (2009)
14. IRSG, News Items: Latest RSB, world rubber industry outlook, International Rubber Study Group (IRSG) (2010), `http://www.rubberstudy.com` (accessed August 15, 2010)
15. Jones, K.P.: Rubber and the environment: Opportunities and constraints for the internalization of environmental costs and benefits into the price of rubber (1997)
16. International Rubber Study Group and the Secretariat of the United Nations Conference on Trade and Development (UNCTAD/IRSG) Workshop: 8–19
17. Kaiyoorawong, S.: Thailand: Rubber prices fluctuate, how can farmers benefit? World Rain Forest Movement (WRM) Bulletin (37), 16–17 (2008)

18. Khin, A., Chong, E.C., Shamsudin, M.N., Mohamed, Z.A.: Natural rubber price forecasting in the world market. In: Agricultural Sustainability Through Participative Global Extension (AGREX 2008), Kuala Lumpur, University of Putra Malaysia (2008)
19. Kittipol, L.: Natural rubber production and outlook of Thailand. Power Point Presentation on Natural Rubber Production and Outlook of Thailand, May 28-29. The Thai Rubber Association, Songhai (2008)
20. Li, W.K., Ling, S., McAleer, M.: Recent theoretical results for time series models with GARCH errors. Journal of Economic Surveys 16(3), 245–269 (2002)
21. Lim, C., McAleer, M., Min, J.C.: ARMAX modeling of international tourism demand. Mathematics and Computers in Simulation 79(9), 2879–2888 (2009)
22. Lim, C., McAleer, M.: A seasonal analysis of Asian tourist arrivals to Australia. Applied Economics 32(4), 499–509 (2000); Ling, S., McAleer, M.: Necessary and sufficient moments condition for the GARCH (r,s) and asymmetric power GARCH (r,s) models. Econometric Theory 18(3), 722–729 (2002)
23. Ling, S., McAleer, M.: Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19(2), 280–310 (2003)
24. Ling, S., Li, W.K.: On fractionally integrated autoregressive moving-average models with conditional heteroskedasticity. Journal of the American Statistical Association 9(439), 1184–1194 (1997)
25. Maddala, G.: Introduction to Econometrics, 2nd edn. MacMillan Publishing Company, New York (1992)
26. McAleer, M., Chan, F., Marinova, D.: An econometric analysis of asymmetric volatility: Theory and application to patents. Journal of Econometrics 139(2), 259–284 (2007)
27. McAleer, M.: Automated inference and learning in modeling financial volatility. Econometric Theory 21(1), 232–261 (2005)
28. QMS, User's Guide Eviews 6 (vol. II). Quantitative Micro Software (QMS), LLC, USA (2007)
29. Saidur, R., Mekhilef, S.: Energy use, energy savings and emission analysis in the Malaysian rubber producing industries. Applied Energy 87(8), 2746–2758 (2010)
30. Shareef, R., McAleer, M.: Modeling the uncertainty in monthly international tourist arrivals to the Maldives. Tourism Management 28(1), 23–45 (2007)
31. Somboonsuke, B., Shivakoti, G.P.: Small holders of rubber-based farming systems in Songkhla Province Thailand: Problems and potential solutions. Kasetsart Journal (Soc. Sci.) 22, 79–97 (2001)
32. Somboonsuke, B., Cherdchom, P.: Socio-economic adjustment of smallholding rubber-based farming system: Case study in Southern Region of Thailand. Kasetsart Journal (Soc. Sci.) 21, 158–177 (2000)
33. Stifel, L.D.: The Growth of the Rubber Economy of Southern Thailand. JOSTER: Journal of Southeast Asian Studies 4(1), 107–132 (1973)
34. Stubbs, R.: Malaysia's rubber smallholding industry: Crisis and the search for stability. Pacific Affair 56(1), 84–105 (1983)
35. TERRA, Plantations are not forest, Towards Ecological Recovery and Regional Alliance (TERRA) (Editorial). Watershed: Commercial Tree Plantations in the Mekong Region 9,1–3 (2004)
36. TRA, Para rubber development strategy 2009-2013. The Thai Rubber Association, TRA (2007), http://www.thainr.com/en (accessed September 15, 2010)

37. UNCTAD, Information on rubber: economic policies. Infocomm: Market information in the commodities area, United Nations Conference on Trade and Development (UNCTAD) (2007),
`http://www.unctad.org/infocomm/anglais/indexen.html`
(accessed January 10, 2011)
38. Viswanathan, P.: Emerging smallholder rubber farming systems in India and Thailand: A comparative economic analysis. Asian Journal of Agriculture and Development 5(2), 1–20 (2006)
39. Yang, C.-H., Lin, H.-L., Han, C.-C.: Analysis of international tourist arrivals in China: The role of world heritage sites. Tourism Management 31(6), 827–837 (2010)

Trade Liberalisation, Labour Productivity Growth and Skilled Labour Complement: Evidence from the Thai Manufacturing Sector

Piyapong Sangkaew and Kankesu Jayanthakumaran

Abstract. Trade liberalisation in Thailand raised two wider questions regarding the labour market–one with regards to the link with labour productivity and the other the link with skilled workers. This outcome provides a link between (1) trade liberalisation and labour productivity growth, and, (2) skilled employment and labour productivity growth. Trade liberalisation is also correlated with skilled employment. This type of evidence matches conventional explanations for the beneficial allocation of trade liberalisation and demanding skills training for potential future industrial growth.

1 Introduction

Trade liberalisation policy has been implemented by countries to stimulate economic and employment growth. However, there have been longstanding concerns about the possible job displacement effects of trade liberalisation and other measures introduced to lift productivity.

Beaudry and Collard [2], for example, explain that reducing controls on trade causes a drastic technological change, and makes human capital the factor for determining the growth in labour productivity while also raising the demand for more skilled workers. Davis and Harrigan [5] explained that trade liberalisation and productivity have been found to be biased toward skilled workers. Thailand accelerated trade liberalisation in the early 1990s such that simple average tariff rates on industrial products decreased from 43.5 per cent in 1991 to 14.6 per cent in 1999 [15], and the next stage of reforms started in 1999 after a brief setback during the Asian

Piyapong Sangkaew · Kankesu Jayanthakumaran
School of Economics,
University of Wollongong,
NSW2522, Australia
e-mail: ps650@uowmail.edu.au
kankesu@uow.edu.au

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 203–213.
DOI: 10.1007/978-3-642-35443-4_14

crisis in 1997/98. This paper intends to study the link between trade liberalisation, labour productivity, and skilled employment. The next section shows trade liberalisation and labour productivity in Thailand, the third section shows the method and data used in this study, the fourth section shows the empirical results, and the fifth section provides the conclusion.

2 Trade Liberalization and Labour Productivity

The first round of tariff restructuring began in the early 1990s and was completed in 1997. Subsequently, the structure of Thailand's tariff was reduced from 39 tariff rate categories to only six in this period. It has restructured customs tariff on nine product categories covering a total of 2,990 items, or 39.52 per cent of all customs tariff items. The second round of tariff restructuring was in 1999 and was implemented immediately after the Asian crisis (Appendix 1). Tariffs on capital goods, raw materials, and other products, including more than 630 items, were either reduced or exempted on a permanent basis, for example (i) the 10 per cent import duty surcharge was removed, (ii) tariffs on machinery and mechanical appliances and parts were reduced, and (iii) tariffs on electrical machinery and equipment parts were reduced from 5 and 20 per cent to 3 per cent for 326 items.

A visual inspection of Figure 1 shows that the labour productivity of the manufacturing sector has, in general, been increasing since 2001, although the fall in 2009 and 2011 may be associated with a fall in output due to the global crisis. Labour productivity in 1999 was almost two times higher than in 1991 [12]. Phan [15] found that trade liberalisation had increased both the labour and total factor productivity growth.

However, gains in labour productivity (higher output per worker) resulting from labour saving technologies may lead to job destruction ([3], [7], [11]) because competitive pressures can drive investment, innovation, skills upgrading, and other factors in the overall development process. Even higher productivity spurs economic growth and expands employment overall, although labour saving technological changes and the relative growth and decline of specific sectors results in job losses in some places and some industries, for actual workers, enterprises, and communities.

Upon careful examination of Figure 2, even though the percentage of employment between 1997 and 2000 had increased, employment in the manufacturing sector gradually decreased from 15.37 per cent to 14.06 per cent respectively, between 2000 and 2007. This trend had changed from the percentage of manufacturing GDP to overall GDP, which had increased over the same period from 1991 to 2010. The percentage of employment in the Thai manufacturing sector had not kept up with an increase in the share of manufacturing GDP. Thus, employment-productivity trade-off problems should be taken into account mainly because the policy of stimulating trade and productivity might decrease employment in this sector.

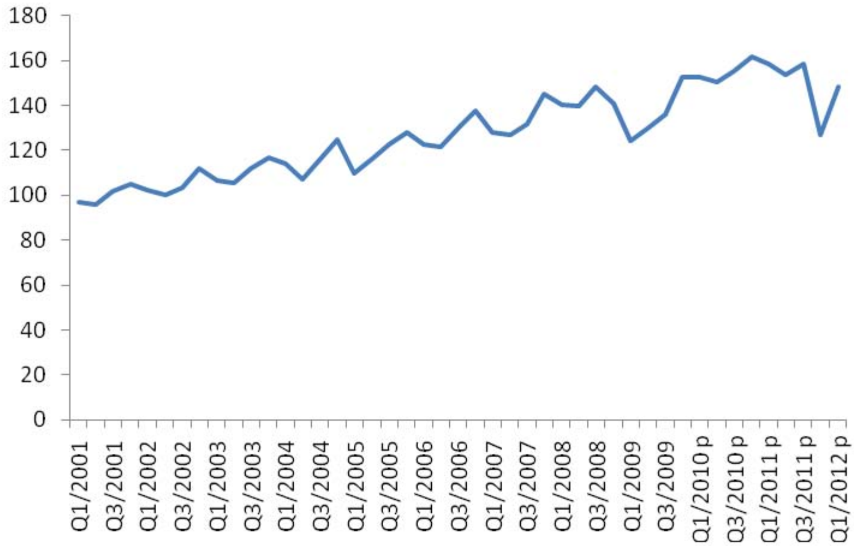


Fig. 1 Labour Productivity Per Employed Person between 2001 and 2010

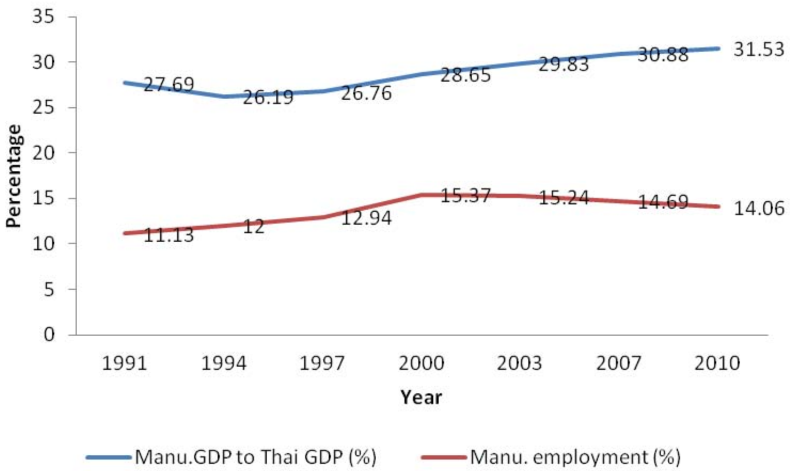


Fig. 2 Labour Productivity Per Employed Person between 2001 and 2010

Table 1 shows the correlation of variables; skilled-total employment, labour productivity (LP) and tariffs in the Thai manufacturing sector. There are negative correlations between (i) tariff and labour productivity, and (ii) tariff and productivity, which are - 0.131 and -0.057 respectively. This value implies that a decreasing tariff

Table 1 Correlation between Manufacturing Tariffs and Labour Productivity between 1991 and 2007

	Skilled-Total Employment Ratio	LP	Tariff
Skilled-Total Employment Ratio	1.000		
Real LP*	0.178	1.000	
Tariff**	-0.057	-0.131	1.000

Source: Calculated by the author from the Industrial Surveys [20]. *: The value used in a natural log form. ** Tariff data is provided in Appendix 1.

correlates to an increase in the LP and skilled-total employment ratio in the manufacturing sector. This should basically support the idea that trade liberalisation increases manufacturing labour productivity and changes the structure of employment by raising the demand for skilled labour in the manufacturing sector.

3 Method and Data

Labour productivity *(lp)* can be defined as value-added per worker. The dependent variable is the growth of labour productivity $(\dot{lp})$ during the period 1991, 1994, 1997, 2000, 2003 and 2007. Independent variables have been chosen to reflect trade policy, employment, and structural and technological changes. The labour productivity growth model can be constructed as follow:[1]

$$\begin{aligned} \dot{lp} = {} & \alpha_0 + \alpha_1 \dot{k}_{it} + \alpha_2(\dot{skill}_{it}) + \alpha_3(\dot{emp}_{it}) + \alpha_4(x_{it}) \\ & + \alpha_5(i\_im_{it}) + \alpha_6(FDI_{it}) + \alpha_7(D_{it}) + \varepsilon_{it} \end{aligned} \tag{1}$$

Three trade policy variables–tariffs, intermediate tariffs, and exports–have been used to capture the effect of trade liberalisation on $\dot{lp}$. The reductions in tariffs reflects the response by local firms, and local firms tend to match a new set of relative prices that are closer to international prices and which stimulate resources in line with comparative advantages. The reduction in tariffs, therefore, are expected to have a negative effect on $\dot{lp}$. Firms target greater technical change in an open trading environment in order to achieve sustainable long term rates of growth. As a result, intermediate-input import per worker is expected to have a positive effect on labour productivity growth.

Exports per worker are normally used as a proxy variable of trade liberalisation in empirical studies to explain productivity. Exports are found to stimulate firms or industrial labour productivity. Jonsson and Subrmanian [9] and Sjoholm [19] explain that export firms tend towards new technology and produce higher quality products. Moreover, exporting firms have a higher price margin than non-exporting

[1] The expected signs of $\dot{k}_{it}$, $\dot{skill}_{it}$, $\dot{emp}_{it}$, x_{it}, i_imit, FDI_{it} are $+,+,-,+,+,+,+/-$, respectively.

firms, which raises an export firms' labour productivity. Therefore, export growth is expected to have a positive effect on industrial labour productivities in this study [13]. If there are expected signs and significant associations between one or more of the trade variables and $\dot{lp}$ then there will be support for a positive impact of liberalisation on the performance of labour.

Two employment variables, share of skilled employment and overall employment, have been used to reflect the effect of the labour market on $\dot{lp}$. Skilled employment growth is expected to have a positive effect on $\dot{lp}$. Generally, skilled workers refers to workers who spend more years in school, so they tend to have high human capital accumulation, and as a result they are more productive than unskilled labour. An increase in the number of skilled workers will raise firms' productivity and efficiency [21]. The share of manufacturer employment to the total manufacturing employment is expected to be negatively associated with labour productivity growth due to the effect of "the law of diminishing marginal return". An increase in the number of workers will increase total output diminishingly [22], and as such, increases in a number of workers will cause a decrease in labour productivity.

Two variables, output growth and FDI have been chosen to reflect the structural changes. Output growth is expected to stimulate growth in labour productivity because industry will benefit from an economy of scale [23]. The share FDI per worker reflects an increase of FDI over time. FDI can stimulate growth in labour productivity in many ways, such as (i) providing better knowledge, (ii) giving firms more opportunity to export. One would expect a positive association between output growth and $\dot{lp}$, and FDI per worker and $\dot{lp}$.

The capital growth variable represents technological change and is expected to stimulate growth in labour productivity. Capital growth per worker is expected to have a positive effect on the growth of industrial labour productivity.

The data used in this study is from (i) Manufacturing Industrial Survey conducted by the National Statistical Office (NSO), Thailand, and (ii) World Integrated Trade Solution (WITS) which is an organisation under the World Bank. The scope of the manufacturing surveys are firms primarily engaged in manufacturing industries which are classified according to the International Standard Industrial Classification (ISIC), have 10 or more persons engaged in the business, and cover the whole country. The criticism of Thai manufacturing data by Ramstetter [17], and Ramstetter and Sjoholm [18], have been taken into account such that the following observations have been used; for the years 1991, 1994, 1997, 2000, 2003 and 2007, the observations are 970, 969, 2,558, 2,285 3,765, 18,620 respectively[2].

[2] Overall there are 29,167 firms left that are useful for this study (22.86 per cent); (i) 10, 268 firms which did not report the number of workers, (ii) one firm which did not report working hours and days, (iii) 23, 129 firms which did not report wage bills, (iv) 18,831 firms which did not report their income, (vi) two firms which did not have ISIC code, (vii) 1,432 firms which did not report their fixed assets, (viii) 13, 266 firms which did not report their fixed assets (machines), (ix) 5488 firms which have a negative value added, (x) 26, 784 firms which have no skilled workers, and (xi) 223 firms which did not report the number of unskilled workers.

In this study the Thai manufacturing labour productivities (LP) were calculated from Thai industrial surveys in the selected years 1991, 1994, 1997, 2000, 2003, and 2007. After the labour productivity growth $(\dot{lp})$ for each industry and for all the years have been obtained, they become a manufacturing panel data. Overall there were 138 observations of Thai manufacturing labour productivities (23 industries in 6 years). Because the calculated growth in the rate of labour productivity caused this study to lose a year of data (23 observations), the actual observations are 115.

There are inconsistencies in the availability of data on intermediate-input import, export, and FDI in Thai industrial surveys. To overcome this problem, this study will incorporate the effect of intermediate-input import on $\dot{lp}$ between 1994-2007, while the effect of export and FDI on $\dot{lp}$ will cover between 1997 to 2007.

4 Empirical Results

To detect multi–collinearity problems, the correlation metrics between independent variables have been checked (Appendix 1). The correlation is low in all cases except for (i) intermediate input import (i_im), and export (x) and (ii) capital per worker growth $(\dot{k})$ and per cent change in tariff $(\dot{tariff})$, which are 0.6542 and -0.3179, respectively. As the high correlation among the variables could lead to multi-collinearity problems, this study applies the variance-inflating factor (*VIF*) to examine whether this would be a critical problem [6] . The calculated *VIF* is 6.20 which is less than ten per cent, which implies that even though there are multi-collinearity problems, they are acceptable [6].

Table 2 provides empirical results from this study. There are six models which begin with the simple ordinary least squared (OLS) shown in the first column. This is to compare the results and to examine the consistency of the coefficient of variables. Then the fixed effected (FE) and random effected (RE) are shown in columns 2 and 3, respectively [6]. After that, new variables will continually be added into the model to examine whether they are significant in emplaning labour productivity growth $(\dot{lp})$ in the Thai manufacturing sector. The best model will be selected based on econometric reasons.

The model 1 in the first column is the OLS pooling the data over the period 1991 to 2007 without industrial effects. The result shows that $\dot{y}$, $\dot{tariff}$, and $\dot{emp}$ significantly affect $\dot{lp}$ at one, ten and one per cent significance level, respectively. $\dot{y}$ has a positive effect on $\dot{lp}$ while $\dot{emp}$ and $\dot{tariff}$ have a negative effect on $\dot{lp}$. According to this study, $\dot{y}$ is found to play an important role in contributing to $\dot{lp}$ since a one per cent increases in $\dot{y}$ will increase $\dot{lp}$ around 0.5195 per cent.

A proxy variable of trade liberalisation $\dot{tariff}$ has a correct and negative sign and is significant at the 10 per cent level. A one per cent decrease in the tariff will increase $\dot{lp}$ in the Thai manufacturing sector around 0.2348 per cent. Therefore, it can be said that trade liberalisation increases the Thai manufacturing labour productivity.

Table 2 Regression Results for the Thai Manufacturing Sector

Variables	OLS (1)	FE (2)	RE (3)	RE2INPUT (4)	RE3EXPORT (5)	RE4FDI (6)
$\dot{k}$	0.0351	-0.07783	0.0351	-0.3622	0.1323	0.2844**
	-0.0779	-0.0958	-0.0779	-0.0778	-0.0946	-0.1262
$\dot{y}$	0.5195***	0.5062***	0.5195***	0.5142***	0.4898***	0.3436***
	-0.0418	-0.0462	-0.0418	-0.042	-0.0483	-0.0802
$\dot{tariff}$	-0.2348*	-0.3584**	-0.2349*	-0.2608**	-0.0528	0.0909
	-0.1318	-0.1512	-0.1318	-0.1338	-0.1725	-0.1891
$\dot{skill}$	0.1717	0.2421**	0.1717*	0.1850**	0.2222*	0.1924
	-0.1037	-0.1184	-0.1037	-0.1043	-0.1174	-0.1495
$\dot{emp}$	-0.1353***	-0.1129***	-0.1354***	-0.1333***	-0.0935**	-0.0918*
	-0.0322	-0.0356	-0.0322	-0.0322	-0.0392	-0.0498
i_im				-0.0463	-0.1066	0.0931
				-0.0426	-0.0734	-0.0879
x					-0.0012	-0.0374
					-0.0812	-0.1608
FDI						0.0489**
						-0.0179
constant	-0.0827	-0.1009	-0.0827	-0.5158	1.4517	-1.3337
	-0.0671	-0.0709	-0.0671	-0.5554	-0.8953	-1.0564
Ind.effect.	NO	YES	YES	YES	YES	YES
N	85	85	85	85	66	51
F-test	38.4971	30.4081	.	.	.	.
R-squared	0.709	0.7238	0.7149	0.7123	0.7879	0.8387
AIC	147.816	126.983	.	.	.	.

Source: Calculated by the author. Note: Dependent variable: Growth of labour productivity. $*, **, ***$; Estimate coefficient is significant at the 10, 5 and 1 percent level, respectively.

Employment ($\dot{emp}$) is found to have a negative effect on $\dot{lp}$ and a one per cent increase in $\dot{emp}$ will decrease $\dot{lp}$ by 0.1353 per cent. The negative effect of this variable on $\dot{lp}$ would be a diminishing return or an employment-productivity trade-off, where more employment would reduce the productivity of labour in manufacturing. If other variable constants are held, expanding employment will decrease the labour productivity growth in the Thai manufacturing sector.

Because manufacturers are in various businesses and use different technologies, this may cause them to have different labour productivity growth. This study applied the fixed (FE) and random effect (RE) techniques to control the industry effect, but if the constant value in the model is not systematically changed, the fixed effect is more suitable. However, if it is the random effect becomes more reliable [6]. To examine whether FE or RE is better, the Hausman test is applied and used to answer this question [6]. The result from the calculated Hausman test shows $prob > 0.171$ so a null hypothesis is accepted [8], and therefore the RE model provides a better explanation than the fixed effect model. For this reason the following models will be based on the random effect model to explain how trade liberalisation effects the labour productivity growth in the Thai manufacturing sector.

The RE model 3 shows that $\dot{y}$, $\dot{emp}$ and $\dot{tariff}$ are still significant and have almost the same coefficients as the OLS model. In model 4, i_im is added into the model, which means that all the explanatory variables remain almost the same. The new variable has an expected sign but it is not significant.

In model 5 the export variable is added into the model and yields a better result. Compared with model 2, r-squared in e model 5 is higher, increasing from 0.72 in model 2 to 0.78 in model 4. However, the new variable added into this model is not significant so in this model, only $\dot{y}$, $\dot{skill}$ and $\dot{emp}$ remain significant but have lower coefficients.

In model 6, FDI is added into the model. The result shows that FDI is positively significant in giving an explanation for in the Thai manufacturing sector. In this model the r-squared increases significantly from around 78 per cent in the model 5 to 83 per cent in model 6. This result shows that an increase in FDI of one per cent will increase the $\dot{lp}$ by around 0.0489 per cent. After putting FDI in the model, the coefficient of $\dot{k}$ becomes positive and significant, although $\dot{y}$ remains and $\dot{emp}$ has the same correct sign, it is a little bit smaller.

In the labour productivity growth model, it is important to consider that output may have endogenous problems [10]. According to this study, output might be affected by the world economy and the level of competition among producers [24]. Therefore, this study uses a dummy variable to capture the effect of the world economic down turn [4] which began in the year 2000. After testing for endogenous problems, the Hausman test does not reject the null hypothesis that variables are exogenous . The result is that a null hypothesis has not been rejected, which is consistent with the study of Quandt and Rosen [16] who mentioned that the exogenous[3] variable can produce results that are just as good as those generated by the more theoretical assumption of endogeneity.

5 Conclusion

Since the early 1990s, Thailand has consistently reduced tariffs and non-tariff barriers and lifted restrictions on FDI. A temporary setback occurred during the Asian crisis in 1997/98, but it was corrected afterwards. Statistics in general show that (1) the labour productivity of the manufacturing sector has been increasing over time, (2) there are widespread mismatches among the share of manufacturing growth and the share of manufacturing employment, indicating a possible expansion of skilled employment.

Trade liberalisation in Thailand raised two wider questions regarding the labour market, one with regard to the link with labour productivity and the other the link

[3] Ho: Variables are exogenous. The test of endogeneity shows that (i) Durbin (score) Chi2 (1) = 0.141861 (p= 0.7064) and (ii) Wu-Hausman F (1, 45) = 0.125521 (p = 0.7248) and the null hypothesis is not rejected.

with skilled workers. Regarding the first question, the results tend to show that there is some indication of a link between trade liberalisation and labour productivity growth. With respect to the second question, a positive correlation has been recorded in (1) trade liberalisation and skilled employment (Table 1), and (2) skilled employment and labour productivity growth. The growth in skilled employment is a contributor to labour productivity growth (Table 2) in the Thai manufacturing sector while the overall growth in manufacturing employment has a negative effect on labour productivity. This type of evidence matches with conventional explanations for the beneficial allocation of trade liberalisation and demanding skills training for potential future industrial growth. However, the models linking trade liberalisation, skilled employment, and labour productivity growth at a micro level would give more concrete results.

Appendix

Table 3 Appendix 1: Manufacturing Average Tariff Rate from 1991 to 2007

Industrial	1991	1994	1997	2000	2003	2007
15 Food and beverage	43.33	42.5	41.6	39.7	32.78	31.15
16 Tobacco products	NA	60	51.4	60	60	60
17 Textiles	60	53.33	30.1	20.2	24.35	20.32
18 Wearing apparel	75	65	41.3	46.9	36.79	27.4
19 Dressing of leather	100	70	28	19.4	21.1	18.33
20 Wood and products of wood and cork	15	40	17.9	16.1	15.2	9.3
21 Paper and paper products	10	10	18.4	15.2	12.66	5.06
22 Publishing, printing and reproduction of recorded media	NA	17.5	20	17.1	15.02	3.66
23 Coke refined petroleum products and nuclear fuel	30	27.5	NA	5.7	3.44	5.13
24 Chemicals and chemical products	30	57.5	15.9	10.1	6.49	4.15
25 Rubber and plastics products	30	55	33.7	25.3	23.53	8.6
26 Other non-metallic mineral products	20	20	24.4	17.2	14.75	9.98
27 Basic metals	30	18	10.8	9	9.66	2.68
28 Fabricated metal products, except machinery and equipment	30	32.5	22.9	18.7	NA	11.74
29 Machinery and equipment n.e.c.	41.67	47.33	10	8.5	NA	5.13
30 Office, accounting and computing machinery	30	30	NA	NA	7.72	2.15
31 Electrical machinery and apparatus n.e.c	40	40	16.5	13	NA	6.59
32 Radio, television and communication equipment and apparatus	45	45	NA	NA	NA	6.68
33 Medical, precision and optical instruments, watches and clocks	40	35	NA	NA	NA	4.65
34 Motor vehicles, trailers and semi-trailers	21.67	31.67	NA	NA	43.44	30.24
35 Other transport equipment	32.5	32.5	26.3	25.6	16.45	13.25
36 Furniture	70	60	40	20	NA	15.17
37 Recycling	NA	NA	NA	NA	NA	NA

Source: World Integrated Trade Solution (WITS, 2001)

Table 4 Appendix 2: Correlation Matrix of Explanatory Variables

Variables	$\dot{k}$	$\dot{y}$	$\dot{tariff}$	$\dot{skill}$	$\dot{emp}$	*i_im*	*x*	*FDI*
$\dot{k}$	1							
$\dot{y}$	-0.1149	1						
$\dot{tariff}$	-0.3179	0.174	1					
$\dot{skill}$	-0.0219	0.1367	-0.0123	1				
$\dot{emp}$	0.0186	0.2205	0.1167	-0.1605	1			
i_im	0.2467	-0.1544	-0.2681	0.0091	0.0903	1		
x	0.0477	0.0517	-0.2266	0.0857	0.1565	0.6542	1	
FDI	-0.2433	0.6415	-0.1037	0.1394	0.0311	0.0421	0.2173	1

Source: Calculated by the author

References

1. Bank of Thailand, Macro Economic Indicators. Bank of Thailand, Bangkok (2011)
2. Beaudry, P., Collard, F.: Why Has the Employment-Productivity Tradeoff among Industrialized Countries Been So Strong? NBER Working Paper, Massachusetts NBER, pp. 1–43 (2002)
3. Blanchard, O.J., Diamond, P.: The Flow Approach to Labor Markets. The American Economic Review 82(2), 345–359 (1992)
4. Crotty, J.: Structural Contradictions of the Global Neoliberal Regime, Economics Department, University of Massachusetts, Massachusetts (2000)
5. Davis, D.R., Harrigan, J.: Good Jobs, Bad Jobs, and Trade Liberalization. Journal of International Economics (84), 26–36 (2011)
6. Gujarati, D.N.: Basic Economietrics. McGraw-Hill Companies, New York (2003)
7. Hine, R.C., Wright, P.W.: Trade with Low Wage Economies, Employment and Productivity in Uk Manufacturing. The Economic Journal 108(450), 1500–1510 (1998)
8. Ing, L.Y.: Lower Tariff, Rising Skill Premium in Developing Countries: Is It a Coincidence? The World Economiy, 1115–1133 (2009)
9. Jonsson, G., Subramanian, A.: Dynamic Gains from Trade: Evidence from South Africa. International Monetary Fund 48(1), 197–244 (2001)
10. Leon-Ledesma, M.A.: Economic Growth and Verdoorn's Law in the Spanish Regions, 1962-91. International Review of Applied Economics 40(1), 55–69 (2000)
11. Levinsohn, J.: Employment Responses to International Liberalization in Chile. Journal of International Economics 47, 321–344 (1999)
12. Manprasert, S.: Overview of the Thai Economy, `http://pioneer.netserv.chula.ac.th/~msompraw/Overview.pdf` (accessed July 2012)
13. Melitz, M.J.: The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica 71(6), 1695–1725 (2003)
14. National Social and Economic Development Board, National Social and Economic Development Plans (2009), `http://www.nesdb.go.th/Default.aspx?tabid=62` (accessed February 1, 2010)
15. Phan, P.: Trade Liberalization and Manufacturing Performance in Thailand 1990-2000. School of Economics. The University of Wollongong, Wolllongong (2004)

16. Quandt, R.E., Rosen, H.S.: Endogeneous Output in an Aggregate Model of the Labour Market. National Bureau of Economic Research (74), 394–400 (1989)
17. Ramstetter, E.D.: Labor Productivity, Wages, Nationality, and Foreign Ownership Shares in Thai Manufacturing, 1996–2000. Journal of Asian Economics 14, 861–884 (2004)
18. Ramstetter, E.D., Sjoholm, F.: Multinational Corpoerations in Indonesia and Thailand. Palgrave Macmillan, New York (2006)
19. Sjholm, F.: Exports, Imports and Productivity: Results from Indonesian Establishment Data, 33 August 1997. Working Paper No, pp. 1–33. Stockholm School of Economics, Stockholm (1997)
20. Statistical Office, Industrial Survey, National Statistical Office. National Statistical Office, Bangkok (2007)
21. Verhoogen, E.: Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector. The Quarterly Journal of Economics 123(2), 489–530 (2008)
22. Wachtel, H.M.: Labor and the Economy. Academic Press, Inc., Florida (1984)
23. Weiss, J.: Export Response to Trade Reform: Recent Mexican Experience. Development Policy Review 10(1), 43–60 (1992)
24. Wilkinson, B., Gamble, J., Humphrey, J., Morris, J., Anthony, D.: The New International Division of Labour in Asian Electronics: Work Organization and Hu. Journal of Management Studies 38(5), 675–695 (2001)
25. World Bank, World Integrated Trade Solution (2011), `http://wits.worldbank.org/wits/index.html` (accessed September 6)

Modeling Dependence Dynamics of Air Pollution: Time Series Analysis Using a Copula Based GARCH Type Model

He Zhanqiong, Songsak Sriboonchitta, and Dai Jing

Abstract. This paper investigates the dependence structure between the Air Pollution Index (API) of Shenzhen and corresponding regional, national levels based on copula based GARCH models. In particular, time varying normal copula and time varying SJC copula are compared and employed to model the dependence structure. Comparison with the results of DCC-GARCH model is made. We find that there exists significant asymmetric upper and lower tail dependence between Shenzhen and regional, national levels; tail dependence captures the change in dependence better; dependence structure change across time. Our findings have implications for environmental management.

1 Introduction

Air pollution in china is attracting the focus of not only the Chinese government and people but also researchers worldwide. There are disputes about the reliability of the Air Pollution Index (API) the government released and the pollutant detected and included to get the API. Despite of the disputes, since the composite and method of computing the integrated air pollution APIs stay unchanged during our observation, exploring spatial dependence dynamic through examining the conditional dependences of urban API and regional, national levels are feasible and meaningful.

Some previous studies examined spatial contagion of air pollution. But they mostly focus on some pollutants and dust (Yongxin Zhang et al. 2010; Tracey Holloway et al. 2008; Lee et al. 2010; Chung-MingLiu et al. 2006; F. Cousin et al. 2005),

He Zhanqiong
Faculty of Management & Economics, Kunming University of Sci.& Tech.,
Faculty of Economics, Chiang Mai University, Thailand, 650053, Kunnming, China
e-mail: hezhanqiong@gmail.com

Songsak Sriboonchita
Faculty of Economics, Chiang Mai University, P.O. Box 650224, Thailand
e-mail: songsak@econ.cmu.ac.th

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 215–226.
DOI: 10.1007/978-3-642-35443-4_15

studies concerning about API, an integrated index are not noticed. And for geographical scale, most previous studies examined contagion between regions within one county, or between countries (Feng Xiao et al. 2006; Guor-Cheng Fang et al. 2010; Paul J. Miller et al. 1998; Suhejla Hoti et al. 2005), spatial contagion between city, regional and national API is yet to be carried out.

In our previous studies, based on the DCC-GARCH model (Engle, 2002), we investigated the dynamic correlation of Air Pollution Indices (APIs) between 42 Chinese sample cities and their corresponding regional and national levels for a duration of 10 years. Some meaningful findings were drawn which were shared by most sample cities, for example, the correlations of local APIs between regional and national levels are time varying; most cities exhibit positive conditional corre-lations with both regional and national APIs, and the conditional correlations of most cities with regional and national APIs are only slightly different and are mostly stable. Whats interesting is that, the behavior of Shenzhen and Zhuhai, two cities with only 128km, the shortest distance among the subjected cities, exhi-bit unique characteristic between each other: a decrease of dynamic correlation with both regional and national levels after spring 2001, and then increase again after autumn 2004. Its not surprising that Shenzhen and Zhuhai behave similarly. Since they are so close to each other geometrically, with 56.4 kilometers of direct distance, and according to our integrating method of city related regional and national APIs, they are expected to have similar regional and national APIs. To further explore the dynamic spatial contagion feature of theses too cities, we focus on Shenzhen in this study. Shenzhen lies in Pearl River Delta, one of the three key regions required to carry out inter-region cooperation to cut and improve air quality. The reason why we choose Shenzhen is that Shenzhen has higher population density, higher GDP per capita (18 thousand USD for Shenzhen while 14 thousand for Zhuhai in 2011), and experiences higher API.

DCC-GARCH model, as a conventional linear-based correlation method is somewhat restrictive due to its requirements of normality for the joint distribution and of linear relationships among variables. More flexible copula-based models have become a common practice to cope with dependence between random variables but was mostly used to study the financial market. But this method has not been used in the API dependence study. To assess the changing dependence structures over time, following our previous research, this paper attempts to investigate time varying air pollution dependence between Shenzhen and its corresponding regional, national levels. Comparison between the DCC-GARCH model based result and copula based result will be made.

This study contribute to the existing literature not only by focusing on the dependence structure of urban and regional, national air pollution, but also by trying to apply the copula based GARCH type models to air pollution co-movement study.

The rest of the paper is set up in the following manner. Section 2 presents the econometric model. Section 3 contains the description of the data. The empirical results are in Section 4, followed by conclusion in the last section.

2 Model

2.1 DCC-GARCH Model

Time varying correlations are often estimated with multivariate generalized auto-regressive conditional heteroskedasticity (MGARCH) models. DCC models proposed by Engle which can be estimated very simply with univariate or two step methods based on the likelihood function, is an important one. In this paper, we employed the DCC model, after Engle (2002) to examine the existence of volatili-ty in each series and the dynamic correlations between urban APIs, regional APIs and national API.

Let us consider the APIs $Y_t = (Y_{1t}, \ldots, Y_{kt})'$, for $t = 1, 2, \ldots, T$. The following mean equation was estimated for each series given as:

$$Y_{it} = \mu_i + Y_{it-1} + \varepsilon_{it}, \qquad where \quad \varepsilon_{it} \sim N(0, H_t) \tag{1}$$

where Y_{it} is API in series i at time t, i is either city, regional or national API, ε_{it} is the error term for the API i at time t. All estimated series exhibited evidence of ARCH effects. DCC (Engle 2002) parameterization of conditional covariance metrics is given as:

$$H_(t) = D_t R_t D_t \tag{2}$$

where D_t is the $k \times k$ diagonal matrix of time varying standard deviations from univariate GARCH models with $\sqrt{h_{it}}$ on the i th diagonal, and R_t is the time varying correlation matrix. The elements of D_t is$\sqrt{h_{it}}$. For simplicity,h_{it} can be expressed for the univariate form as:

$$h_{it} = \omega_i + \sum_{p=1}^{p_i} \alpha_{ip} r_{it-p}{}^2 + \sum_{q=1}^{q_i} \beta_{iq} h_{it-q} \tag{3}$$

for $i = 1, 2, \ldots, k$ with the usual GARCH restrictions for non-negativity and stationarity being imposed, such as non-negativity of variances and $\sum_{p=1}^{p_i} \alpha_{ip} + \sum_{q=1}^{q_i} \beta_{iq} < 1$. To investigate the seasonal effect of mean and variance, and the effect on the dynamic correlation between local API, regional API and national API, we set three seasonal dummy in both mean and variance equations, so that equation (1) now becomes:

$$Y_{it} = \mu_i + S_2 D_2 + S_3 D_3 + S_4 D_4 + \alpha Y_{it-1} + \varepsilon_{it} \tag{4}$$

Equation (3) becomes:

$$h_{it} = \omega_i + S_2' D_2 + S_3' D_3 + S_4' D_4 + + \sum_{p=1}^{p_i} \alpha_{ip} r_{it-p}{}^2 + \sum_{q=1}^{q_i} \beta_{iq} h_{it-q} \tag{5}$$

So that D is seasonal effect vector where D_2, D_3, D_4 equals 1 when t is in summer, autumn, or winter respectively, other equations same. Spring includes March, April and May; summer includes June, July and August; autumn includes September, October and November; winter includes December, January and February.

2.2 Copula Concept

Copulas are functions that join or couple multivariate distribution functions to their uniform one-dimensional marginal distribution functions (Roger B. Nelsen, 2006). Sklar(1959) showed that a joint distribution can be factored into the margins and a dependence function called a copula. For bivariate case, let X and Y be two continuous random variables with margins $F(x)$ and $G(y)$ and with a joint distribution function $H(x,y)$, Sklar's theorem states that the standard representation for the joint distribution H is:

$$H(x,y) = C\Big(F(x), G(x)\Big) \tag{6}$$

where $C(u,v)$, $u = F(x)$ and $v = G(y)$ is the copula that captures the dependence structure between X and Y. If the margins are continuous, then C is uniquely determined, otherwise, the copula C is uniquely determined on $Ran(F) \times Ran(G)$. Thus, copulas can be used to link margins to a multivariate distribution function, which, in turn, can be decomposed into its univariate marginal distributions and a copula capturing the dependence structure between the two variables. Patton (2006) extended Sklar's theorem for conditional distributions. By extending Sklar's theorem, the conditional copula function can be written as:

$$H_(x,y|w) = C(F(x|w), G(y|w)|w) \tag{7}$$

where W is the conditioning variable, $F(x|w)$ is the conditional distribution of $X|W = w$, $G(y|w)$ is the conditional distribution of $Y|W = w$ and $H(x,y|w)$ is the joint conditional distribution of $(X,Y)|W = w$. Given the condition that F and G are differentiable, H and C are twice differentiable, the unconditional and conditional joint densities are given by:

$$f(x,y) = f(x) \cdot g(y) \cdot c(u,v) \tag{8}$$

$$f(x,y|w) = f(x) \cdot g(y) \cdot c(u,v|w) \tag{9}$$

2.3 The Models for the Marginal Distributions

APIs series in this study exhibit volatility clustering feature. To capture the most important features of air pollution index, such as fat tails or leverage effects, and seasonality of the first and the second moments, the marginal models of the APIs are estimated by three most widely used models and then choose which one outperforms other two: In this paper, volatility models to be estimated are associated with a stationary AR (1) conditional means given by:

$$Y_t = \mu + \theta Y_{t-1} + \varepsilon_t, \qquad where \quad |\theta| < 1 \tag{10}$$

where, Y_t is Air Pollution Index, ε_t is shock to API. Conditional variance covariance equations we examines in this study are Generalised autoregressive conditional heterocedasticity (GARCH) model (Bollerslev, 1986), GJR-GARCH, the threshold

GARCH (TGARCH) (Glosten, Jaganathan and Runkle, 1993) which is a simple extension of the GARCH scheme with extra term(s) added to account for possible asymmetries and EGARCH model of Nelson(1991) which can also accommodate asymmetry and species the conditional variance in a different way. To capture the seasonal effect in our data, we include seasonal dummy D in both mean equations and variance equations so that D is seasonal effect vector where D_2, D_3, D_4 equal 1 when t is in summer, autumn, or winter respectively, other things equal.

2.4 Copula Models

Copula methods have advantages over linear correlation in that the copula-based GARCH models allow for better exibility in joint distributions than bivariate normal or Student-t distributions. In this study, we are interesting in the time varying dependence of air pollution, especially time varying dependence of the propensity of air pollution to improve or deteriorate. So we focus on the conditional Symmetrized Joe-Clayton copula and conditional Gaussian copula of Patton (2006). The conditional Gaussian copula function is the density of the joint standard uniform variables (u_t, v_t), as the random variables are bivariate normal with a time-varying correlation, β_t. Moreover, let $x_t = \Theta^{-1}(u_t)$ and $y_t\Theta^{-1}(v_t)$, where $\Theta^{-1}()$ denotes the inverse of the cumulative density function of the standard normal distribution. The density of the time-varying Gaussian copula is then:

$$c_t^{Gau}(u_t, v_t|\rho_t) = \frac{1}{\sqrt{1-\rho_t}} \exp\{\frac{2\rho_t x_t y_t - x_t^2 - y_t^2)}{2(1-\rho_t^2)} + \frac{x_t^2 + y_t^2}{2}\} \tag{11}$$

The dependent process of the time varying Gaussian copula has the follow-ing form:

$$\rho_t = \Lambda_1\left(\omega + \beta\Lambda_1^{-1}(\rho_{t-1}) + \alpha\frac{1}{m}\sum_{i=1}^{m}\phi^{-1}(U_{1,t-1})\phi^{-1}(U_{2,t-1})\right) \tag{12}$$

$$\Lambda_1(x) = \frac{1-\exp(-x)}{1+\exp(-x)} \tag{13}$$

where$\Lambda_1()$ is a transformation function which holds the correlation parameter ρ_t in the interval $(-1,1)$, $\phi()$is the standard normal cdf and m is an arbitrary window length. The upper and lower-tail dependences of the conditional SJC copula is as:

$$T^U = \prod(\beta_U^{SJC} T_{t-1}^U + \omega_U^{SJC} + \gamma_U^{SJC}\frac{1}{10} + \sum_{i=1}^{10}|u_{t-1} - v_{t-1}| \tag{14}$$

$$T^L = \prod(\beta_U^{SJC} T_{t-1}^L + \omega_L^{SJC} + \gamma_L^{SJC}\frac{1}{10} + \sum_{i=1}^{10}|u_{t-1} - v_{t-1}| \tag{15}$$

where $\prod$ is the logistic transformation to keep T^U and T^L within the $(0,1)$ interval.

3 Data Description

The data series for this study comprises of 3 series of daily average Air Pollution Index (APIs) during the period from June 5th, 2000 to March 04th, 2010: API of Shenzhen; API of region to which Shenzhen belongs and national API with 3560 observations each series. Data on Shenzhen API comes from the data base of Ministry of Environmental Protection of the Peoples Republic of China (http://www.zhb.gov.cn//) (MEPPRC). The data of regional and national levels are integrated from APIs of the other cities within the region and nation respectively, by calculating inverse distance weighted average of city APIs for all other cities in the region and in the nation. The plot of APIs for Shenzhen and corresponding regional, national levels show that. Obvious volatility clustering feature can be noticed in all the three series. JB test shows that normality hypothesis is significantly rejected. Both the ADF unit root tests and PP test show that all the series are statistically significant. Rejecting the hypothesis that there exists unit root.

4 Results

Table 1 reports the estimation result of the two-step DCC model based on the univariate GJR-GARCH (1, 1) for each series, with the error skewed-t distribution assumption in all cases(We estimated with normal and student-t distribution assumption, skewed-t distribution outperform the other two). Results indicate that the assumption of constant conditional correlation for all shocks to APIs is not supported empirically. Both the condition mean and variance of spring are significantly lower than summer, but are higher than that of winter. The autumn is special, higher in mean but lower in variance compare with spring.

Table 2 reports the model specification for marginal distributions. Based on the log-likelihood value and Akaike, Schwarz information criteria, AR(1)-EGARCH model for Shenzhen and AR(1)-GJR-GARCH model for regional and national series outperform other models. So an AR-skewed-t-EGARCH model was employed for the marginal distributions of Shenzhen API and an AR-skewed-t-GARCH model employed for regional and national APIs. ARCH effect tests of residual didnt reject the null hypothesis of no serial correlation in the squared standardized residuals at 1% level, suggesting that the models listed capture the time varying volatility in the data very well.

Table 5 reports the univariate estimation result for each series chosen in previous step. Except several seasonal dummies, other estimates are significant at 1% level. The asymmetric effects in three series are all significant. The parameter of the conditional volatility equation in GJR-GARCH model is negative and highly significant, implying that negative shocks(good news) exert smaller impact on regional and national air pollution volatility than positive shocks (bad news) of the same magnitude. Similarly, in EGARCH model in Shenzhen API is positive and highly significant, implying that positive shock(bad news) exert bigger impact on Shenzhen air pollution.

Table 1 DCC estimation results

API	Shenzhen		Regional		National	
	coefficient	t-ratios	coefficient	t-ratios	coefficient	t-ratios
c	53.44***	-38.13	49.43***	-40.67	50.89***	-41.9
S_2	-11.27***	-6.44	-10.70***	-7.24	-10.68***	-7.25
S_3	6.75***	-3.44	1.891	-1.16	1.47	-0.91
S_4	10.28***	-4.96	9.70***	-5.82	9.15***	-5.52
AR	0.68***	-49.15	0.72***	-54.55	0.73***	-56.6
ω	80.90***	-3.86	62.66***	-4.78	52.31***	-4.71
S'_2	-28.84***	-2.90	-26.84***	-3.57	-23.44***	-3.65
S'_3	-7.98	-0.99	-10.53	-1.60	-10.12*	-1.84
S'_4	35.91***	-2.86	5.23	-0.76	4.5	-0.77
ARCH(α)	0.26***	-4.91	0.33***	-6.1	0.33***	-6.09
GARCH(β)	0.50***	-4.33	0.24*	-1.86	0.27**	-2.19
GJR(Gamma)	-0.31***	-5.88	-0.28***	-4.91	-0.28***	-4.97
$\Theta 1$	0.99***	-358.8				
$\Theta 2$	0.01***	-4.7				
Q(5)		3.06***		12.32***		5.65***
Q(10)		12.32***		26.71		33.84

Notes: This table reports the estimation results of DCC-GARCH models for the city APIs against regional and national APIs. The Q(5) and Q(10) are, respectively, the LjungBox autocorrelations test (1978) of five and lags in the standardized squared residuals from the regression. ***, **, * denote statistical significance at 1%, 5% and 10% level respectively.

From Table 4, we notice that the log-likelihood of time varying normal copula is higher than that of time varying SJC copula for both Shenzhen-regional estimation and Shenzhen-national estimation. But we hope to examine whether the feature of linear dependence also exists in tail dependence, so we focus on time varying conditional SJC copula, and compare the dependence behaviors implied by DCC, time varying normal copula and time varying SJC copula.

Figure 1 presents the plot of time varying dependence implied by DCC-GARCH model. We notice decline of dependence both between Shenzhen and regional, Shenzhen and national air pollution from the end May, 2001 and reach a bottom in October 2004 and then gradually increase to a high level, over 0.7.

In figure 2, we present plots of conditional dependence based on the time varying normal copula. Compare with figure 1, the time varying dependence implied by time varying normal copula exhibits a lower dependence during June 2001 to November 2004, but is not that significant as in figure 1.

Figure 3 is the plot of conditional tail dependence implied by the time varying SJC copula model. The dynamics of conditional upper and lower tail dependence were confirmed. Further, we notice that both upper and lower tail depen-dence exhibit a decline of dependence in late spring 2001, and then increase again autumn 2004, which is in conform to what we notice from the dependence implied by DCC-GARCH. This feature is not such clear for the dependence from the time varying normal copula.

Table 2 Model specification for the marginal distributions

	AR(1)-Garch(1,1) Skew T	AR(1)-Garch(1,1) - T	AR(1) -Egarch Skew T	AR(1) Egarch(1,1) t	AR(1) -JGR(1,1) Skew T	AR(1)-JGR(1,1) T
Shenzhen						
Log-likelihood	-14359.15	-14374.31	-14301.84	-14326.9	-14312.31	-14338.17
Akaike	8.07	8.08	8.04	8.06	8.05	8.06
Schwarz	8.1	8.1	8.07	8.08	8.07	8.09
ARCH 1-2 test	0.43[0.64]	0.88[0.41]	0.46[0.63]	-	1.57[0.20]	1.16[0.31]
ARCH 1-5 test	1.20[0.30]	2.18[0.05]	0.38[0.86]	-	1.24[0.28]	1.14[0.33]
ARCH 1-10 test	1.40[0.17]	1.63[0.08]	1.03[0.40]	-	1.23[0.26]	1.21[0.27]
Regional						
Log-likelihood	-12937.2	-12951.75	-	-	-12907.44	-12930.25
Akaike	7.28	7.28	-	-	7.26	7.27
Schwarz	7.3	7.3	-	-	7.28	7.29
ARCH 1-2 test	0.12[0.88]	0.089[0.9]	-	-	0.47[0.620]	0.47[0.621]
ARCH 1-5 test	0.26[0.93]	0.28[0.92]	-	-	0.37[0.865]	0.34[0.888]
ARCH 1-10 test	1.42[0.16]	1.38[0.17]	-	-	1.81[0.053]	1.73[0.068]
National						
Log-likelihood	-12700.72	-12712.35	-12664.63	-	-12670.05	-12688.92
Akaike	7.14	7.15	7.12	-	7.13	7.14
Schwarz	7.17	7.17	7.15	-	7.15	7.16
ARCH 1-2 test	0.06[0.93]	0.04[0.95]	0.10[0.90]	-	0.35[0.70]	0.35[0.70]
ARCH 1-5 test	0.27[0.92]	0.30[0.91]	0.33[0.89]	-	0.35[0.87]	0.31[0.90]
ARCH 1-10 test	1.35[0.19]	1.34[0.20]	1.45[0.15]	-	1.82[0.05]	1.76[0.06]

Note: - no convergence. Standard errors are indicated in bracket.

Table 3 Results for the marginal distributions

Parameters	Shenzhen	Regional	National
Cst(M)	49.46^{***}(1.18)	50.79^{***} (1.20)	52.96^{***}(1.37)
d2(M)	-11.72^{***}(1.35)	-11.55^{***}(1.36)	-11.58^{***}(1.69)
d3(M)	2.028(1.69)	1.72(1.70)	7.97^{***}(2.10)
d4(M)	10.61^{***}(1.61)	10.10^{***}(1.67)	11.60^{***}(2.07)
AR(1)	0.73^{***}(0.01)	0.75^{***}(0.01)	0.69^{***}(0.01)
Cst(V)	57.51^{***}(9.07)	48.51^{***}(7.94)	5.35^{***}(0.08)
d2(V)	-30.38^{***}(6.23)	-26.08^{***}(5.37)	-0.48^{***}(0.11)
d3(V)	-6.77(5.53)	-6.86(4.63)	-0.02(0.10)
d4(V)	8.92(6.25)	7.79(5.31)	0.48^{***}(0.09)
ARCH(Alpha1)	0.43^{***}(0.06)	0.42^{***}(0.06)	0.13(0.19)
GARCH(Beta1)	0.27^{***}(0.09)	0.30^{***} (0.09)	0.56^{***} (0.14)
GJR(Gamma)	-0.40^{***}(0.06)	-0.39^{***} (0.06)	
EGARCH(Gamma)			0.40^{***} (0.03)

Notes: Standard errors for the estimators are included in parentheses. ***indicate signicant at the 1% level.

Table 4 Copula estimation results

	Shenzhen-Regional	Shenzhen-National
Time-varying normal copula		
Constant	0.0244	0.0205
α	0.2681	0.2638
β	1.9531	1.9756
LL	-595.0485	-624.5461
Time-varying SJC copula		
Constant^U	1.2044	1.1301
α^U	-11.1693	-11.2216
β^U	0.3192	0.4682
Constant^L	-1.6446	-1.7493
α^L	-1.4649	-1.0727
β^L	3.7598	3.8549
LL	-657.3164	-697.9039

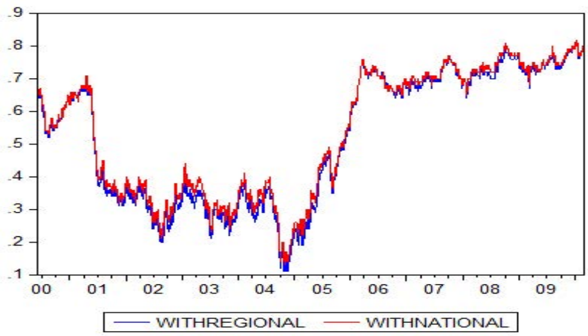


Fig. 1 Dynamic correlation estimated from DCC-GARCH model. Note: Red line displays the implied time paths of the conditional dependence between Shenzhen API and national API, blue line between Shenzhen API and regional API.

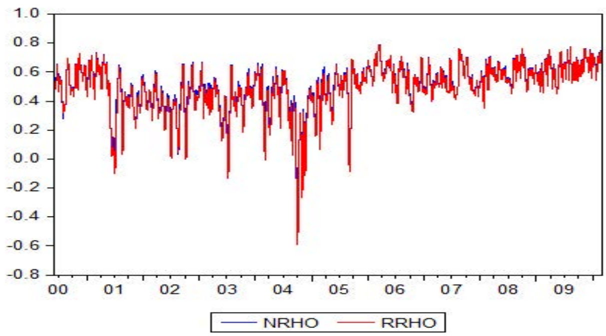
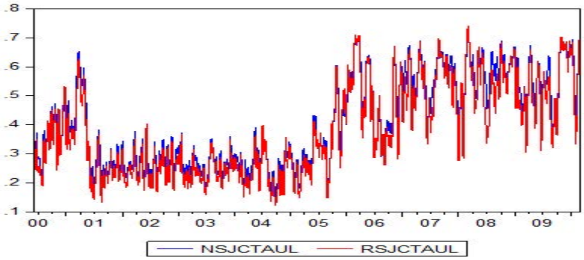
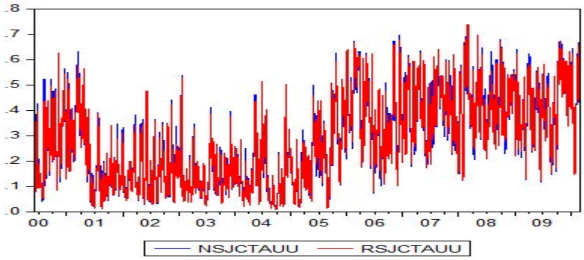


Fig. 2 Conditional dependence implied by time varying normal copula. Note: Red line displays the implied time paths of the conditional dependence between Shenzhen and regional, blue line between Shenzhen and national.

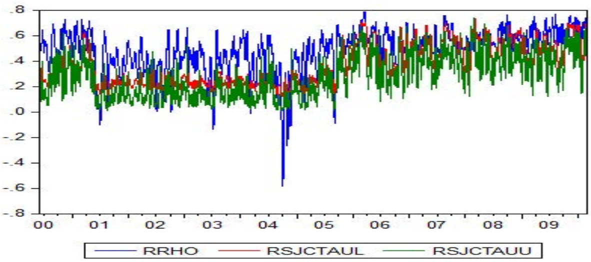


(a) Conditional dependence in lower tail implied by time varying SJC copula

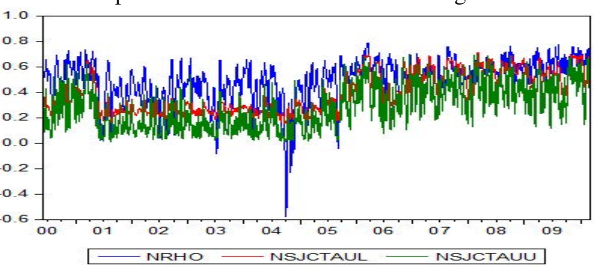


(b) Conditional dependence in upper tail implied by time varying SJC copula

Fig. 3 Conditional dependence implied by time varying SJC copula



(a) Conditional dependence between Shenzhen and regional API



(b) Conditional dependence between Shenzhen and national API

Fig. 4 Conditional dependence estimate from the copula models

5 Conclusions

In this study, we examine time varying dependence between air pollution of Shenzhen and corresponding regional, national levels by modeling the conditional dependence structure via copula time varying SJC copula. The Engle DCC and time varying normal copula are estimated as comparison. Univariate estimations reveal that in all three seasons, seasonal effect exists in both mean and variance equations, with lower mean and variance in summer and higher in winter. The asymmetric effects are all significant, bad news exert bigger impact on air pollution of Shenzhen, regional and national levels. The change in dependence during the time period from end May 2001 to October 2004 we found in DCC model also takes place in dependence implied by time varying normal copula and time varying SJC copula. This feature is not very obvious in time varying normal copula, but it is very clear in both upper and lower tail dependence implied by time varying SJC copula. This may imply that the change come mostly from extreme value. Further, the existence of asymmetry is confirmed. We notice that lower tails are higher than upper tails in both Shenzhen-regional and Shenzhen-national relationships, indicating Shenzhen will benefit from the improved regional and national air quality; the decline of regional and national air quality will affect the contemporaneous air quality of Shenzhen, but with lower impact. Asymmetry increase after October 2004 and increasing with the level of dependence, suggesting the change of dependence structure over time. DCC, time varying copula and time varying SJC copula all reveal that the conditional dependence between Shenzhen and national is slightly higher than that of Shenzhen and regional.

These results have strong policy implications. When capturing the regional and inter-region relationship, seasonal variation should be taken into consideration. Spring and Winter exhibit higher volatility, which means higher uncertainty; regional or single city settlement in air pollution control is important but not enough, inter-region cooperation and national decision are important; The cooperation mechanism should be able to respond the time varying nature of conditional correlation; regional heterogeneity should be considered in cooperation policy decision, cooperation among regions with higher correlation and similar correlation feature is better.

References

1. Bollerslev, T., Engle, R.F., Wooldridge, J.M.: A capital asset pricing model with time varying Covariances. Journal of Political Economy 96, 116–131 (1988)
2. Bollerslev, T.: Modeling the coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model. Review of Economics and Statistics 72, 498–505 (1990)
3. Liu, C.-M., Young, C.-Y., Lee, Y.-C.: Influence of Asian dust storms on air quality in Taiwan. Science of the Total Environment 368, 884–897 (2006)
4. Nelson, D.B.: Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59(2), 347–370 (1991)
5. Engle, R., Kroner, F.: Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150 (1995)

6. Engle, R.F.: Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH, http://weber.ucsd.edu/~ksheppar/research
7. Engle: Dynamic conditional correlation: asimple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Business Econ. Stat. 20, 339–350 (2002)
8. Cousin, F., Tulet, P., Rosset, R.: Interaction between local and regional pollution during Escompte 2001: impact on surface ozone concentrations (IOP2a and 2b). Atmospheric Research 74, 117–137 (2005)
9. Xiao, F., Brajer, V., Mead, R.W.: Blowing in the wind: The impact of China's Pearl River Delta on Hong Kong's air quality. Science of the Total Environment 367, 96–111 (2006)
10. Glosten, Lawrence, R., Jagannathan, et al.: On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. Journal of Finance, American Finance Association 48(5), 1779–1801 (1993)
11. Fang, G.-C., Chang, S.-C.: Atmospheric particulate (PM10 and PM2.5) mass concentration and seasonal variation study in the Taiwan area during 2000–2008. Atmospheric Research 98, 368–377 (2010)
12. Jondeau, E., Rockinger, M.: The copula GARCH model of conditional dependencies: an international stock market application. Journal of International Money and Finance 25, 827–853 (2006)
13. Longmore, R., Robinson, W.: Modelling and Forecasting Exchange Rate Dynamics: An Application of Asymmetric Volatility Models. Bank of Jamaica. Working Paper WP 2004/03 (2004)
14. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
15. Patton, A.J.: Modelling Asymmetric Exchange Rate Dependence. International Economic Review 47, 527–556 (2006a)
16. Miller, P.J., Grumet, J.: Bi-national and regional issues in air quality science and policy. Environmental Science & Policy 1, 223–230 (1998)
17. Hoti, S., McAleer, M., Chan, F.: Modelling the spillover effects in the volatility of atmospheric carbon dioxide concentrations. Mathematics and Computers in Simulation 69, 46–56 (2005)
18. Holloway, T., Sakurai, T., Han, Z., et al.: MICS-Asia II: Impact of global emissions on regional air quality in Asia. Atmospheric Environment 42, 3543–3561 (2008)
19. Lee, Y.C., Yang, X., Wenig, M.: Transport of dusts from East Asian and non-East Asian sources to Hong Kong during dust storm related events 1996-2007. Atmospheric Environment 44, 3728–3738 (2010)
20. Zhang, Y., Olsen, S.C., Dubey, M.K.: WRF/Chem simulated springtime impact of rising Asian emissions on air quality over the U.S. Atmospheric Environment 44, 2799–2812 (2010)

Estimating Time-Varying Systematic Risk by Using Multivariate GARCH

Muttalath Kridsadarat

Abstract. The purpose of this study is aim to estimate the time-varying systematic risk or beta by using multivariate GARCH models. Since there are several researches found that estimating systematic risk by market model using traditional regression approach violated classical assumptions in both stationary assumption and independent identically distributed of the innovations. Then, the study focuses on using multivariate GARCH to improve the beta estimation since GARCH model is the popular model used in volatility clustering data. There are three type of Multivariate GARCH used in this study to compare the forcasting ability of each model of Multivariate GARCH. The results show from the plots of beta that Multivariate GARCH model can catch up volatility of risk quicker and better than Ordinary Least Square model and from model performance evaluation, vech model Multivariate GARCH confirms the superiority in capturing Time-Varying Systematic risk among the other Multivariate GARCH.

1 Introduction

Since the recent economic recession in 2008 has drawn attention in the fragile of financial market to both investors and financial market participants. Therefore, understanding the market risk precisely is the advantage for all of them to coup with the volatility of the market and prompts for the investment opportunity. One of the most important risks of financial market is systematic risk which means risk that associate with market returns. Systematic risk is the market risk that investor cannot avoid by diversification (Maginn, Tuttle, McLeavey and Pinto (2007)). From the view of modern theory of finance, systematic risk is an extremely essential risk because it is the only one type of risk that should be rewarded (for example, Sharp

Muttalath Kridsadarat
National Institute of Development Administration,
71/113 Sukumwit Rd., Muang, Samutprakarn
e-mail: muttalath@gmail.com

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 227–239.
DOI: 10.1007/978-3-642-35443-4_16

(1964); Black (1972); Lintner (1965) and Mossin (1966)). Therefore, the proper model that can estimate the systematic risk is the essential financial instrument for investors and market participants. The commonly used of systematic risk estimation is market model under ordinary least squares which provides the simplicity of estimation. Ordinary Least Square Regression is widely known in the term of simplicity. However, it violated stationary assumption. The nonstationarity in parameters, nonstationary of error terms and the intertemporal dependences in the number of outliers are founded in the ordinary least square regression model (e.g., Bey and Pinches, 1980)

These imply that the model is non-Gaussian. However, there is an alternative hypothesis that allows the systematic risk varying through time (Bos and Newbold (1984); Fabozzi and Francis (1978)) There are various new different econometric methods that have been used to estimate time-varying beta and one of the most famous method is multivariate GARCH.

By using multivariate GARCH model, the model allows forecasting variance of return to vary systematically along the periods that is consistent with this alternative hypothesis. Therefore, systematic risk of the market can be estimated more precisely by using multivariate GARCH model. The market participants will understand more on beta that represent the characteristics of the firms and industries. Koutmos and Knif (2002) did the research about Estimating systematic risk using time varying distributions. The research has used the vector GARCH model presented by Bollerslev, Engle and Wooldridge (1988) to estimate the time varying beta. The data used in their study are collected from the daily closed prices of five sector portfolios and market index of the stock market in four countries which are Germany, Japan , UK and USA. They found that estimating beta by traditional regression violated the assumpmion of the Ordinary Least Squared method and they estimated the single beta model by using bivariate GARCH model. The results has shown that all portfolios present time varying variance. However, there are several forms of multivariate GARCH model. Choudhry Taufiq and Wu Hao (2009) has studied about time-varying beta forecasting ability by using data from UK Companies. They estimated time-varying beta and compared the ability to forecast of three types of GARCH models and Kalman filter method. Another article that supports the proof of time-varying beta is the study of Mergner and Bulla (2008). This paper investigated the time varying behavior of systematic risk for eighteen pan-European sectors. The paper has contributed an investigation of time-varying betas for pan-European industry portfolios.. And they also compared the forecasting accuracy of these three GARCH models which are the bivartate GARCH, BEKK GARCH, and GARCH-GJR. They evaluated the performance of the model by calculating the mean squared error (MSE) and mean absolute error (MAE). Among the GARCH models, the GARCH-GJR model appears to be more accurate forecasts than the bivariate GARCH and the BEKK models, followed by bivariate GARCH.

From the Literature reviews that are mentioned above, since the ordinary least squared need many assumption and most of ths studies found that the beta from market model estimated by ordinary least square violate these assumption. This difficulty becomes the motivation of this study. This study will contribute the time varying systematic risk of each industry index in Thailand Stock Market by using the proper model which is Multivariate GARCH model. However, this study thus focuses on three type of multivariate GARCH which are vech model Multivariate GARCH, BEKK model Multivariate GARCH, and CCC model Multivariate GARCH. This study will compare both advantages and disadvantages of three potential GARCH models in calculating more precise time-varying beta. In addition the GARCH model that concern of asymmetric effect should be considered in this paper as well.

2 Methodology and Data

The step of studies in this paper separated into five steps. Firstly, the study provides the evidence that betas follow time-varying by estimating the market model using standard regression approach to show the problem of this model. Secondly, the model was changed to the multivariate GARCH in three models which are vech model, BEKK model and CCC model. The forms of them will be mentioned in this following section. The results are expected to see the time-varying in variance. The coefficients that link current variance to its own past history as well as past innovations should be statistically significant. Next, the study investigates asymmetries in covariance by examining the statistically significant of $\delta_{i,m}$ in vech model. The results are expected to illustrate the negative coefficient to explain that there exist asymmetric pattern in covariance. Next, the study plots the beta that is estimated by these various methods to see the pattern and trend of systematic risk in each industry and to see how each model capture the change in systematic risk. Finally, the performance of multivariate GARCH and the ordinary least square are compared. The study will calculate the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) to examine it.

The scope of this study focuses on Thailand Stock Exchange systematic risk. Therefore, the data used are collected from closed prices of the industry index of Thailand Stock Exchange and Market Index (SET). There are eight industries Index separated by Thailand Stock Market which are Agriculture Product and Food Industry, Consumer Products, Financials, Industrials, Property and Construction, Resource, Services and Technology. The frequency of data collected is daily and the period of data started from January 2008 to June 2012. Therefore, there are 953 Observations provided for the test. This paper uses closed price of each industry to compute the return of each industry and uses the closed price of SET to compute the market portfolio return. This follows market or single index model. And the return of industry index (i) and Market are calculated by continuous compound return method. Then the daily returns are computed by this following formula

$$R_i = 100 * log(P_{i,t}/P_{i,t-1})$$

R_i is the compound return of industry i and $P_{i,t}$ and $P_{l,t-1}$ stand for the closed price of industry index i at time t and time t-1 respectively. The first model used in this study is vech model Multivariate GARCH modell followed Bollerslev (1990). The diagonal vech model or vech model specification has an important advantage on captures the contemporaneous correlation between the various error terms. Then the coefficients that are estimated by this extended multivariate GARCH model will be more efficient than using a set of single equation estimations. The vech model diagonalizes the system by let the variance and covariance equations contain only lags of itself and the cross product of residuals ($\varepsilon_{i,t}$, $\varepsilon_{j,t}$).

The vech model used in this study is modified to capture asymmetric in covariance can be shown by these following set of equations.

$$R_{i,t} = \mu_{i,t} + \sigma_{i,t} Z_{i,t} \tag{1}$$
$$R_{m,t} = \mu_{m,t} + \sigma_{m,t} Z_{m,t} \tag{2}$$
$$\sigma^2_{i,t} = \alpha_{i,0} + \alpha_{i,1}\varepsilon^2_{i,t-1} + \alpha_{i,2}\sigma^2_{i,t-1} + \varepsilon_i\, S_{i,t-1}\delta^2_{i,t-1} \tag{3}$$
$$\sigma^2_{m,t} = \alpha_{m,0} + \alpha_{m,1}\varepsilon^2_{m,t-1} + \alpha_{m,2}\sigma^2_{m,t-1} + \delta_m\, S_{m,t-1}\varepsilon^2_{m,t-1} \tag{4}$$
$$\sigma_{i,m,t} = \lambda_0 + \lambda_1\varepsilon_{i,t-1}\varepsilon_{m,t-1} + \lambda_{2,}\sigma_{i,m,t} + \delta_{i,m} S_{m,t-1}\varepsilon_{m,t-1} \tag{5}$$

$R_{i,t}$ and $R_{m,t}$ are the continuous compound return of industry i and market respectively. $\mu_{i,t}$ and $\mu_{m,t}$ are the conditional means. $\sigma^2_{i,t}$ and $\sigma^2_{m,t}$ are conditional variances and $\sigma_{i,m,t}$ is the conditional covariance. $\varepsilon_{i,t}$ and $\varepsilon_{m,t}$ are the error terms or innovations and $Z_{i,t}$ and $Z_{m,t}$ are the standardized innovations which can be calculated be $Z_{i,t} = \varepsilon_{i,t}/\sigma_{i,t}$ and $Z_{m,t} = \varepsilon_{m,t}/\sigma_{m,t}$. However, there is the different thing on normal form of vech model multivariate GARCH on the term $S_{i,t-1}$ and $S_{m,t-1}$ that is designed to capture potential asymmetry in conditional variance.

Where	$S_{j,t-1}$	=	1	; if	$\varepsilon_{j,t-1} < 0$
And	$S_{j,t-1}$	=	0	;	otherwise

The next Multivariate GARCH model used is BEKK model which is popularized by Engle and Kronos (1995). The model ensures that the conditional variances are always positive by putting the model in quadratic forms. Therefore, the conditional variances and conditional covariance equation depend on the square of residual or innovation and cross product of residuals. However, the drawback of BEKK model is the complicated of the model since there are large number of coefficients need to be estimated. So, this study put the restriction that all of the matrixes need to be diagonal matrix to reduce the number of coefficient and make the form of model simplify. So the model can be illustrated as shown and the explanation of the variables is the same as vech model.

$$R_{i,t} = \mu_{i,t} + \sigma_{i,t} Z_{i,t} \tag{6}$$
$$R_{m,t} = \mu_{m,t} + \sigma_{m,t} Z_{m,t} \tag{7}$$
$$\sigma^2_{i,t} = \gamma^2_{i,i} + \alpha^2_{i,1}\varepsilon^2_{i,t-1} + \alpha^2_{i,2}\sigma^2_{i,t-1} \tag{8}$$
$$\sigma^2_{m,t} = \gamma^2_{m,m} + \alpha^2_{m,1}\varepsilon^2_{m,t-1} + \alpha^2_{m,2}\sigma^2_{m,t-1} \tag{9}$$
$$\sigma_{i,m,t} = \alpha_{i,1}\alpha_{m,1}\varepsilon_{i,t-1}\,\varepsilon_{m,t-1} + \alpha_{i,1}\alpha_{m,1}\varepsilon_{i,t-1}\varepsilon_{m,t-1} + \alpha_{i,2}\alpha_{m,2} \tag{10}$$

The last one of Multivariate GARCH model in this study is constant conditional correlation model or CCC model. The model requires correlation coefficient to be constant. This formulation is a special case of the more general Multivariate GARCH model. And the same as BEKK model, CCC model need to estimate many parameters. Therefore, this study restricts the coefficient to be scalar and the formations of the model are shown below and the explanation of the variables is the same as vech momel.

$$R_{i,t} = \mu_{i,t} + \sigma_{i,t} Z_{i,t} \quad (11)$$
$$R_{m,t} = \mu_{m,t} + \sigma_{m,t} Z_{m,t} \quad (12)$$
$$\sigma^2_{i,t} = \gamma_{i,i} + \alpha_{i,1}\varepsilon^2_{i,t-1} + \alpha_{i,2}\sigma^2_{i,t-1} \quad (13)$$
$$\sigma^2_{m,t} = \gamma_{m,m} + \alpha_{m,1}\varepsilon^2_{m,t-1} + \alpha_{m,2}\sigma^2_{m,t-1} \quad (14)$$
$$\sigma_{i,m,t} = \gamma_{i,m}\sigma_{i,t}\sigma_{m,t} \quad (15)$$

The equation Eqs. 1, 2, 6, 7, 11, and 12 are the mean equations of each form of models and the equation Eqs. 3, 4, 8, 9, 13 and 14 are the set of variance equations. Moreover, since this study uses multivariate GARCH, then the set of equation in each type of model requires the covariance equation which are on equation Eqs. 5, 10 and 15. The specification of the model used in this study follow Bollerslev et al. (1988) exception that it allows for testing on asymmetric responses to up and down trend of the market which represent by $S_{j,t-1}$ in vech model.

3 Results

Firstly, the study tests the ordinary least square estimation of beta from the market model. All of the betas are significantly different from zero in all industries group at 99 percent confident interval which means that the systematic risk or market risk is significantly effect on industry return. Next, the study tests Q^2-stat (Ljung-Box Statistic)) to examine the ARCH effect in the estimation to investigate that whether estimating GARCH is proper or not. The results of estimation are shown in Fig. 1 below.

The results show that most of Ljung-Box Statistic for all five, ten and twenty lags show significance in estimation except Consumer Product Industry. Therefore, most of industries present ARCH effect or conditional heteroskedasticity except the Consumer Product Industry. Overall, estimating GARCH(1,1) in these equations can be used to improve precise of beta estimations by using Multivariate GARCH models.

Next, the study estimates vech model Multivariate GARCH. The results from this estimation are shown in Fig. 2.The results show that not surprisingly, all indexes show time varying variance. You can see fiom the significant of $\alpha_{i,2}$ that happen in all industries group. The coefficient that link current variance and its own past variance are all significantly different from zero. For the evidence that show whether past innovation can influence on current variance, the results show that all of the cases are true. The coefficient $\alpha_{i,1}$ of every industry is significantly different from zero.

Moreover, for the results of asymmetries, if the sign of most coefficients $\delta_{i,m}$ is negative and the coefficients $\delta_{i,m}$ are significant, it follows asymmetric responses

			Qsqrt(5)		Qsqrt(10)		Qsqrt(20)	
Industry Group	Beta	Prob.	Qsqrt(5)	Prob.	Qsqrt(10)	Prob.	Qsqrt(20)	Prob.
Agro&Food	0.609512***	0.0000	27.16	0.0000	32.88	0.0000	55.03	0.0000
Consumption	0.272076***	0.0000	0.62	0.9870	1.33	0.9990	2.53	1.0000
Financial	1.110659***	0.0000	64.99	0.0000	92.14	0.0000	162.56	0.0000
Industrial	1.036868***	0.0000	115.31	0.0000	128.62	0.0000	203.29	0.0000
Prop&Cons	0.939007***	0.0000	50.48	0.0000	76.27	0.0000	147.91	0.0000
Resource	1.235534***	0.0000	43.84	0.0000	44.48	0.0000	58.01	0.0000
Service	0.690750***	0.0000	88.89	0.0000	108.30	0.0000	127.51	0.0000
Technology	0.729566***	0.0000	14.44	0.0130	21.99	0.0150	51.20	0.0000

** Significant at 90% confident interval*
*** Significant at 95% confident interval*
**** Significant at 99% confident interval*
Q2(n) is the Ljung-Box Statistic calculated for the squared standardized residuals using n lags.

Fig. 1 Results of traditional regression approach estimation

Industry Group	αi,0	αi,1	δi	αi,2	λ1	δi,m	λ2
Agro&Food	0.003871	0.051027***	0.804050	0.091917***	0.073906	0.089339***	0.828667
	0.0000	*0.0043*	*0.0004*	*0.0002*	*0.0001*	*0.0085*	*0.0000*
Consumer	0.000020	0.069498***	0.491499	0.243565***	0.053649	0.064722***	0.867426
	0.0000	*0.0067*	*0.0000*	*0.0000*	*0.0007*	*0.0000*	*0.0000*
Financial	0.000017	0.060647***	0.059593	0.848056***	0.059536	0.062655***	0.855981
	0.0000	*0.0000*	*0.0016*	*0.0000*	*0.0000*	*0.0003*	*0.0000*
Industrial	0.000013	0.077538***	0.026272	0.876279***	0.075214	0.06407***	0.857640
	0.0000	*0.0000*	*0.0732*	*0.0000*	*0.0000*	*0.0011*	*0.0000*
Property&Construction	0.000006	0.07599***	0.013232	0.888074***	0.051125	0.042218***	0.890947
	0.0000	*0.0000*	*0.3673*	*0.0000*	*0.0000*	*0.0039*	*0.0000*
Resource	0.000011	0.037378***	0.077445	0.889228***	0.044461	0.889228***	0.873205
	0.0000	*0.0002*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Service	0.000008	0.066285***	0.065425	0.836***	0.042575	0.836009***	0.882257
	0.0000	*0.0000*	*0.0009*	*0.0000*	*0.0001*	*0.0000*	*0.0000*
Technology	0.000027	0.043656***	0.100733	0.778336***	0.051159	0.058922***	0.883471
	0.0000	*0.0243*	*0.0001*	*0.0000*	*0.0000*	*0.0002*	*0.0000*

** Significant at 90% confident interval*
*** Significant at 95% confident interval*
**** Significant at 99% confident interval*

Fig. 2 Result from multivariate GARCH estimation

of the market which specify that covariance will be higher during market decline. However, the results show that the coefficient $\delta_{i,m}$ of all industries are significantly different from zero but all of coefficients are positive. Consequently, from the results above, the study cannot conclude about the pattern of asymmetric of beta in this model.

Next, the study estimates multivariate GARCH in the form of BEKK model. The results show that all of industry coefficient $\alpha_{i,2}$ are significantly different from zero which means that all indexes show time varying variance. The finding of BEKK model is consistent with vech model. As same as vech model, the study examines the effect of past innovation to current variance and found that all industries provide

Industry Group	αi,1	αi,2	δi,m
Agro&Food	0.220630***	0.957577***	0.221471***
	0.0000	*0.0000*	*0.0000*
Consumer	0.142862***	0.976463***	0.207058***
	0.0000	*0.0000*	*0.0000*
Financial	0.266825***	0.958365***	0.170722***
	0.0000	*0.0000*	*0.0000*
Industrial	0.282237***	0.951748***	0.189514***
	0.0000	*0.0000*	*0.0000*
Property&Construction	0.271323***	0.958962***	0.138243***
	0.0000	*0.0000*	*0.0004*
Resource	0.220244***	0.961345***	0.246715***
	0.0000	*0.0000*	*0.0000*
Service	0.277828***	0.947477***	0.223314***
	0.0000	*0.0000*	*0.0000*
Technology	0.241959***	0.954045***	0.178816***
	0.2957	*0.0000*	*0.0000*

** Significant at 90% confident interval*
*** Significant at 95% confident interval*
**** Significant at 99% confident interval*

Fig. 3 Result from multivariate GARCH estimation (BEKK model)

significant coefficient $\alpha_{1,1}$. Consequently, the results explain that past innovation can influence current variance.

Finally, the study estimate by using CCC model. All of industries present Time-Varying variance since the result table shows that all industries coefficient$\alpha_{i,2}$ are significantly different from zero. Therefore, the results from CCC model are consistent with the result from vech model and BEKK model. Moreover, the study also observes that past innovation can influence on current variance by using CCC model. The results found that most of industry indexes show the pattern that past innovation has an effect on current variance except Technology industry which its coefficient $\alpha_{i,1}$is not significantly different from zero.

Next, the study will compare the beta in each industry by plotting the beta by five methods of calculating. The first method is Ordinary least square. The returns of each industry index and return of the market from January 2008 to June 2012 are used to calculate the beta. Therefore, this method will provide one value of beta and the graph will be the horizontal line along the period. The second method is OLS rolling estimate that improved from OLS. In this method, data will be separated into small period or window that each window contains 60 observations of data. In each window, one value of beta will be generated by OLS and roll the window along the period. Therefore, by this method beta plotting is not the horizontal line as OLS but it will change all the time. The rest methods are Multivariate GARCH with various models that are vech model, BEKK model and CCC model. The Time-Varying beta

Industry Group	αi,1	αi,2	Ri,m
Agro&Food	0.050259**	0.776268***	0.730828
	0.0415	*0.0000*	*0.0000*
Consumer	0.083345**	0.38004***	0.441852
	0.0111	*0.0000*	*0.0000*
Financial	0.049427***	0.833435***	0.898579
	0.0006	*0.0000*	*0.0000*
Industrial	0.087376***	0.846044***	0.817324
	0.0000	*0.0000*	*0.0000*
Property&Construction	0.08793***	0.88181***	0.909619
	0.0000	*0.2223*	*0.0000*
Resource	0.023423**	0.871996***	0.934491
	0.0137	*0.0000*	*0.0000*
Service	0.057457**	0.799242***	0.819696
	0.0149	*0.0000*	*0.0000*
Technology	0.022993	0.714548***	0.651051
	0.2957	*0.0000*	*0.0000*

** Significant at 90% confident interval*
*** Significant at 95% confident interval*
**** Significant at 99% confident interval*

Fig. 4 Result from multivariate GARCH estimation (CCC model)

from these three models of Multivariate GARCH can be calculated by this method as formula shown below.

$$\beta_{i,t} = \sigma_{i,m,t} / \sigma_{m,t}^2$$

The time-varying market variance and covariance with the industry i are estimated by the conditional covariance equations divided by the conditional variance equation of the market in each model. The beta calculated by multivariate GARCH method will move up and down along the period and is expected that it wlll move along the beta calculated by OLS.

From the plots, they show that three series of beta from these three Multivariate GARCH models and the beta from rolling OLS move up and down along OLS beta. However, by using Multivariate GARCH and rolling OLS, the trend of beta can be observed the clearly. The obvious one is Industrial Industry, the betas from Multivariate GARCH and rolling OLS provide the increasing trend. By OLS, the beta of Industrial is 0.97 that is very close to market index. When considering the Time-Varying beta calculated by OLS Rolling estimate and Multivariate GARCH the study found that Time-Varying Beta of both methods has large fluctuation and can observe increasing trend better. They fluctuate from around 0.6 to 1.9 and present the increasing trend as you can observe from graph and using OLS beta estimation cannot observe it. This is the advantage of Portfolio adjustment for market participants by using the precise model. It can show the rapid move of systematic risk and

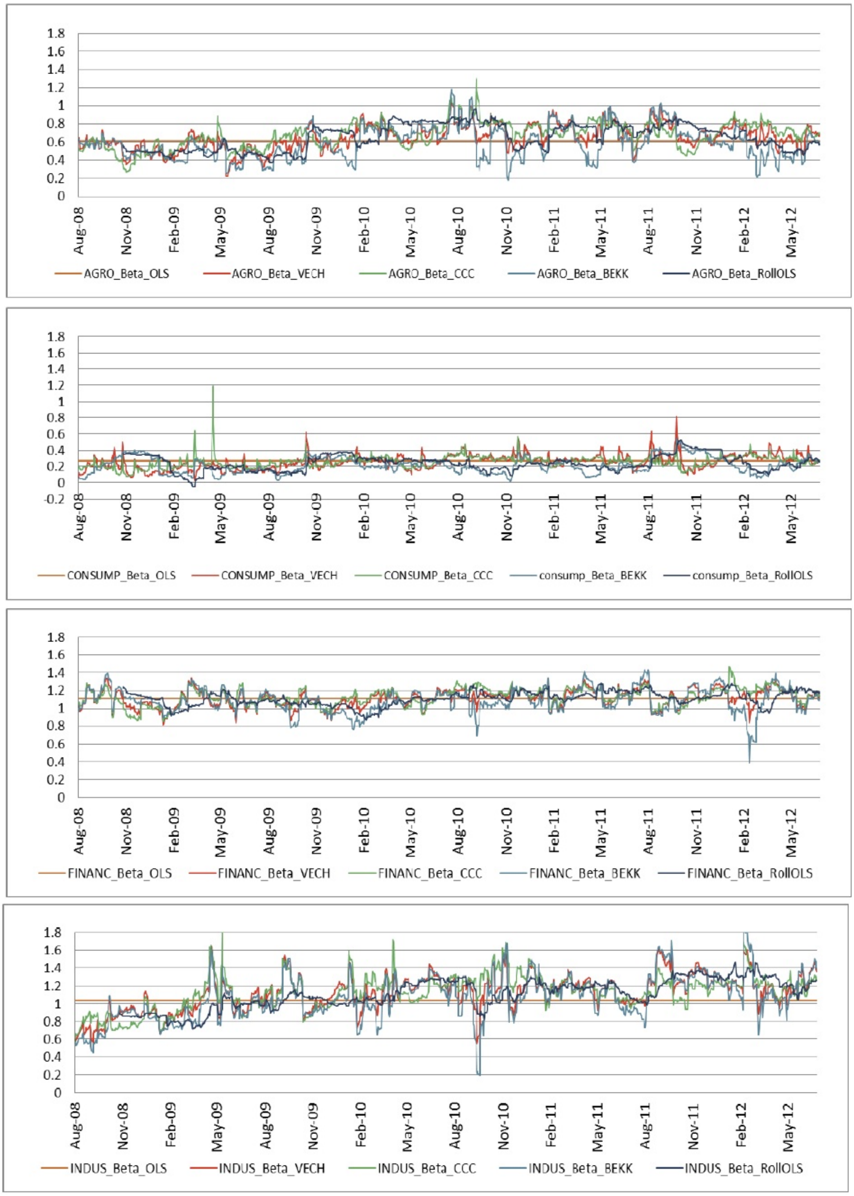


Fig. 5 Beta Plotted of AGRO, CONSUMP, FINANC and INDUS

trend of the particular industry. Moreover, Multivariate GARCH can also catch up the volatility quicker than rolling OLS as you can see from the graphs that the three models of Multivariate GARCH show the spike of the beta along the period while

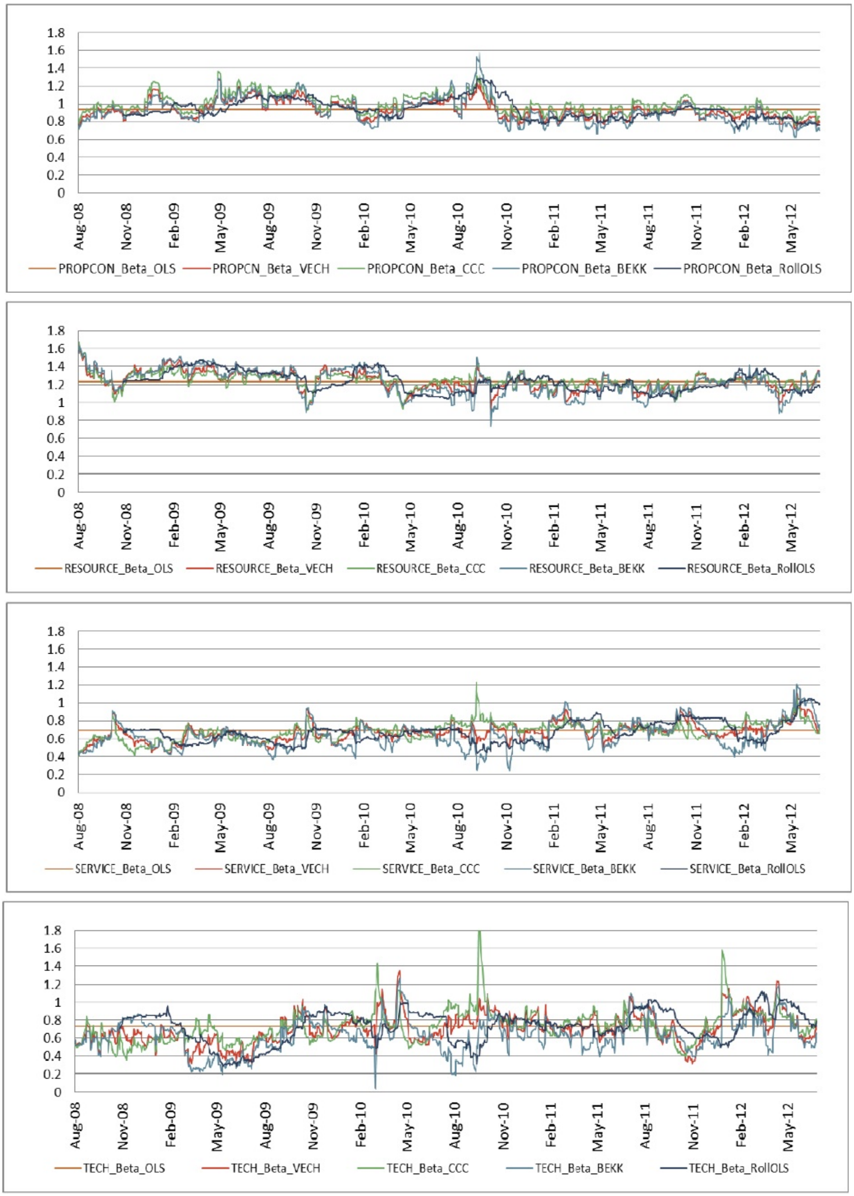


Fig. 6 Beta Plotted of PROPCON, RESOURCE, SERVICE and TECH

rolling OLS beta does not present large move of beta. In addition, from the plots, CCC model tend to provide highest volatile beta.

Finally, this study evaluates the performance of the models by calculating root mean square error and mean absolute error. The root mean square error is a square root of total quadratic error divided by the number of observation and mean square error or MAE is sum the absolute value of error divided by the number of observation. This study calculates RMSE and MAE of Ordinary Least Square, vech model Multivariate GARCH, BEKK model Multivariate GARCH and CCC model Multivariate GARCH and the results table can be illustrated as shown.

	Root Mean Squre Error (RMSE)				Mean Absolute Error (MAE)			
Index	OLS	VECH	CCC	BEKK	OLS	VECH	CCC	BEKK
AGRO	0.00864	0.00834	0.00839	0.00848	0.00642	0.00636	0.00643	0.00641
CONSUMP	0.00669	0.00666	0.00694	0.00674	0.00442	0.00440	0.00452	0.00442
FINANC	0.00781	0.00771	0.00770	0.00781	0.00582	0.00572	0.00571	0.00583
INDUS	0.01089	0.01064	0.01075	0.01078	0.00812	0.00788	0.00792	0.00792
PROPCON	0.00620	0.00618	0.00625	0.00621	0.00457	0.00454	0.00459	0.00455
RESOURCE	0.00659	0.00652	0.00656	0.00658	0.00495	0.00491	0.00493	0.00495
SERVICE	0.00662	0.00653	0.00660	0.00662	0.00499	0.00488	0.00493	0.00494
TECH	0.01142	0.01132	0.01154	0.01139	0.00814	0.00818	0.00822	0.00826

Fig. 7 Comparing RMSE and MAE of the models

From the table, in most industries, vech model provide least value of both RMSE and MAE which mean that this model is superior in Time-Varying betas. However, for CCC model and BEKK model, they provide the RMSE and MAE similar to RMSE and MAE of OLS that means they do not perform better than OLS. This may come from the restriction in this study that reduces the number of parameters that need to estimate in these two models. Then the ability of precisely forecast may drop.

4 Conclusion

The results prove that simple ordinary least square estimation exhibits heteroskedasticity problem. Therefore, the model that can handle with heteroskedasticity is GARCH model. The results of multivariate GARCH estimation in all three models that are vech model, BEKK model and CCC model show the pattern of time-varying in beta in all industry indexes. The results in all cases explain that one factor that influence in the current variance is its own variance in the past period. Moreover, the past period of the innovation or error terms also influence the conditional variance of the returns. Moreover, the study plotted the pattern of beta by these model compare to beta from Ordinary Least Square regression and rolling window OLS. The pattern of beta from these models of Multivariate GARCH in all industries moves up and down along the beta from OLS. In addition, using Multivariate GARCH models can capture the increasing or decreasing trend of systematic risk that occurs in some industries while OLS cannot. Multivariate GARCH model can capture the change in systematic risk quicker and better than rolling OLS and among these three models, CCC model present the most rapid move of beta. For asymmetries in covariance, we can observe from only vech model, the study expects to see negative and significant in coefficient $\delta_{i,m}$ that represent that covariance would be higher if both industry index and market index returns are negative compared to the case when they both are positive. However, even the results present that all cases coefficients are significant, but all coefficient show positive sign of $\delta_{i,m}$. Therefore, this study can conclude that there is no clear pattern of asymmetries in covariance of returns in these eight industries index. From the model evaluation, vech model performs the best since it provides lowest value of RMSE and MAE. Therefore, the vech model of Multivariate GARCH is the good choice for Time-Varying systematic risk estimation. Overall, the study found that by using Multivariate GARCH in estimating Time-Varying Systematic risk provide understanding in systematic risk precisely that is useful both academic and stock market investment and enhanced knowledge of Thailand Stock market. The further study objective can be scope of the data expanding since the establishment of ASEAN Trading Link. Therefore, the scope of data in further study can be ASEAN Stock market industry index, estimating the industry indexes in each country precisely can develop the opportunity in investment for market participants and improved knowledge in Thailand academic.

References

1. Bey, R., Pinches, G.: Additional evidence of heteroskedasticity in the market model. Journal of Financial and Quantitative Analysis 15, 299–322 (1980)
2. Black, F.: Capital market equilibrium with restricted borrowing. Journal of Business 20, 8–18 (1993)
3. Bollerslev, Engle, Wooldrige: A Capital Atset Pricing Model with Time-Varying Covariances. Journal of Political Economy 96, 116–131 (1988)
4. Bos, T., Newbold, P.: An empirical investigation of possibility of stochastic systematic risk in the market. Journal of Business 57, 35–41 (1984)
5. Choudhry, T., Wu, H.: Forcasting the Time Varying Beta of UK Companies GARCH method and Kalman filter method. The European Journal of Finance 15(4), 437–444 (2009)
6. Ender, W.: Applied Econometric Time Series, 3rd edn. John Wiley and Sons, Hoboken (2004)
7. Fabozzi, F., Francis, J.: Beta as a Random Coefficient. Journal of Financial and Quantitative Analysis 13, 101–116 (1978)
8. Grundy, K., Malkiel, B.G.: Reports of beta's death have been greatly exaggerated. Journal of Portfolio Management 15, 277–290 (1992)
9. Hinich, Melvin, Roll: Measuring Nonstationarity in parameters of market model. Research in Finance 1, 1–51 (1981)
10. Koutmos, G., Knif, J.: Estimative Systematic Risk Using Time Varying Distributions. European Financial Management 8, 59–73 (2002)
11. Lintner, J.: The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37 (1965)

12. Mergner, S., Bulla, J.: Time-varying Beta Risk of Pan-European Industry Portfolios: A Comparison of Alternative Modeling Techniques. The European Journal of Finance 14(8) (2008)
13. Mossin, J.: Equilibrium in a capital asset market. Econometrica 34, 768–783 (1966)
14. Sharpe, W.: Capital apset pricing: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442 (1965)

Forecasting Using Nonlinear Long Memory Models with Artificial Neural Network Expansion

Chaleampong Kongcharoen*

Abstract. We compare a number of models for obtaining h-step ahead minimum mean square error forecasts for nonlinear long memory processes. The forecasts from a proposed approximate nonlinear long memory, Fractionally Integrated Artificial Neural Network (*FI-ANN*) model, are compared to pure long memory models, e.g., $ARFIMA(1,d,0)$ and Local Whittle, pure nonlinear, i.e., Artificial Neural Network (*ANN*), and high order autoregressive model. Consider several nonlinear specifications in nonlinear long memory processes, the one- or two-step ahead forecasts of *FI-ANN* model generally perform better than *ANN* and other alternative models in the Monte Carlo simulation. The model is used to forecast series of inflation.

Keywords: Evaluating forecasts, Long memory time series, Neural networks, Nonlinear time series, Simulation.

1 Introduction

This paper considers prediction from a model that combines both long memory and nonlinear feature. The aim of the paper is to examine the forecast performance of Fractionally Integrated Artificial Neural Network (*FI-ANN*) model which incorporates both long memory and pure nonlinear characteristics approximated by Artificial Neural Network (*ANN*). In particular, we compare prediction performance from

Chaleampong Kongcharoen
Faculty of Economics,
Thammasat University, Bangkok, 10200, Thailand
e-mail: chaleampong@econ.tu.ac.th.

* The author is grateful to Richard Baillie and George Kapetanios for allowing me to use their *ARFIMA* gauss code, to P.H. Franses and D. Van Dijk for posting their gauss code for *ANN* estimation in their book's website and to seminar participants at the Michigan State University for helpful comments.

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 241–254.
DOI: 10.1007/978-3-642-35443-4_17

FI-ANN model, in term of Mean Square Errors (*MSEs*) with ones from pure long memory models, pure nonlinear model, and autoregressive model.

The long memory behavior is associated with hyperbolically decaying autocorrelations and impulse response weights. Empirical evidence on long memory was found in many economic and financial time series, see the reference in [4], and [24]. However, theoretical and simulation works, such as, [17] and [28], have addressed the problem that a linear process with structural break can mimic long memory process. On one hand, researchers, e.g., [33], develop a test to distinguish a true long memory process from a spurious one. On the other hand, researchers, e.g., [37] and [6] propose a model that can capture both long memory and structural break phenomena and found that many economic time series have both characteristics.

Few studies, e.g., [26], [14], [16], and [29], have considered forecasting performance from a model with both long memory and nonlinear features using actual time series data. Literature on forecasting with a combined model focuses on the parametric nonlinear model, such as, a Smooth Transition Autoregressive (*STAR*) model. However, as suggested in literature, the true specification of nonlinear part in a process is rarely unknown. The *ANN* model is widely used to approximate various nonlinear processes, see [18] for Monte Carlo simulation of the usefulness of *ANN* approximation in various nonlinear Data Generating Process (*DGP*). Recently, several works, e.g. [8] and [3] propose the flexible form for level of series to capture nonlinear part.[1].

This paper proposes a model that combines the long memory feature with an *ANN* expansion and investigates the usefulness of the model in term of long memory parameter estimation and forecasting a long memory process with nonlinear specification, such as, *STAR*, Self-Exciting Threshold Autoregressive (*SETAR*), and Markov-Switching (*MSW*). The main question is how well prediction from *FI-ANN* perform comparing to alternative models, such as, autoregressive, pure long memory (both parametric and semiparametric) and pure nonlinear (approximated by *ANN*) models.

From Monte Carlo simulation, we found that *FI-ANN* model generally provides lower biases of long memory parameter than semiparametric estimations especially with *SETAR* and *STAR* specification in the nonlinear part. Further, for one- and two-step head forecast, we found that *FI-ANN* model performs better than *ANN* model, pure long memory model and autoregressive model. There is also detailed examples of the methodology applied to monthly CPI inflation series.

2 Model and Method

2.1 Nonlinear-Long Memory Models

Long memory, or fractionally integrated processes were proposed by [20], and [25] to capture the slow hyperbolic rates of decay in an impulse response weight and

[1] There are also works that focus on the nonlinear in variance of series, e.g. [10]

autocorrelation of a series. A univariate process with fractional integration in its conditional mean is represented by

$$(1-L)^d y_t = u_t, \quad t = 1, ..., T, \tag{1}$$

where L is the lag operator and u_t is a short-memory, $I(0)$ process. Note that polynomial $(1-L)^d$ can be expressed in terms of its Binomial expansion, i.e., $(1-L)^d = \sum_{j=0}^{\infty} \binom{d}{j} = 1 - dL + \frac{d(d-1)L^2}{2!} + ...$ Then, y_t is said to be a fractionally integrated process of order d, $I(d)$. The degree of "long memory", or persistence is represented by the d parameter. The process is stationary if $d < 0.5$ and invertible if $d > -0.5$. Moreover, if $d > 0$ the process is associated with long memory. If the short memory is represented by a white noise process, then (1) becomes the fractional white noise model,

$$(1-L)^d y_t = \varepsilon_t, \quad t = 1, ..., T, \tag{2}$$

where $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2$, and $E(\varepsilon_t \varepsilon_s) = 0, s \neq t$. If $u_T t$ is $ARMA(p,q)$, i.e. $\phi(L)u_t = \theta(L)\varepsilon_t$, y_t becomes $ARFIMA(p,d,q)$. The infinite-order moving average representation is given by

$$y_t = \sum_{i=0}^{\infty} \psi_i(d)\varepsilon_{t-i}, \tag{3}$$

where $\psi(L) = 1 + \sum_{j=1}^{\infty} \psi_j L^j = (1-L)^d \theta(L)^{-1} \phi(L)$ and the infinite-order autoregressive representation is given by

$$y_t = \sum_{i=1}^{\infty} \pi_i(d) y_{t-i} + \varepsilon_t, \tag{4}$$

where $\pi(L) = 1 - \sum_{j=1}^{\infty} \pi_j L^j = \theta(L)^{-1}\phi(L)(1-L)^d$. For large i, the moving average and autoregressive coefficients decay at very slow hyperbolic rate, $\psi_k \approx c_1 k^{d-1}$ and $\pi_k \approx c_2 k^{-d-1}$ where c_1 and c_2 are constant ([4]).

The nonlinear process, such as, *MSW* and Threshold Autoregressive (*TAR*), can be mistaken for a long memory process [17, 28]. Some researchers try to propose the tests to distinguish the real long memory from the spurious one. See [34] and [9] for reviews about spurious long memory. Using Monte Carlo simulation, [11] conclude that, even a structural break process behaves similar to long memory process, it is dangerous to forecast the structural break series with the long memory model. However, this paper focuses on the other route supported by [6]'s results that many economic and financial time series have both nonlinear and long memory components. The useful model should combine both phenomena. Following Baillie and Kapetanios' strategy, this paper considers a nonlinear process as a short memory part in long memory model. The general nonlinear process is defined as

$$u_t = F(u_{t-1}, ..., u_{t-p}) + \varepsilon_t. \tag{5}$$

The nonlinear part can be a parametric nonlinear model, such as, the *STAR* model in [37] and [7]. The joint estimation between the long memory parameter and nonlinear parameters is advocated by appealing Monte Carlo and empirical results in [7]. However, the true form of nonlinear part is hardly known. The flexible functional form of nonlinear model by the *ANN* expansion is widely used in the literature and should be suitable in the joint estimation of long memory parameter and nonlinear part. The modeling of nonlinear part by nonparametric specification is proposed by [6].

[6] shows that the *MLE* of the *FI-ANN* model produces $T^{1/2}$-consistent and asymptotic normal for d parameter. In literature, the *ANN* model is usually used as the approximated model for nonlinear process, such as, smooth transition, Markov switching, or bilinear models. The estimation of the *ANN* model is based on misspecification models. The model and method of estimating *FI-ANN* model are discussed in following section. We conjecture that a long memory parameter from *FI-ANN* model have a nice asymptotic result. The usefulness of this model can be investigated by Monte Carlo simulation in following section. However, the formal proof of asymptotical result is beyond the scope of this paper.

2.2 Time Domain MLE

In this section we focus on the nonlinear part approximated by the *ANN* model. The *ANN* model is obtained by assuming that the conditional mean of u_t depends on the value of a linear combination of p lagged values $u_{t-1},...,u_{t-p}$. Moreover, in order to adequately capture the nonlinear relationship between u_t and lagged values, a hidden unit, $G(\cdot)$, is included up to q. Hence, the $ANN(p,q)$ part of the joint model is represented as

$$u_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j u_{t-j} + \sum_{j=1}^{q} \beta_j G\left(\gamma_{0j} + \sum_{i=1}^{p} \gamma_{ij} u_{t-i}\right) + \varepsilon_t. \tag{6}$$

This equation also includes the linear part, which is common in econometric application. The nonlinear part of Equation (6) can approximate any function to any degree of accuracy, provided that the number of nonlinear components q is sufficiently large ([18], p.208 and reference herein). We call the joint estimation of Equations (1) and (6) *FI-ANN*. By assuming that the white noise process ε_t is Gaussian, the $\theta = [d, \alpha_0, \alpha_j, \beta_j, \gamma_{0j}, \gamma_{ij}]$, where $i = 1,...,p, j = 1,...,q$ parameters can be estimated by minimizing the conditional sum of squared residuals (*CSS*) $\sum_{t=1}^{T} \tilde{\varepsilon}_t{}^2$, where $\tilde{\varepsilon}_t = \tilde{u}_t(d) - \alpha_0 + \sum_{j=1}^{p} \alpha_j \tilde{u}_{t-j}(d) + \sum_{j=1}^{q} \beta_j G\left(\gamma_{0j} + \sum_{i=1}^{p} \gamma_{ij} \tilde{u}_{t-i}(d)\right)$ and $\tilde{u}_t(d) = y_t - \sum_{l=0}^{t-k} \pi_l(d) y_{t-i} \approx (1-L)^d y_t$. Note that k is the truncation lag. In various settings, *CSS* is asymptotically equivalent to the approximate time domain *ML* estimation and easy to compute (See [15]).

The *ANN* model, generally, is considered as an approximate model for nonlinear processes, such as, *TAR*, *STAR*, or *MSW*, not the true *DGP*. Hence, we consider *SETAR*, *STAR*, and *MSW* as nonlinear part in our nonlinear long memory *DGP*.

As we mentioned, *FI-ANN* estimation from these *DGP* is inherently misspecified. However, according to [18], the properties of *CSS* estimation in misspecified model is well developed. We expect consistent estimates from the *CSS* estimation of the *FI-ANN* model. The small sample properties of the *FI-ANN* model will be investigated in following section.

Generally, the performance of the *FI-ANN* model depends heavily on selections of the function $G(\cdot)$, the number of hidden unit, q, and the number of lags, p. The logistic function is used almost invariably in applied literature. The appropriate choice of p and q is determined by a model selection criterion, such as, the Schwarz information criterion (*SIC*), where $p \in \{1,2,...,p*\}$ and $q \in \{1,2,...,q*\}$. An alternative strategy can be found in [30].

2.3 Semiparametric Estimations

The semiparametric estimation of the long memory parameter is one of growing literature, many estimation procedures are proposed and refined. See [31] for an overview about long memory estimations. [7] suggest the possibility of two-step estimation procedure for process with both long memory and break. Firstly, a semiparametric estimator for d parameter is estimated. Then, the nonlinear model is fitted to the fractional filtered series, $u_t = y_t - \sum_{t=1}^{t-p} \pi_l(\hat{d}) y_{t-l}$. However, an accuracy of second-step estimation relies heavily on the semiparametric estimate in first-step estimation. There is a wide variety of semiparametric methods for the estimation d and this study uses Local Whittle, which is know to have a desirable properties; see [32]. The Local Whittle semiparametric estimator for d is obtained by maximizing the objective function, $\ln\left[\frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I(\lambda_j)\right] - \frac{2d}{m}\sum_{j=1}^{m} \ln(\lambda_j)$. with respect to d, where $I(\lambda_j)$ is the periodogram given by $I(\lambda_j) = \frac{1}{2\pi T}|\sum_{j=1}^{T} y_t e^{i\lambda_j t}|^2$. The estimator depends on the choice of bandwidth, m. The discussion of optimal bandwidth or bias reduction which is beyond scope of this paper can be found in [2], [23], and [1]. For all simulation study in this study, the bandwidth is chosen as $m = \lfloor T^{0.5} \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x.

3 Simulation Study

In this section we investigate the finite sample performance of *FI-ANN*, $FI-TA$, *GPH*, and *LW*. In particular, the objective is to compare the biases and mean square errors of long memory parameter from *FI-ANN* and semiparametric estimations under the nonlinear long memory process.

3.1 Monte Carlo Setup

For each *DGP* we generate nonlinear process, u_t, then, we impose long memory characteristic by using moving average representation, $y_t = \sum_{l=0}^{t-l} \psi_l(d) u_l$. In order to obtain the sample size of 200, we generated 400 observations and dropped first

200 observations. Number of replication is 1,000. For long memory parameters, we investigate two main categories: (1) a pure structural break process ($d = 0$), and (2) a process with both structural break and long memory property ($d = 0.2$ and $d = 0.4$).

We consider three well-known nonlinear specifications which are *SETAR*, *STAR*, and *MSW*. The shifts between regimes in *SETAR* and *STAR* are observable and endogenous. However, the regime switching in *SETAR* relies on discontinuous function, while one in *STAR* is governed by continuous functions, e.g. exponential or logistic functions. On the contrary, the changes in *MSW* are created by unobservable variable. The detail and parameter values are outlined in following section.

3.1.1 Self-Exciting Threshold Autoregressive (*SETAR*) Process

We consider a simple *SETAR* with two regimes and $AR(1)$ in each regime, following [18], which is defined as

$$u_t = \begin{cases} \nu_{0,1} + \nu_{1,1} u_{t-1} + \varepsilon_t & \text{if } u_{t-1} \le c, \\ \nu_{0,2} + \nu_{1,2} u_{t-1} + \varepsilon_t & \text{if } u_{t-1} > c, \end{cases} \tag{7}$$

where ε_t is assumed to be i.i.d. white noise sequence conditional upon the history of the time series, i.e. $E(\varepsilon_t|\Omega_{t-1}) = 0$, $E(\varepsilon_t^2|\Omega_{t-1}) = \sigma^2$, and $\Omega_{t-1} = y_{t-1}, y_{t-2}, \ldots$. The delay parameter is set to be one. All simulations in this study, the error term, ε_t, is set as $N(0,1)$.

The stationary conditions of *SETAR* process are derived by [12] and [13]. In our simulation, we consider (1) the case that both regimes are stationary (experiment 1.1), and (2) the case that one regime is nonstationary (experiments 1.2 - 1.4). The parameter values are shown in Table 1.

3.1.2 Exponential Smooth Transition Autoregressive (*ESTAR*) Process

The *STAR* process with exponential transition function, following [6], is given by

$$u_t = \alpha_0 + \alpha_1 u_{t-1} + \beta_1 \left[1 - \exp(-\gamma_1 (u_{t-1} - \gamma_0)^2)\right] u_{t-1} + \varepsilon_t, \tag{8}$$

where γ_0 and γ_1 are threshold and smoothness parameters, respectively. The parameter values is presented in Table 2.

3.1.3 Markov Switching Process

We consider a simple two-state Markov switching process. The time varying parameters are governed by an unobservable random variable, s_t. Following [21], the Markov switching process with an $AR(1)$ process in each regime are defined as

$$u_t = \begin{cases} \nu_{0,1} + \nu_{1,1} u_{t-1} + \varepsilon_t & \text{if } s_t = 1, \\ \nu_{0,2} + \nu_{1,2} u_{t-1} + \varepsilon_t & \text{if } s_t = 2. \end{cases} \tag{9}$$

The process s_t is a Markov chain with a transition probability $\mathbf{P}$ given by the following matrix,

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix}. \tag{10}$$

A sufficient condition for geometric ergodicity of Markov switching autoregressive models is given by [38]. In our simulation, we set the parameter $v_{0,1} = 0.5$ and $v_{0,2} = -0.5$. We consider both low persistence, $p_{11} = p_{22} = 0.1$ (Experiments 3.1 and 3.2), and high persistence, $p_{11} = p_{22} = 0.9$ (Experiment 3.3 and 3.4) with moderate and high autoregressive parameters.

3.2 Long Memory Parameter Estimations

The performance of four estimations are presented in Tables 1 through 3 and evaluated by the average and Root Mean Square Errors (*RMSEs*) over the replications. For *FI-ANN* , optimal orders of p and q are determined by *SIC*. A selection of the simulations that were performed are reported in the text, and the results reported are typical of the general findings. Full details are available from the authors on request.

For the *FI-SETAR DGP*, the estimated d from *FI-ANN* are close to the true d parameter. There are substantial biases in the estimators from *LW* estimators. Please see Table 1. The bias of *LW* estimation in the $FI-SETAR$ *DGP* Experiment positively varies with degree of persistence in *DGP* process. Moreover, both semiparametric estimations perform unsatisfactorily in term of the *RMSEs*.

In the *FI-ESTAR* case, the estimated d from *FI-ANN* models are less bias than ones from semi parametric estimations in Experiments 2.3 and 2.4. In general, there are substantial biases for the estimators from the *LW* and *GPH* for the process with both structural break and long memory, i.e., $d > 0$. Please see Table 2.

In the case of $FI-MSW$, *FI-ANN* model performs slightly different. *FI-ANN* models do not suggest long memory in pure short memory *DGP* in Experiments 3.1 and 3.2 but *LW* model does. *FI-ANN* performs well only the case of high persistence process (Experiments 3.3. and 3.4). This result is consistent with the literature that *ANN* does not capture *MSW* process well, (see [37], p.242). The *LW* always give the positive and substantial biases for *FI-MSW DGP*. Please see Table 3.

We conclude from the results in Tables (1) (2) and (3) that applying the popular semiparametric estimations in the possibility of a structural break can provide an adverse result and misleading conclusion. On the contrary, the approximations both Taylor approximation and *ANN* expansion seems to be appealing in practice for long memory series with break and pure structural break especially the break that is induced by endogenous variables.

3.3 Forecasting

This subsection considers the prediction from *FI-ANN* model. From equation (1), $(1-L)^d y_t = u_t$, we can rewrite y_t as $y_t = \psi(L)u_t = (1-L)^{-d}u_t = \sum_{j=0}^{\infty} u_{t-j}$. Note that $u_t = F(u_{t-1}, u_{t-2}, ..., u_{t-p}) + \varepsilon_t$. The optimal h-step-ahead forecast of y_{t+h} at time t is given by

$$\hat{y}_{t+h|t} = E(y_{t+h}|\Omega_t), \tag{11}$$

where Ω_t is the information up to and including time t. The optimal one-step-ahead forecast is easily obtained as

$$\hat{y}_{t+1|t} = F(u_t, ..., u_{t+1-p}) + \psi_1 u_t + \psi_2 u_{t-1} + ... \tag{12}$$

As mentioned in literature, e.g., [37], [35], the derivation of optimal forecasts for period longer than one are more complicated. For example, two-step-ahead forecast can be derived as

$$\hat{y}_{t+2|t} = E(F(F(u_t, ..., u_{t+1-p}) + \varepsilon_{t+1}, u_t, u_{t-1}, ..., u_{t+2-p})|\Omega_t) + \psi_1 F(u_t) + \psi_2 u_t + ... \tag{13}$$

The conditional expectation can be exactly calculated by numerical integration. However, this method is time consuming (see [37], p.119-120). An alternative approach is to use Monte Carlo or Bootstrap method to approximate conditional expectation. Following [37], the two-step-ahead Monte Carlo forecast is given by

$$\hat{y}^{mc}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(F(u_t, ..., u_{t+1-p}) + \varepsilon_i, u_t, u_{t-1}, ..., u_{t+2-p}) + \psi_1 F(u_t) + \psi_2 u_t + ... \tag{14}$$

where k is large number and ε_i is drawn from assumed distribution of ε_{t+1}. Similarly, the two-step-ahead bootstrap forecast is obtained by

$$\hat{y}^{b}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(F(u_t, ..., u_{t+1-p}) + \hat{\varepsilon}_i, u_t, u_{t-1}, ..., u_{t+2-p}) + \psi_1 F(u_t) + \psi_2 u_t + ... \tag{15}$$

where $\hat{\varepsilon}_i$ is drawn from residuals in the estimated model. A practical method for calculating h-steps-ahead when $h \geq 2$ is outlined in [35]. The 'naive' forecasts, where ε_{t+h} are set at 0, are considered in our Monte Carlo simulation. In practice, predictors are calculated from T observations.

In Monte Carlo simulation, we compare performance of nonlinear long memory (*FI-ANN*) with pure long memory, pure nonlinear (*ANN*), higher order Autoregressive model, and $ARFIMA(1,d,0)$. For each iteration we simulate a series of 205 observations. The first 200 observations are used to estimate FI-ANN and alternative models.

The results for the mean square forecast errors (*MSFEs*) of the nonliear long memory process are reported in Table 4. For *FI-SETAR DGP*, the one- and two-step ahead forecasts from the *FI-ANN* model are generally better than *ANN* model and other alternatives. For the *FI-ESTAR DGP*, the forecasts from *FI-ANN* model perform better than ones from *ANN* model in short-period ahead predictions. For $FI-MSW$, the *FI-ANN* model gives the best forecasting among alternative models for high long memory parameter. The *FI-ANN* models perform worse than *ANN* models for the case of low long memory parameter and a pure MSW. However, the $MSFE$ from *FI-ANN* and *ANN* models are close.

4 Empirical Application

Inflation in many countries appear to be characterized by long memory with probable nonlinearity due to regime switching (e.g., [22], [5], and [8]. In this section, we focus on the inflation of G-7 countries, i.e. the U.S., Canada, the U.K, France, Germany, Japan, and Italy. The monthly inflation series are calculated from Consumer Price Index (CPI) obtained from the OECD statistical website. [8] suggested that there is an evidence that inflation for the U.S., Japan, and the U.K. have long memory and nonlinearity.

We separate data into estimation period and forecasting period. We employ the rolling forecast technique where the sample size is kept at 600 months and we repeat this procedure for 80 times. The one- to five-period ahead forecasts are compared to the actual data and the mean square errors from each estimation model are computed. The result is presented in Table 5. We find that the forecasts from *FI-ANN* perform close to *ARFIMA* and better than *ANN* for Canada, France, Germany, Japan and the U.K.. Thus, in case of G7 inflation, combining nonlinear in long memory model does not improve the forecasting performance. Moreover, the pure nonlinear model provide the inferior result. One striking result is the high order autoregressive model provides the superior performance for all countries except Canada and the U.S.

5 Conclusion

The prediction from *FI-ANN* models has been considered by comparing mean square forecast errors with *ANN* models and other alternative models. Using Monte Carlo simulation and several nonlinear specifications for nonlinear long memory process, we found that *FI-ANN* models generally work well and better than *ANN* models especially one- or two-step ahead forecasts. This result provides a useful suggestion for practitioners when they facing the series that may have both long memory and nonlinear structure. The performance of *FI-ANN* model is close to *ARFIMA* and suggested that nonlinear part does not help improving inflation forecasts for G7 countries.

Appendix

Table 1 Properties of the estimated long memory parameter for *FI-SETAR* Data Generating Process, $d = 0.4$

Experiment	Parameter values				Average of Estimated d		*RMSE* of estimated d	
	$v_{0,1}$	$v_{0,2}$	$v_{1,1}$	$v_{1,2}$	*FI-ANN*	*LW*	*FI-ANN*	*LW*
1.1	-0.3	0.1	-0.5	0.5	0.324	0.525	0.218	0.224
1.2	-0.5	0	1	0.1	0.407	0.876	0.268	0.552
1.3	-0.9	0	1	0.1	0.379	0.846	0.236	0.477
1.4	0.5	-0.5	1	1	0.474	1.096	0.413	0.707

Note: Number of observations is 200. Number of replications is 1,000.

Table 2 Properties of the estimated long memory parameter for *FI-ESTAR* Data Generating Process, $d = 0.4$

Experiment	Parameter values			Average of Estimated d		*RMSE* of estimated d	
	α_1	β_1	γ_1	*FI-ANN*	*LW*	*FI-ANN*	*LW*
2.1	0.8	-1.0	0.01	0.172	0.602	0.446	0.290
2.2	0.8	-1.0	0.05	0.092	0.496	0.506	0.224
2.3	1.3	-1.0	0.01	0.356	1.101	0.405	0.702
2.4	1.3	-1.0	0.05	0.361	1.068	0.238	0.687

Note: See Table 1.

Table 3 Properties of the estimated long memory parameter for *FI-MSW* Data Generating Process, $d = 0.4$

Experiment	Parameter values		Average of Estimated d		*RMSE* of estimated d	
	$v_{0,1} = -v_{0,2}$	$\pi_{11} = \pi_{22}$	*FI-ANN*	*LW*	*FI-ANN*	*LW*
3.1	0.5	0.01	0.222	0.667	0.283	0.311
3.2	0.9	0.05	0.330	0.707	0.167	0.343
3.3	0.5	0.01	0.266	0.684	0.436	0.339
3.4	0.9	0.05	0.426	0.826	0.230	0.473

Note: See Table 1.

Table 4 *MSEs* of Different Forecast Horizons, $d = 0.4$

Expe-riment	Forcast Horizon(s)	*AR*	*ARFIMA*	*LW*	*FI-ANN*	*ANN*	Expe-riment	Forcast Horizon(s)	*AR*	*ARFIMA*	*LW*	*FI-ANN*	*ANN*	Expe-riment	Forcast Horizon(s)	*AR*	*ARFIMA*	*LW*	*FI-ANN*	*ANN*
1.1	1	5.962	5.478	7.774	5.372	5.670	2.1	1	1.156	1.007	1.759	1.006	1.028	3.1	1	1.737	2.844	1.708	1.512	1.553
	2	6.184	5.571	5.899	5.479	5.697		2	2.711	2.417	3.367	2.432	2.439		2	1.835	3.441	1.847	1.659	1.764
	3	7.260	6.308	6.758	6.136	6.654		3	4.107	3.675	4.567	3.723	3.696		3	1.767	3.592	1.849	1.598	1.765
	4	7.084	6.505	6.897	6.309	6.757		4	5.300	4.820	5.530	4.928	4.841		4	1.929	3.953	1.906	1.753	1.920
	5	6.805	6.150	6.633	5.940	6.339		5	6.280	5.871	6.378	6.022	5.939		5	1.839	3.986	1.835	1.656	1.809
1.2	1	8.616	7.777	10.631	7.200	7.614	2.2	1	1.213	1.111	1.771	1.132	1.139	3.2	1	3.428	18.354	6.406	3.086	3.642
	2	20.295	17.796	23.265	17.167	18.348		2	2.411	2.084	2.769	2.191	2.127		2	4.525	21.286	9.096	4.391	5.072
	3	31.689	27.911	35.929	28.084	29.133		3	3.205	2.913	3.535	3.109	2.977		3	4.328	20.975	5.903	4.368	4.926
	4	44.872	39.861	50.115	41.608	43.152		4	3.831	3.501	4.004	3.787	3.605		4	4.488	19.307	4.614	4.531	4.687
	5	58.211	52.052	63.063	55.032	76.425		5	4.314	4.049	4.419	4.301	4.150		5	5.505	17.884	6.575	5.185	5.443
1.3	1	11.153	11.855	10.721	8.557	9.925	2.3	1	1.257	1.356	1.147	1.190	1.179	3.3	1	1.738	1.658	1.752	1.647	1.612
	2	21.996	25.020	21.055	17.855	20.407		2	2.565	3.174	2.385	2.444	2.479		2	2.272	2.225	2.177	2.211	2.066
	3	30.929	38.531	30.233	26.407	29.618		3	3.756	5.106	3.519	3.640	3.714		3	3.156	3.139	3.011	3.243	2.921
	4	37.105	47.769	35.105	34.021	40.244		4	4.872	7.307	4.703	4.808	5.070		4	3.706	3.497	3.308	4.092	3.324
	5	44.331	58.926	41.224	42.283	119.134		5	6.045	9.466	5.797	6.121	6.450		5	3.957	3.969	3.711	5.951	3.682
1.4	1	6.232	5.702	6.413	5.705	5.916	2.4	1	1.275	1.154	1.249	1.147	1.222	3.4	1	6.149	5.693	10.644	5.735	6.047
	2	16.577	15.252	17.065	15.356	15.829		2	2.867	2.563	2.728	2.591	2.684		2	7.450	6.916	8.374	6.955	7.048
	3	31.036	29.132	31.388	29.606	29.603		3	4.945	4.372	4.574	4.410	4.511		3	13.210	12.247	16.199	12.363	12.633
	4	47.771	46.075	47.541	48.177	47.230		4	7.296	6.651	6.878	6.750	6.779		4	15.969	15.005	16.349	15.277	15.540
	5	66.244	65.136	65.989	71.291	71.684		5	9.807	8.945	9.187	9.358	9.076		5	20.910	19.890	21.748	20.669	25.526

Note: See Table 1 and naive forecasts are reported for *FI-ANN* and *ANN*.

Table 5 MSEs of different forecast horizons, G7 CPI inflation

Country	Forecast Horizon(s)	*AR*	*ARFIMA*	*LW*	*FI-ANN*	*ANN*
Canada	1	0.159	0.163	0.220	0.164	0.160
	2	0.160	0.165	0.227	0.166	0.168
	3	0.160	0.172	0.259	0.174	0.180
	4	0.168	0.181	0.300	0.184	0.195
	5	0.161	0.177	0.314	0.180	0.194
France	1	0.071	0.097	0.154	0.101	0.113
	2	0.074	0.097	0.147	0.099	0.118
	3	0.072	0.092	0.150	0.094	0.118
	4	0.073	0.093	0.152	0.096	0.129
	5	0.076	0.096	0.189	0.100	0.140
Germany	1	0.091	0.127	0.254	0.128	0.139
	2	0.087	0.108	0.148	0.108	0.115
	3	0.088	0.110	0.168	0.111	0.118
	4	0.085	0.111	0.148	0.110	0.116
	5	0.086	0.116	0.155	0.115	0.121
Italy	1	0.032	0.033	0.050	0.034	0.040
	2	0.035	0.036	0.054	0.036	0.049
	3	0.041	0.040	0.061	0.042	0.064
	4	0.045	0.044	0.078	0.046	0.080
	5	0.045	0.045	0.089	0.048	0.092
Japan	1	0.086	0.090	0.115	0.103	0.138
	2	0.089	0.093	0.124	0.104	0.162
	3	0.089	0.097	0.131	0.107	0.178
	4	0.089	0.095	0.119	0.120	0.180
	5	0.089	0.095	0.109	0.854	0.180
U.K.	1	0.088	0.151	0.224	0.151	0.164
	2	0.092	0.141	0.178	0.140	0.154
	3	0.095	0.145	0.203	0.145	0.167
	4	0.094	0.139	0.172	0.137	0.165
	5	0.098	0.148	0.206	0.148	0.181
U.S.	1	0.209	0.203	0.199	0.204	0.199
	2	0.280	0.273	0.324	0.274	0.276
	3	0.295	0.287	0.370	0.289	0.287
	4	0.296	0.285	0.381	0.288	0.285
	5	0.295	0.282	0.396	0.285	0.278

References

1. Andrews, D.W.K., Guggenberger, P.: A bias–reduced log–periodogram regression estimator for the long–memory parameter. Econometrica 71, 675–712 (2003)
2. Andrews, D.W.K., Sun, Y.: Adaptive local polynomial whittle estimation of long-range dependence. Econometrica 72, 569–614 (2004)

3. Aladag, C.H., Egrioglu, E., Kadilar, C.: Improvement in Forecasting Accuracy Using the Hybrid Model of ARFIMA and Feed Forward Neural Network. American Journal of Intelligent Systems 2, 12–17 (2012)
4. Baillie, R.T.: Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59 (1996)
5. Baillie, R.T., Han, Y.W., Kwon, T.G.: Further long memory properties of inflationary shocks. Southern Economic Journal 68, 496–510 (2002)
6. Baillie, R.T., Kapetanios, G.: Testing for neglected nonlinearity in long-memory models. Journal of Business & Economic Statistics 25, 447–461 (2007)
7. Baillie, R.T., Kapetanios, G.: Nonlinear models for strongly dependent processes with financial applications. Journal of Econometrics 147, 60–71 (2008)
8. Baillie, R.T., Morana, C.: Adaptive ARFIMA models with Application to Inflation. Economic Modelling 29, 2451–2459 (2012)
9. Banerjee, A., Urga, G.: Modelling structural breaks, long memory and stock market volatility: an overview. Journal of Econometrics 129, 1–34 (2005)
10. Bildirici, M., Ersin, O.: Nonlinear volatility models in economics: smooth transition and neural network augmented GARCH, APGARCH, FIGARCH, and FIAPGARCH models, MPRA Paper 40330, University Library of Munich (2012)
11. Bisaglia, L., Gerolimetto, M.: Forecasting long memory time series when occasional breaks occur. Economics Letters 98, 253–258 (2008)
12. Chan, K.S., Tong, H.: On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. Advances in Applied Probability 17, 666–678 (1985)
13. Chan, K.S., Petruccelli, J.D., Tong, H., Woolford, S.W.: A multiple-threshold ar(1) model. Journal of Applied Probability 22, 267–279 (1985)
14. Chung, S.-K.: The Out-of-sample Forecasts of Nonlinear Long-memory Models of the Real Exchange Rate. International Journal of Finance and Economics 11, 355–370 (2006)
15. Chung, C.-F., Baillie, R.T.: Small sample bias in conditional sum-of-squares estimators of fractionally integrated arma models. Empirical Economics 18, 791–806 (1993)
16. Clements, M.P., Franses, P.H., Swanson, N.R.: Forecasting economic and financial time-series with non-linear model. International Journal of Forecasting 20, 169–189 (2004)
17. Diebold, F.X., Inoue, A.: Long memory and regime switching. Journal of Econometrics 105, 131–159 (2001)
18. Franses, P.H., van Dijk, D.: Non-linear time series models in empirical finance. Cambridge University Press (2000)
19. Geweke, J., Porter-Hudak, S.: The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238 (1983)
20. Granger, C., Joyeux, R.: An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–39 (1980)
21. Hamilton, J.D.: A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384 (1989)
22. Hassler, U., Wolters, J.: Long memory in inflation rates: International evidence. Journal of Business & Economic Statistics 13, 37–45 (1995)
23. Henry, M., Robinson, P.M.: Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence. In: Robinson, P., Rosenblatt, M. (eds.) Athens Conference on Applied Probability and Time Series Analysis, Time Series Analysis in Memory of E.J. Hannan. Lecture Notes in Statistics, vol. II, 115, pp. 220–232. Springer, Berlin (1996)
24. Henry, M., Zaffaroni, P.: The Long-Range Dependence Paradigm for Macroeconomics and Finance. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (eds.) Long Range Dependence: Theory and Applications, pp. 417–438. Birkhauser, New York (2003)

25. Hosking, J.: Fractional differencing. Biometrika 65, 165–176 (1981)
26. Hyung, N., Franses, P.H.: Forecasting Time Series with Long Memory and Level Shifts. Journal of Forecasting 24, 1–16 (2005)
27. Kunsch, H.R.: Statistical aspects of self-similar processes. In: Yu, P., Sazanov, V.V. (eds.) Proceedings of the First World Congress of the Bernoulli Society, pp. 67–74. VNU Science Press, Utrecht (1987)
28. Kuswanto, H., Sibbertsen, P.: A study on "spurious long memory in nonlinear time series models". Applied Mathematical Sciences 2, 2713–2734 (2008)
29. Lahiani, A., Scaillet, O.: Testing for threshold effect in ARFIMA models: Application to US unemployment rate data. International Journal of Forecasting 25, 418–428 (2009)
30. Medeiros, M.C., Terasvirta, T., Rech, G.: Building neural network models for time series: A statistical approach. Journal of Forecasting 25, 49–75 (2006)
31. Nielsen, M.O., Frederiksen, P.H.: Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. Econometric Reviews 24, 405–443 (2005)
32. Robinson, P.M.: Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23, 1630–1661 (1995)
33. Shimotsu, K.: Simple (but effective) tests of long memory versus structural breaks. Working Papers 1101, Queen's University, Department of Economics (December 2006)
34. Sibbertsen, P.: Long-memory versus structural change: An overview. Statistical Papers 24, 465–515 (2004)
35. Teräsvirta, T.: Forecasting economic variables with nonlinear models. In: Elliott, G., Granger, C., Timmermann, A. (eds.) Handbook of Economic Forecasting, pp. 413–457. Elsevier (2006)
36. Teräsvirta, T., Lin, C., Granger, C.: Power of the neural network linearity test. Journal of Time Series Analysis 14, 209–220 (1993)
37. van Dijk, D., Franses, P.H., Paap, R.: A nonlinear long memory model, with application to us unemployment. Journal of Econometrics 110, 135–165 (2002)
38. Yao, J.F., Attali, J.G.: On stability of nonlinear AR processes with Markov switching. Advances in Applied Probability 32, 394–407 (2000)

Modeling Dependency of Crude oil Price and Agricultural Commodity Prices: A Pairwise Copulas Approach

Phattanan Boonyanuphong, Songsak Sriboonchitta, and Chukiat Chaiboonsri

Abstract. This study examines the dependency between the return of crude oil future prices and the agricultural commodity future prices as well as provides flexible models for dependency and the conditional volatility GARCH. Therefore, this paper used copula-based GARCH models, which consists in estimating the marginal distributions of the return of the crude oil price and agricultural commodity prices and then estimates the copula parameters by static and time-varying copula models. The results revealed that the co-movement between crude oil price and agricultural commodity prices are generally strong and there exists symmetric tail dependence between crude oil and agricultural commodity prices in all pairs. However, its tail dependence is relatively weak. The dependence parameters are very volatile over time and deviate from their constant levels. Our findings have important implications for policy makers, producers and traders, which could be used to implement a better policy to optimize and stabilize the markets or their portfolio management in the agricultural commodity markets.

1 Introduction

In recent years, the oscillation of agricultural commodity prices in the international market has created a widespread debate. The increase in agricultural commodity prices started sharply in 2007; the price had grown to its highest level in thirty

Phattanan Boonyanuphong · Songsak Sriboonchitta · Chukiat Chaiboonsri
Faculty of Economics,
Chiang Mai University,
Chiang Mai 50200 Thailand
e-mail: bphattanan@bauga.pn.psu.ac.th
{songsak,chukiat}@econ.cmu.ac.th

Phattanan Boonyanuphong
Department of Social Sciences,
Prince of Songkla University, Pattani 94000 Thailand

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 255–267.
DOI: 10.1007/978-3-642-35443-4_18

years by June 2008. The main agricultural commodity prices such as corn, soybeans, wheat and rice rose up to 75 percent, 116 percent, 78 percent and 166 percent, respectively. After that, commodity prices fell sharply in the next six months. However, the prices of agricultural commodities are still above their historic levels [1, 2]. In 2010, the prices began rising again and peaked in early 2011. In the meantime, the prices of commodity energy experienced a high level of fluctuations which has been unprecedented in recent history. For instance, the crude oil prices began increasing in 2004 and reached their highest price in 2008 and then fell steadily in late 2008. However, these prices are still higher than the price levels in 2004 [3]. And they increased steeply again to nearly 100 dollars per barrel in 2010.

The co-movement between agricultural commodity prices and crude oil price has attracted the interest of many researchers to investigate what factors drive this co-movement. There are several ways to examine the transmission from the crude oil price to agricultural commodity prices. First, the energy-agriculture linkage; the rise in oil prices causes an increase in agricultural commodity prices by driving costs of production through its impact on chemicals, fertilizers, and other inputs. Hanson [4] showed that soaring crude oil prices drive higher costs of production which in turn cause agricultural commodity prices to increase. Also, Baffes [5] found evidence that the fertilizer index has strongly impacted the pass-through from the energy prices to non-energy prices as followed by agriculture. Moreover, the close linkage between energy and agricultural markets, which is the result of the production of biofuels that has surged since 2006, has altered prices also. Ethanol and biodiesel are substitutes for gasoline and diesel, thereby creating a higher demand for agricultural commodities in the biofuels' industry in combination with the high oil prices. Tang and Xiong [6] reveal that the increase in the biofuels industry has probably caused prices of grains and oil seeds such as corn, soybeans and wheat to co-move with oil prices. Chen et al. [7] investigated the interaction between the prices of corn, soybean and wheat and the crude oil price. They argue that the demanding growth of grain production based on biofuels is significantly due to higher crude oil prices. Second, the indirect effect of energy price on agricultural commodity prices through the depreciation/appreciation of the exchange rate results in an increase in agricultural prices [8].

Most empirical studies focus on investigating the relationship between the crude oil price and agricultural commodity prices. Chang and Su [9] point out that the substitutive effect can be represented in the period of the high crude oil price due to the increasing use of biofuels. Chen, et al. [7] confirm that the increase in the crude oil price has significantly affected world agricultural grain production and its prices. Also, Nazlioglu and Soytas [8] examined the dynamic relationship between world oil prices and twenty four world agricultural commodity prices by employing panel co-integration and Granger causality methods. The empirical results revealed that changes in world oil prices significantly resulted in the prices of several agricultural commodities. In contrast, Yu, Bessler and Fuller [10] analyzed the long-run interdependence between the crude oil price and four major traded edible oil prices including soybean, sunflower, rapeseed and palm oils and could not detect an influence of crude oil price on edible oils prices over the period from January 1999 to March 2006. More recently, Gilbert [11] and Saban [2] found that the crude oil prices and the agricultural commodity prices did not affect each other by linear causality. In addition, they argue that macroeconomic and financial factors were seen as the main determinants of changes in overall agricultural prices. Therefore, the discussion above shows that the analysiss of the relationship between oil price and agricultural commodity prices are widely differing and complex in their empirical evidence.

There are numerous methods that have been used to explore the co-movement between random variables. For example, the Johansen cointegration and Vector Error Correction Models which was employed by Natanelov et al. [12] Cifarellli and Paladino [13] used a multivariate CCC-GARCH-M model. Modeling time-varying volatility such as multivariate GARCH models were used by Chang et al. [14] However, this model which is based on some strong assumptions was often used in order to obtain a desirable variance-covariance matrix. Furthermore, the VAR model and multivariate GARCH models were assumed to have a linear relationship with multivariate normal distribution or student-t [15]. These assumptions may be considered as strong assumptions in empirical studies. In several data sets it was found that the data were skewed, fat-tailed and leptokurtic with dissimilar marginal distribution, as well as the distributions had degrees of freedom need that were not the same for each marginal distribution. In addition, previous empirical research involving the issue of co-movement between crude oil prices and the agricultural commodity prices mostly used the co-integration theory, the Granger causality test, the vector autoregressive model and the vector error correction.

In this research, we attempt to fill the gaps and handle the drawbacks of the traditional models by the Copula-based GARCH model. This model provides for better flexibility in joint distributions as well as transformation invariant correlations; hence, it does not need to assume linear correlation. More specifically, we try to answer three questions: What is the dependence structure between crude oil price and agricultural commodity prices? Is the dependence symmetric or asymmetric? Is there existence extreme tail dependence between crude oil and agricultural commodity markets? By answering these questions, we hope to enhance the understanding of the dependence structure between the crude oil price and agricultural commodity prices.

The rest of the paper is structured as follows. In Section 2 we provide a brief review of the copula model, marginal model and the estimation method used for this paper. In Section 3 we describe the data and in Section 4 and 5 we discuss our results and include the discussion about the co-movement between crude oil price and agricultural commodity prices, and to be finally followed by conclusions in Section 6.

2 Econometrics Models

2.1 Copula Models

Here we consider both static copulas and time varying copulas models to describe the different tail dependence structure. Static copulas models used in our work include Gaussian copula, Student-t copula, Frank copula, Clayton copula and Gumbel copula.

The bivariate Gaussian copula is given by $C_{Gaus}(u,v;\rho) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$, where the variables u and v are CDFs or ECDFs of the standardized residuals from the marginal models, and $0 \le u, v \le 1$ and where $\Phi^{-1}(u)$ and $\Phi^{-1}(v)$ are standard normal quantile functions. Gaussian copula cannot capture tail dependence. Similarly, the Student-t copula is defined by $C_T(u,v;\rho,\nu) = T_{\nu,\rho}(t_\nu^{-1}(u), t_\nu^{-1}(v))$, where $T_{\nu,\rho}$ is the bivariate student-t distribution with degrees of freedom ν and the linear correlation coefficient ρ and $t_\nu^{-1}(u)$ and $t_\nu^{-1}(v)$ are the quantile function of student-t distribution. The Student-t copula provides symmetric structures non-zero tail dependence with the same probability of occurrence, λ_U, $\lambda_L > 0$ both positive and negative side. Moreover, we also choose non-elliptical copulas such as Frank, Clayton, and Gumbel which consider negative dependence and upper (lower) tail dependence. The Frank copula is defined by $C_F(u,v;\theta) = \frac{-1}{\theta}\ln(1 + ((e^{-\theta u} - 1)(e^{-\theta v} - 1))/(e^{-\theta} - 1))$, where $\theta \in (-\infty, +\infty)$. The Frank copula permits negative dependence between the marginal distributions. Similar to elliptical copulas, dependence is symmetric in both tails and also it is most appropriate for data that exhibit weak tail dependence. The Clayton copula is defined as follows $C_{CL}(u,v;\theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$, where $\theta \in (0,\infty)$. The Clayton copula exhibits strong lower tail dependence. On the contrary, the Gumbel copula can catch upper tail dependence. The Gumbel is defined by $C_G(u,v;\theta) = exp(-((-\ln(u))^{1/\theta} + (-\ln(v))^{1/\theta})^{\theta})$, where $\theta \in [1,\infty)$.

In case of time varying dependence, the alternative way to capture time variation is to use a regime switching copula function. In Patton [16], they proposed the upper and lower tail dependence parameters to follow ARMA(p,q) process. The following are some time varying copula candidates. By following Patton [16], the time varying Gaussian copula can be defined as:

$$\rho_t = \Lambda(\psi_0 + \psi_1\rho_{t-1} + \cdots + \psi_n\rho_{t-p} + \gamma_0\frac{1}{q}\Sigma_{j=1}^{q}\Phi^{-1}(u_{t-j})\Phi^{-1}(v_{t-j})) \quad (1)$$

where $\Lambda = (1-e^{-x})(1+e^{-x})^{-1}$ is a logistic transformation which is to keep the correlation coefficient, ρ_t, belonging to (-1,1) at all times. For the Student-t copula with time varying, $\Phi^{-1}(x)$ is replaced by $t_\nu^{-1}(x)$.

Time varying Clayton copula and Time varying Gumbel copula also assumed the tail dependence parameters to follow ARMA(p,q) process. We propose the time varying Clayton and Gumbel copulas are as following:

$$\tau_t = \Lambda(\psi_C + \psi_{C1}\tau_{t-1} + \cdots + \psi_{Cn}\tau_{t-p} + \gamma_C \frac{1}{q}\Sigma_{j=1}^{q}|u_{t-j} - v_{t-j}|), \quad (2)$$

$$\delta_t = \Lambda(\psi_G + \psi_{G1}\delta_{t-1} + \cdots + \psi_{Gn}\delta_{t-p} + \gamma_G \frac{1}{q}\Sigma_{j=1}^{q}|u_{t-j} - v_{t-j}|) \quad (3)$$

where Λ is the logistic transformation to ensure that τ_t and δ_t within the (0,1) interval.

2.2 Marginal Models

The first step in estimating the parameters of the copula model is to estimate the marginal models. In this study, the model used for the marginal distribution is ARMA (p,q)-GARCH(1,1) and standardized residuals satisfying student-t distribution. The ARMA (p,q)-GARCH(1,1) model is a common approach to model time series with the fat-tail and conditional heteroskedastic errors. The models employed for the marginal distributions are followed from Ling [17]and we will denote the log-difference of the crude oil future price or the agricultural commodity future prices as the variable r_t. Thus, the marginal model for the crude oil future prices or the agricultural commodity future prices, r_t, can be formed as:

$$r_t = \omega + \Sigma_{i=1}^{p}\phi_i r_{t-1} + \Sigma_{i=1}^{q}\psi_i \varepsilon_{t-i} + \varepsilon_t, \quad (4)$$

$$\varepsilon_t = \eta_t\sqrt{h_t} \text{ and } h_t = \alpha_0 + \Sigma_{i=1}^{r}\alpha_i\varepsilon_{t-i}^2 + \alpha_0 + \Sigma_{i=1}^{q}\beta_i h_{t-i}, \quad (5)$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\Sigma_{i=1}^{k}\alpha_i + \Sigma_{i=1}^{l}\beta_i < 1$. η_t is the standardized residual, which can be assumed for any distribution. In general, we assume that it is Gaussian, student-t or skewed-t distribution. For example, the Student-t innovation was focused by Bollerslev and Wooldridge [18], while Hansen [19] modeled skewed Student-t distribution.

The correctness and usefulness of the joint copula model is important. If the marginal distributions are an invalid specification, then the copula will also be invalid in the joint copula model. For performance with misspecification, we used the Lagrange multiplier tests and the Kolmogorov-Smirnov test to confirm the empirical adequacy of the marginal models. (Discuss in detail in section 4.1)

2.3 Estimation Method

The set of unknown parameters of the copula function can be estimated by the full maximum likelihood (FML), the inference function for margins (IFM) and the canonical maximum likelihood method (CML). The FML and IFM are two parametric estimation methods. They might have shortcomings in estimating θ is that they are likely to be inconsistent if they misspecified the assumption of marginal distribution. The CML is semiparametric method, which marginal distributions are estimated without specifying the marginal, and then use a parametric model for estimating the parameter of the copula. We use two steps of CML for estimating the parameterθ of a copula (see Cherubini, Luciano and Vecchiato [20]). In the first step, estimating the marginals $\hat{u}_t$and $\hat{v}_t$ using the empirical distributions of $\hat{X}_t$ and $\hat{Y}_t$. Then, estimate via MLE estimators the copula parameters θ

$$\hat{\theta} = argmax\Sigma_{t=1}^{N} \ln c(\hat{u}, \hat{v}; \theta) \tag{6}$$

To evaluate the best fit of the different copula models are the following ways: (1) using the Akaike Information Criterion (AIC) and Bayesian information criteria (BIC) are followed from Brechmann [21] to measure the goodness of fit and the copula would be selected with smallest AIC and/or BIC; and (2) using the test based the empirical copula to perform the goodness of fit of the copulas, which calculates the Cramer-von Mises statistics as well as the according p-values using bootstrapping [22]. The goodness-of-fit tests based the empirical copula process are comparing a "distance" between C_n and an estimation C_{θ_n} of C under null hypothesis, $H_0 : C \in \{C_\theta\}$ The formulas are following as: $C_n = \sqrt{n}(C_n - C_{\theta_n})$ where C_n is the distribution of empirical copula function, and $C_{\theta_n}$is the distribution of copula function C_θ.

3 Data

We used daily data from February 28, 2008 to December 15, 2011 from three energy crops future prices: corn, soybean oil, and palm oil. And for crude oil prices we employed the Brent future prices as a proxy for the world oil prices. We chose data from February 28, 2008 to December 15, 2011 because their joint swings had greater intensity than in the previous period. The data for the agricultural commodity and crude oil future prices are compiled from the EcoWin database.

Table 1 Descriptive statistics for daily oil price and agricultural commodity returns

	Mean	Max	Min	Std.Dev.	Skewness	Kurtosis	Jarque-Bera	Prob
Brent	-7.0e-06	0.1271	-0.1095	0.026	-0.1428	5.8766	340.8683	0.0000
Corn	4.1e-05	0.1276	-0.1041	0.0237	-0.0292	4.7367	123.1705	0.0000
Soybean oil	-0.0003	0.0740	-0.0777	0.0187	-0.0813	5.099	180.8011	0.0000
Palm oil	-0.0003	0.0976	-0.1104	0.0208	-0.2486	6.7960	597.8861	0.0000

The descriptive statistics for crude oil and agricultural commodity daily futures returns are reported in Tables 2. The skewness statistics show that most return series are obviously skewed. With respect to the excess kurtosis statistics, data reveals that all return series are highly leptokurtic with respect to the normal distribution. Similarly, the Jarque-Bera statistics are significant, implying that the distributions are both leptokurtic and fat tails.

4 Empirical Results

4.1 Results for the Marginal Models

Table 2 presents the marginal distribution of crude oil return and agricultural commodity returns corn, soybean oil and palm oil. We found that the ARMA(0,2)-GARCH(1,1) specification is the best model for corn, soybean oil and palm oil, respectively. For crude oil return, however, the proper marginal distribution is the ARMA(0,4)-GARCH(1,1). The empirical evidence presented in Table 2 shows that all parameters for variance equation are highly significant. The GARCH term, β is strong as it is in the range of 0.92 to 0.94. Moreover, it satisfies the assumption of convergence, which appears that, $\alpha+\beta$ is close to one for all of the series. The degrees of freedom of the Student-t distribution were significant and ranging from 6.65 to 9.38, implying that the error terms are not normal. In addition, there are no ARCH effects in the residuals.

Table 2 Parameter estimates for the marginal distribution models

	ma1	ma2	ω	α	β	ϑ	LL	ARLM
Brent	-0.0633	-0.016	1.0e-05	0.0570	0.9343	9.3867	2340.0329	0.8477
	(0.001)	(9.1e-04)	(1.3e-11)	(3.6e-04)	(5.1e-04)	(6.961)		
Corn	0.0231	-0.0097	1.0e-05	0.0537	0.9204	6.6567	2325.8748	0.9598
	(9.7e-04)	(6.5e-04)	(1.8e-10)	(4.3e-04)	(0.001)	(1.894)		
Soybean oil	0.0081	0.0156	2.0e-06	0.0488	0.9440	8.6841	2626.0594	0.7818
	(0.001)	(0.001)	(1.7e-12)	(1.8e-04)	(2.2e-04)	(6.632)		
palm oil	-0.0114	0.0920	2.0e-06	0.0602	0.9336	7.7032	2602.7955	0.5853
	(0.001)	(0.001)	(1.7e-12)	(2.2e-04)	(2.3e-04)	(3.179)		

Note: The numbers in parentheses are standard deviations.

The correct specifications of marginal distributions are crucially important to the estimation in the copula model. Diebold, Gunther and Tay [23] argued that if the marginal distributions are correctly specified, then the probability transformations should be i.i.d. uniform (0,1). If we use incorrect marginal distribution models, then their probability transformation will not i.i.d. uniform (0,1) and the copula will automatically be misspecified. We followed Patton [16] by using two steps; the Lagrange Multiplier (LM) test for serial independence of the probability integral transformations and the Kolmogorov-Simirnov (K-S) test of density specification to evaluate the marginal distribution assumption. First, we used the LM test to examine the independence of the first four moments of variables u_t and v_t. We regress $(\hat{u}_t-\bar{u})^k$

and $(\hat{v}_t - \bar{v})^k$ on 20 lags of both variable for $k = 1,2,3,4$. The LM test statistic is $(T-20)R^2$ for each regression and is asymptotically distributed as χ^2_{20} under the null no serial correlation. Second, we test the null hypothesis whether the transformation of variables $\hat{u}_t$ and $\hat{v}_t$ are uniform (0,1) using the K-S test.

Table 3 Goodness-of-fit test for marginal distributions

	First moment	Second moment	Third moment	Fourth moment	K-S test
Brent	0.46	0.97	0.74	0.99	1.00
Corn	0.41	0.73	0.43	0.67	1.00
Soybean oil	0.55	0.77	0.45	0.78	1.00
Palm oil	0.93	0.18	0.82	0.17	1.00

Note: This table presents the p-values from LM tests and K-S tests, respectively.

The ρ values presented in Table 3 suggested that the null hypothesis of no serial correlation could not be rejected at the 5% significance level for all series and also the p-values from K-S tests show that all marginal distribution series can pass at the 5% significance level. The results represented above imply that the marginal distribution models were the correct specification. Therefore, the copula could be capturing the dependence structure between crude oil price and agricultural commodity prices in the right way.

4.2 Results for Copula Models

The results presented in Table 4 are the parameter estimates of constant dependence copulas. All parameters of elliptical copula family, Gaussian and Student-t copula, the correlation coefficient ρ are strong positive and strongly significant. The correlation coefficient is ranging from 0.245 to 0.571. The degrees of freedom of the Student-t copula are ranging from 13.18 to 18.45, indicating temperate extreme co-movements and tail dependence for all pairs.

However, all tail dependence values of the Student-t copula are generally small; tail dependence between crude oil and corn, soybean oil and palm oil is 0.012, 0.069 and 0.003 respectively. The results also show that all parameters for Frank copula are of relatively strong dependence between crude oil price and agricultural commodity prices.

For the asymmetric tail dependence cases, the parameter estimates for Clayton and Gumbel copulas were significant. The lower tail dependence parameters for the Clayton copula are strongly significant, where lower tail dependence value for pair of Brent-soybean oil, Brent-corn and Brent-palm oil are 0.467, 0.259 and 0.098, respectively. Likewise, the upper tail dependence parameters for the Gumbel copula are significant and the upper tail dependence value for pairs of Brent-soybean oil, Brent-corn and Brent-palm oil are 0.434, 0.286 and 0.179, respectively. The result

Table 4 Static Copula estimates of crude oil price-agricultural commodity prices

	parameter	Gaussian	Student t	Frank	Clayton	Gumbel
corn	ρ	0.3786	0.3865	2.4720	0.5136	1.2859
		(0.0006)	(0.0008)	(0.0065)	(0.0016)	(0.0009)
	ϑ		16.6045			
			(0.3148)			
	LL	-75.8978	-77.5329	-74.6675	-65.6629	-61.8003
	AIC	-151.7935	-155.0618	-149.3329	-131.3237	-123.5987
	BIC	-151.7885	-155.0518	-149.3279	-131.3187	-123.5937
soybean oil	ρ	0.5647	0.5719	4.0513	0.9091	1.5462
		(0.0004)	(0.0005)	(0.0070)	(0.0019)	(0.0012)
	ϑ		13.1843			
			(0.1855)			
	LL	-188.3787	-191.3463	-179.4465	-158.8212	-164.7097
	AIC	-376.7553	-382.6886	-358.8909	-317.6404	-329.4174
	BIC	-376.7503	-382.6787	-358.8859	-317.6354	-329.4124
palm oil	ρ	0.2448	0.2529	1.5659	0.2989	1.1567
		(0.0005)	(0.0009)	(0.0063)	(0.0015)	(0.0008)
	ϑ		18.4591			
			(0.3461)			
	LL	-30.3079	-31.9540	-31.8995	-26.3560	-23.8692
	AIC	-60.6139	-63.9039	-63.7969	-52.7099	-47.7364
	BIC	-60.6089	-63.8940	-63.7919	-52.7049	-47.7314

Note: Numbers in parentheses are standard errors.

of tail dependence implies that the crude oil price and each of agricultural prices can be crashing (booming) together at the same time.

Moreover, the goodness of fit of copula models is presented in Table 5 according to the empirical copula process (see Genest, Rémillard, and Beaudoin [22]). The result also showed that Student-t copula is accepted which is consistent with the best choice among the static copulas. As a result, we could not reject the null hypothesis that all static copulas above are suitable among the constant copula models. In addition, from the table 6, the time-varying of dependence parameters are more

Table 5 The Goodness of fit of Cramer-von Mises statistic

	Gaussian	Student-t	Frank	Clayton	Gumbel
Corn	0.233	0.391	0.064	0.005	0.005
Soybean oil	0.094	0.312	0.005	0.004	0.005
Palm oil	0.114	0.094	0.252	0.004	0.005

Note: This table shows the p-value of Cramer-von Mises statistic.

suitable than the constant copula model for all the couples of crude oil and agricultural commodity prices, which is consistence with the AIC and BIC, respectively.

Consequently, the results from the sample period can be concluded as follows: (1) the dependence between crude oil and agricultural commodity prices are strong, and (2) in the dependence structure between crude oil and agricultural commodity prices, there exists extreme dependence for pair of Brent-corn, Brent-soybean oil and Brent-palm oil. The best performing dependence models for all pairs are time-varying Student-t copula. However, the tail dependence for all pairs is generally weak.

5 Discussion

Our empirical results in the sample period show that the crude oil price co-movement with agricultural commodity prices is consistent with the situation of high world oil prices and a response to the high demand for biofuels. The dependence between crude oil price and agricultural price is evident in the relatively high prices of corn, soybean oil and palm oil because these are the main crops used in the production of biofuels (see Reboredo [24]). The subsidies in the biofuels industry affect the demand and prices of agricultural commodities as well as changes in fundamentals or policy actions in the energy crop market will be transmitted to the non-energy crops market. These results suggested that the energy policies, which subsidize the biofuels industry, may indeed have an impact on agricultural prices. Therefore, the energy crop prices are an important determinant in agricultural fundamentals allocation, which is affected by the supply of non-energy crops.

Moreover, there exists extreme tail dependence between the crude oil price and agricultural commodity prices. The tail dependence symmetry generally implies that crude oil and agricultural commodity prices are likely to move together during boom and bust markets. Policy makers should consider this behavior and its extreme effect of oil price on agricultural prices. The higher agricultural prices lead to more gain and benefit in the net agricultural exporting countries, but it will be a loss for net food importing countries, especially the poorer countries which may be negatively affected. Agricultural policies, food subsidies or trade policies should be designed and implemented to reduce the effect of extreme fluctuation in crude oil prices (see (Nazlioglu ans Soytas [8]; Reboredo [24]). Finally, our findings could give investors the ability to obtain a greater and more accurate assessment of the linkage between crude oil and agricultural commodity markets, which can protect their portfolio from loss during extreme market events. The symmetrical tail dependence implies that the alternative of using agricultural commodity markets for risk diversification purposes should be taken into account by investors when crude oil and agricultural markets are exhibit extreme tail dependence.

Table 6 Copula-Time Varying estimates of crude oil price-agricultural commodity prices

	parameter	Gaussian	Student t	Clayton	Gumbel
corn	ψ_0	0.0812	0.0660	-1.3572	2.1335
		(0.0048)	(0.0046)	(0.2742)	(0.0033)
	ψ_1	0.1133	0.0941	-1.3683	-0.8604
		(0.0033)	(0.0025)	(0.3945)	(0.0017)
	γ	1.7953	1.8481	-0.7264	-1.8828
		(0.0157)	(0.0142)	(0.1143)	(0.0017)
	LL	-79.0902	-80.5135	-72.2729	-68.8519
	AIC	-158.1744	-161.0208	-144.5451	-137.6977
	BIC	-158.1594	-161.0059	-144.5435	-137.6828
soybean oil	ψ_0	-0.1313	-0.1491	0.1176	-0.3556
		(0.0041)	(0.0044)	(0.0003)	(0.0231)
	ψ_1	0.1054	0.0836	-0.8479	-0.3044
		(0.0020)	(0.0018)	(0.0019)	(0.0133)
	γ	2.4159	2.4572	0.7946	0.4061
		(0.0088)	(0.0095)	(0.0006)	(0.0144)
	LL	-192.4536	-194.7249	-164.2613	-166.1149
	AIC	-384.9011	-389.4437	-328.5165	-332.2236
	BIC	-384.8861	-389.4287	-328.5015	-332.2086
palm oil	ψ_0	0.3909	0.4109	-0.5003	0.9525
		(0.0164)	(0.0292)	(0.0031)	(0.0073)
	ψ_1	0.1055	0.0425	-0.6807	-0.4829
		(0.0062)	(0.0051)	(0.0038)	(0.0019)
	γ	0.3691	0.3833	0.5557	0.0309
		(0.0680)	(0.1154)	(0.0019)	(0.0225)
	LL	-30.5481	-31.9987	-27.2014	-23.8804
	AIC	-61.0900	-63.9913	-54.3967	-47.7546
	BIC	-61.0751	-63.9763	-54.3818	-47.7396

Note: Numbers in parentheses are standard errors.

6 Conclusion

In this paper, we examine the co-movement between the crude oil futures and agricultural commodity futures, including corn, soybean oil and palm oil by using copula models with time invariant and time varying. We found evidence that the co-movement between crude oil price and agricultural commodity prices in the sample period is well performed by the Student-t copula with time varying for all pairs. Furthermore, we generally found that there was symmetrical tail dependence between crude oil prices and agricultural commodity prices; however, the tail dependence is relatively low for all pairs. And the structure of the time varying copula shows that the dependence parameters are very volatile over time and frequently deviate from their constant levels.

Acknowledgement. We would like to thank the Prince of Songkla University for providing funding of a PhD fellowship for Mr. Phattanan Boonynuphong.

References

1. OECD-FAO, Agricultural Outlook 2010-2019. OECD/FAO (2010)
2. Saban, N.: World oil and agricultural commodity prices: Evidence from nonlinear causality. Energy Policy 39(5), 2935–2943 (2011)
3. Tverberg, G.E.: Oil supply limits and the continuing financial crisis. Energy 37(1), 27–34 (2012)
4. Hanson, K.A., Sherman, R., Schluter, G.E.: Sectoral Effects of a World Oil Price Shock: Economywide Linkages to the Agricultural Sector. Journal of Agricultural and Resource Economics 18(01), 96–116 (1993)
5. Baffes, J.: More on the energy/nonenergy price link. Applied Economics Letters 17(16), 1555–1558 (2010)
6. Tang, K., Xiong, W.: Index Investment and Financialization of Commodities. SSRN eLibrary (2010)
7. Chen, S.T., Kuo, H.I., Chen, C.C.: Modeling the relationship between the oil price and global food prices. Applied Energy 87(8), 2517–2525 (2010)
8. Nazlioglu, S., Soytas, U.: Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. Energy Economics 34(4), 1098–1104 (2012)
9. Chang, T.H., Su, H.M.: The substitutive effect of biofuels on fossil fuels in the lower and higher crude oil price periods. Energy 35(7), 2807–2813 (2010)
10. Yu, T.E., Bessler, D.A., Fuller, S.: Cointegration and causality analysis of world vegetable oil and crude oil prices. Paper Presented at the The American Agricultural Economics Association Annual Meeting (2006)
11. Gilbert, C.L.: How to Understand High Food Prices. Journal of Agricultural Economics 61(2), 398–425 (2010)
12. Natanelov, V., Alam, M.J., McKenzie, A.M., et al.: Is there co-movement of agricultural commodities futures prices and crude oil? Energy Policy 39(9), 4971–4984 (2011)
13. Cifarelli, G., Paladino, G.: Oil price dynamics and speculation: A multivariate financial approach. Energy Economics 32(2), 363–372 (2010)
14. Chang, C.L., Khamkaew, T., McAleer, M., et al.: Modelling conditional correlations in the volatility of Asian rubber spot and futures returns. Mathematics and Computers in Simulation 81(7), 1482–1490 (2011)
15. Wu, C.C., Chung, H., Chang, Y.H.: The economic value of co-movement between oil price and exchange rate using copula-based GARCH models. Energy Economics 34(1), 270–282 (2012)
16. Patton, A.J.: Modelling asymmetric exchange rate dependence. International Economic Review 47(2), 527–556 (2006)
17. Ling, S.: Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. Journal of Econometrics 140(2), 849–873 (2007)
18. Bollerslev, T., Wooldridge, J.M.: Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11(2), 143–172 (1992)
19. Hansen, B.E.: Autoregressive conditional density estimation. International Economic Review 35(3), 705–730 (1994)
20. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. John Willey & Sons, Chichester (2004)

21. Brechmann, E.C.: Truncated and simplied regular vines and their applications. Technische Universitaet Muenchen (2010)
22. Genest, C., Rémillard, B., Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics 44(2), 199–213 (2009)
23. Diebold, F.X., Gunther, T.A., Tay, A.S.: Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review 39(4), 863–883 (1998)
24. Reboredo, J.C.: Do food and oil prices co-move? Energy Policy 49(0), 456–467 (2012)

Charitable Giving Behavior in Northeast Thailand and Mukdaharn Province: Multivariate Tobit Models

Jintanee Jintranun*, Peter Calkins, and Songsak Sriboonchitta

Abstract. As an Asian and Buddhist nation, Thailand's population is particularly concerned with charitable giving, and does so through several channels. This paper quantifies charitable donations for the poorest region of Thailand (the Northeast) and compares the shares of a household's charitable donation expenditures of three types with those allocated to religion-consistent food and religion-antipathetic alcohol consumption. It also tests for significant differences by degree of modernization (GPP/capita) and proximity to a major highway. We employed dataset collected by the National Statistical Office on nineteen Northeastern provinces, including 11,850 observations.The objective was to estimate multivariate tobit models and compare expenditure behavior under Engel's law. The analytical results reveal that the share of religious institutional donations, like the share of food, declines with income (Engel-negative); while alcohol expenditure, direct donations to NGOs display a positive relationship with household income (Engel-positive). Moreover, households located within poorer-than-average Mukdaharn province and along the East-West Economic Corridor province have a higher share of religious institutional donations than otherwise. Other demographic determinants of charitable giving include education level, the characteristics of the household head, and work status. If charitable donations by individuals to individuals (direct donation) increased by 60 percent of their current value, they would be adequate to eradicate poverty without recourse to governmental or NGO programs. Policy suggestions are made in this direction.

Jintanee Jintranun · Peter Calkins · Songsak Sriboonchitta
Faculty of Economics Chiang Mai University,
Chaing Mai Thailand
e-mail: jintanee@hotmail.com
petercalkins2004@yahoo.com
songsak@econ.cmu.ac.th

* For financial support, the first author is grateful to the Office of Higher Education Commission, Ministry of Education, Thailand, for a CHE-PhD 2550 scholarship.

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 269–281.
DOI: 10.1007/978-3-642-35443-4_19

1 The Real-World Problem

Thailand has achieved many of the U.N.'s Millennium Development Goals (MDGs), including the poverty and hunger targets at the national level. As a result, it has gone on to set the MDG plus goal of reducing the proportion of households below the poverty line in the northeast and three southernmost provinces to less than 4 percent by 2009. In addition to government and Non-government Organization initiatives in this direction, charitable giving is a third way in which the populace themselves may voluntarily close that gap. Charity is a virtually universal human behavior that originated in the fundamental principles and expressions of religious faith but has been extended to embrace more general human beliefs, social values and motives, particularly where the gift does not remain anonymous. But does charitable giving of all types increase or decrease as a share of expenditures with rising income? Do proximity to highways and modernization/poverty alleviation (as reflected in GPP/capita) stimulate or depress charitable giving? And what is the relationship between charitable giving and the consumption of gray goods such as alcohol, which may be considered both a part of Engel's prediction regarding food; and a good to be shunned by the same religious households that are expected to give more to charity. If charity increases with income and modernization, to what extent may public poverty-alleviation programs be cut back in favour of private charity? This paper seeks to provide answers to these three questions.

2 The Goals of the Study

To investigate in detail Northeastern Thai household consumption and giving behavior, the present paper therefore aims to

1) Isolate the significantly determinants of overall household expenditures on goods (e.g. food), bad or gray goods (e.g. alcohol), and charitable giving.
2) Determine whether charity is a substitute or complement to food and gray-goods.
3) Test the impacts the East-West Economic Corridor and modernization on giving.
4) Estimate how much of the poverty gap could be erased by charitable giving.
5) Make policy suggestions for efficiently stimulating and channeling charitable giving.

3 Review of Literature

Carroll et al. [4] estimated the impacts of income, dependency, employment, education and economic class on charitable donations in the Republic of Ireland using both the Tobit and double hurdle models. The results implies that non economic factors can be explained by a zero observation, whereby the role of zero is not just the standard corner solution. Showers et al. [8] estimated a charitable giving expenditure model for the U.S. Their model used total household expenditures to represent permanent income under the presumption that it provides a more accurate analysis of giving. The results showed that the response of religious giving was positive but low flexible on income and that religious giving behaved as a basic necessity good. Race, marital status, gender, age of household head, number of persons less than 18 years of age, and education level bore a significant positive relationship with giving. Specifically, black households had a higher probability and level of giving than nonblack households; and married households gave more than single males or females to religious organizations. Brown et al. [3] also modeled charitable donations in response to an unexpected natural disaster (tsunami). The study employed a system Tobit model to analyze panel household data. The study revealed that donations for the victims of a tsunami bore a positive relationship with previous donations to other charitable cause. That means charitable donations to the victims of a tsunami exhibit a complementary rather than a substitution effect.

4 Multivariate Tobit Model

The present paper will employ a multivariate tobit model with simultaneous maximum likelihood calculations. Previous empirical studies that have employed multivariate tobit include study of fluid milk purchases in the United States, multivariate tobit examined post-wildfire reseeding on arid rangeland in a western U.S. state for three types of plant: (1) unwanted invasive grasses, (2) seeded grasses, and (3) sagebrush underlying the measured densities, several years after fires. The details of the methods of data collection and analysis will be shown in later sections [5, 6].

5 The Hypotheses of the Study

The present study seeks to test four (4) empirical hypotheses:

H1. The economic variables income, household employment rate and/or highest education significantly reduce the share of food consumption, but increase the shares of alcohol consumption and overall charitable donations.

H2. Roads (EWEC) and/or modernization (other provinces than Mukdaharn) significantly reduce the Engel shares of food consumption and total charitable donations, but increase the share of alcohol consumption.

H3. Females, the elderly, married people, Thais, the fully employed, the more modern, and Buddhists contribute more than their counterparts to all three types of charity as a percentage of their income.

H4. All three types of donation (Religouse institution, NGOs and Direct to household or person) increase with income, education, and employment rate. This hypothesis tests to what extent the government could focus on education and employment creation policies rather than direct poverty-relief policies.

6 Methods of Modeling and Estimation

Both to model the determinants of household expenditure behavior that involve many choices and to correctly treat dependent variables with-non zero observations, this study applies multivariate Tobit models (MVTOBIT). The multivariate model postulates that the same household may choose among many expenditure channels. This is a reasonable assumption since the same household can simultaneously select both which kind of goods to consume and which charitable channels to employ.

The models used in the present research were originated by Amemiya [1], who created MVTOBIT by extending the univariate Tobit model to multivariate and simultaneous equation models. The utility of the MVTOBIT formulation resides chiefly in the feature that truncated dependent variables are jointly determined. The multivariate model to be termed **model 1** in the present research is designed to explain the choice of relative expenditure shares of three types: the share of food consumption y_{1i}^*, the share of alcohol consumption y_{2i}^* and the share of religious institutional donations y_{3i}^*. These three variables are all unobserved or latent; and depend upon such household characteristics (Xi) as income, province, proximity to EWEC, education, religion, nationality, marital status, gender, and employment. In addition, the error terms ε_{1i}, ε_{2i} and ε_{3i} collect unobserved characteristics that affect expenditure behavior. The term θ indicates the a certain threshold level, it equal to zero in this study. Further details of the model follow the reasoning of Amemiya) [1].

The multivariate Tobit model assumes that the joint density function of and behave as a multivariate normal distribution with zero mean, constant variances and a constant correlation between error terms, where ε is an m-multivariate $\mathrm{N}(0,\Omega)$ variable and is independent of X_i. The parameters β will be estimated by maximum likelihood using mvtobit command in STATA program as coded by Barslund (2009) [2]. The mvtobit command employs a multivariate tobit with SML (Simultaneous Maximum Likelihood) model.

The above variables and terms have been combined to formulate the three components of our research model. The food share consumption equation becomes:

$$\begin{aligned} y_{1i}^* &= X_{1i}\beta + \varepsilon_{1i} \\ y_{1i} = y_{1i}^* &\quad \text{if} \quad y_{1i}^* > \theta \\ y_{1i} = 0 &\quad \text{if} \quad y_{1i}^* \leq \theta \end{aligned} \tag{1}$$

The alcohol share consumption equation may similarly be expressed as:

$$\begin{aligned} y_{2i}^* &= X_{2i}\beta + \varepsilon_{2i} \\ y_{2i} = y_{2i}^* &\quad \text{if} \quad y_{2i}^* > \theta \\ y_{2i} = 0 &\quad \text{if} \quad y_{2i}^* \leq \theta \end{aligned} \tag{2}$$

Meanwhile, the share of religious institutional donations is given by:

$$\begin{aligned} y^*_{3i} &= X_{3i}\beta + \varepsilon_{3i} \\ y_{3i} = y^*_{3i} \quad &\text{if} \quad y^*_{3i} > \theta \\ y_{3i} = 0 \quad &\text{if} \quad y^*_{3i} \leq \theta \end{aligned} \tag{3}$$

In order to capture a different question the choice among donation channels, a second multivariate model **(model 2)** was also formulated. It was composed of three donation equations: those through religious institutional channels y^*_{3i}, through NGOs institutional channels y^*_{4i} and in the form of direct donations to households or individuals y^*_{5i}. Again, these unobserved or latent variables depend on the household characteristics X_i. ε_{3i}, ε_{4i} and ε_{5i} and θ are as defined above. The remaining two components are thus defined as NGOs institutional donations equation:

$$\begin{aligned} y^*_{4i} &= X_{4i}\beta + \varepsilon_{4i} \\ y_{4i} = y^*_{4i} \quad &\text{if} \quad y^*_{4i} > \theta \\ y_{4i} = 0 \quad &\text{if} \quad y^*_{4i} \leq \theta \end{aligned} \tag{4}$$

And the direct donations equation:

$$\begin{aligned} y^*_{5i} &= X_{5i}\beta + \varepsilon_{5i} \\ y_{5i} = y^*_{5i} \quad &\text{if} \quad y^*_{5i} > \theta \\ y_{5i} = 0 \quad &\text{if} \quad y^*_{5i} \leq \theta \end{aligned} \tag{5}$$

We may now show in plain text form the specific expected determinants of the two sets of dependent variables:

Model 1
Food consumption/Alcohol consumption/Religious donations = f [Income, Income2, Location (Non-Mukdaharn, proximity to EWEC), Highest Education (secondary, university, masters), Employment (household employment rate), Other household characteristics (Buddhist, elderly, Thai, married, female)] [1]

Model 2
Religious/NGO/Direct donations= f [Income, Income2, Location (Non-Mukdaharn province, EWEC), Highest Education (secondary, university, masters), Employment (household employment rate), Other household characteristics (Buddhist, elderly, Thai, married, female)] [2].

7 Empirical Results

Table 1 reports the descriptive statistics of the sample of 11,850 household observations in Northeast Thailand. In addition to the minimum, maximum, mean, standard deviation and the coefficient of variation (st.dev/mean). The results from the two multivariate Tobit models (Model 1 and 2) are shown in table 2. model 1 reports a simultaneous multivariate Tobit model for three dependent variables: the shares of food expenditure ***(FOOD)***, alcohol expenditure ***(ALCOHOL)*** and religious donations ***(RELIGIOUS)***. It should be noted that this multivariate model includes one dependent variable (food expenditure) with non-zero observations. The result can be run by the MVTOBIT command in STATA. However, the maximum number of censored equations is 2 less than or equal 2: simulations are not needed. The process continues with conventional maximum likelihood. Moreover, the only instance of multicollinearity in the correlation matrix is between highest educted in bechelor deegree and higher ***(UNIV)*** and work as professional ***(PROFESS)*** (Pearson r = 0.5835), however these two variables were never used in the same equation.

8 Hypothesis Testing

H1. *The economic variables income, household employment rate and/or highest education significantly reduce the share of food consumption, but increase the shares of alcohol consumption and overall charitable donations.* Based on the results of Table 2, we cannot reject this hypothesis. Higher income reduces the food share and donations to religious institutions, but increases the share of both NGO and direct donations. A higher employment rate reduces the food share, but increases alcohol share and donations to all types of charity. Maximum secondary education reduces food consumption and religious donations but increases alcohol consumption and donations to NGOs and direct to others. University and Masters degrees have the same effects, except that there is no significant change in alcohol consumption and religious donations.
Even though food is Engel-negative and the shares of both NGO and direct donations are Engel-positive, there is a) a positive effect of income on alcohol consumption and b) a negative effect on religious institutional donations. We may conclude that food and religious donations are inferior goods while alcohol, NGO charity and direct donations are luxury goods.

H2. *Roads (EWEC) and/or modernization (other provinces than Mukdaharn) significantly reduce the Engel shares of food consumption and total charitable donations, but increase the share of alcohol consumption.* Once again, we fail to reject this hypothesis. However, the results show that with the growing depersonalization of the modern, connected economy, there is a significant shift among the three types of charitable donations. People substitute religious organizations and NGOs for direct donations within their overall declining share of charitable donations. Meanwhile, EWEC increases religious donations but reduces direct donations. There is no significant effect of road on either consumption patterns or NGO donations; but modernization does increase alcohol consumption.

H3. *Females, the elderly, married people, Thais, the fully employed, the more modern, and Buddhists contribute more than their counterparts to all three types of charity as a percentage of their income.* We must also reject this hypothesis. Married people, females, Buddhist, those able to take care of themselves, technicians and employees actually have higher overall donation rates; while the more modern, elderly, Thai, clerks and laborers have lower shares of total donation. This means that a general policy of job creation and education can be contemplated by the

government as a direct replacement to direct aid to poor households to which only the elderly and laborers will not respond as intended. But this is consistent with such a policy, since the elderly and adult laborers would not normally be targeted by either new jobs or higher education.
Separating the results by charity channel, we find that females contribute more than males to NGOs donation as a share of their income. However, there is no significant difference between genders with respect to religious and direct donation.

When we considered household head's age, elderly heads have few or no dependents and a more sufficient life style than the new generation. The results revealed that the elderly contribute a greater share of their income than their counterparts to religious institutions and NGOs, but a lower share in direct donations to households or persons. This implies a close closely relationship between the elderly and Wats or charitable organizations.

Moreover, the results reveal that, while marital status has no significant impact on NGO donations, married people contribute more than their counterparts to religious and direct donation as a share of their income. As if to respond to who does contribute to NGOs, Thai language spoken in household shows a significantly higher contribution to NGOs than non-Thais. In complementary fashion, minority groups pay respect to religious institutions by giving more to them. Economic variables are also significant determinants of charitable giving. The fully employed contribute more than their counterparts to all three types of charity as a percentage of their income. We may surmise that the labor force of such households has a greater power to donate because of their low dependency ratio. The more modern also contribute more than their counterparts to religious institutions and NGOs as a percentage of their income, but give less in direct aid to households or individuals. We may conclude that modernization induces households to contribute more through formal channels and less through informal channels.

Finally, Buddhists contribute more than their counterparts only to religious institutions. However, they display no significant difference from other religions in NGO or direct contributions. This result implies that Buddhist culture pays greater respect to wats than their other religions do to their religions own institutions.

H4. *All three types of donation (to religious institutions, NGOs, and directly to households or persons) increase with income, education, and employment rate.* We may partially accept this hypothesis in the cases of both NGO donations and direct donations to individuals. In fact, NGO donation shares are higher not only for income, employment rate, modernization, and all completed levels of education; but also for the elderly, females and Thai households. Direct donations are higher not only for income, employment, and all levels of education; but also for married households whose heads are employees as opposed to common laborers. And it is interesting that modernization leads to reduced direct donation. But we must reject the hypothesis with respect to contributions to religious institutions. Indeed, the increase in direct and NGO donations appears to be at the expense of reduced religious donations. More specifically, although religious donations increase with the employment rate, proximity to EWEC and modernization (captured by the variable non-Mukdaharn) lower them through increased income and the completion of secondary education. Since the movement toward higher income, higher education, and fuller employment is a normal part of the process of economic development, the government can therefore focus on employment and education programs and gradually step aside from direct transfer payments to the poor. It will be the people themselves who voluntarily take up the slack with contributions to NGOs and needy individuals.

9 Conclusions and Policy Implications

Table 3 presents a policy matrix that summarizes the results of Table 2. It shows which policies are most appropriate; and who should be responsible for implementing them. The results reveal that religious institution donations are, like food, Engel-negative with respect to income. This means they could not serve as a powerful income distribution channel like a progressive tax. Direct donations and NGOs are better redistribution channels. If government launches social programs to strengthen community relationships, this should automatically lead to increased direct contributions to persons and households. Moreover, efficient NGOs seem to a good channel for progressive income redistribution.

The results also indicate that Mukdahan and other provinces located in the East West Economic Corridor have higher religious donations than otherwise. They thus have the potential to promote Buddhist and cultural tourism campaigns to increase religious donations. The coefficient of income on religious donations is negative, while the coefficients of income on NGOs and direct contributions are low compared with the coefficient of employment rate and education. Government should focus on education and employment creation rather than direct poverty-relief policies. Given the differences in behavior by charitable giving channel, policies to promote charity in each channel should also differ. The value of direct donations (informal channel) is higher than that of religious and NGOs institution donations (formal channel). If government wishes to promote the formal charity channel it should improve its income tax policy.

Finally, alcohol consumption is negatively correlated with age. Government is therefore launching a campaign to reduce alcohol consumption among young people. We may now answer the question posed at the outset: “To what extent can poverty-reduction programs rely only on individual charitable donations?” (Table 4). Under Gap 1, voluntary donations with administrative costs could close 213 percent of the poverty gap with a surplus of 123 million baht at the provincial level. Under Gap 2 (without administrative costs) this rises to 428 percent with a surplus three times greater. Moreover, giving exceeds need in each of the three areas: rural, semi-urban and urban and for both gaps 1 and 2, with a minimum of 121 percent in the rural areas under gap 1. These results imply that government and/or NGO programs will not be necessary overall or in any area.

However, current direct donations alone (gap 3) can eliminate only 62 percent of poverty with a shortfall of some 41 million baht at the provincial level. This pattern holds true at varying levels for each of the urban, semi-urban and rural areas. Since giving is less than need, charitable institutions and programs will still be necessary. NGO and religious charities cannot simply disappear.

Acknowledgement. For financial support, the first author is grateful to the Office of Higher Education Commission, Ministry of Education, Thailand, for a CHE-PhD 2550 scholarship. And we wish to thank Professor David Roland-Holst of the University of California at Berkeley and Professor Aree Wiboonpongse of Chiang Mai University for their invaluable help and encouragement with the research reported in this chapter.

Appendix

Table 1 Summary Statistics

Variable	Mean	Std. Dev.	Coeff var Percent	Min	Max	Type
Dependent variables						
NGO	0.001	0.006	849.2	0.00	0.29	Numeric
DIRECT	0.029	0.083	289.0	0.00	0.77	Numeric
ALCOHOL	0.004	0.009	243.7	0.00	0.14	Numeric
RELIGIOUS	0.016	0.019	114.5	0.00	0.63	Numeric
FOOD	0.402	0.152	37.9	0.00	0.92	Numeric
Economic variables						
CLERK	0.019	0.136	718.8	0.00	1.00	Dummy
TECHN	0.026	0.160	608.1	0.00	1.00	Dummy
PROFESS	0.062	0.241	390.0	0.00	1.00	Dummy
NONMUK	0.205	0.404	197.0	0.00	1.00	Dummy
EWEC	0.263	0.440	167.3	0.00	1.00	Dummy
INCOME	16.898	25.044	148.2	0.00	922.96	Numeric
LABOURER	0.747	0.435	58.2	0.00	1.00	Dummy
EMPLRATE	0.587	0.299	51.0	0.00	1.00	Numeric
EMPLOYEE	0.924	0.265	28.6	0.00	1.00	Dummy
Education						
MASTERS	0.015	0.122	809.8	0.00	1.00	Dummy
UNIV	0.109	0.311	286.5	0.00	1.00	Dummy
SECONDARY	0.992	0.090	9.0	0.00	1.00	Dummy
Socioeconomic variables						
AGE	50.849	14.270	28.1	12.00	99.00	Numeric
FEMALE	0.304	0.460	151.3	0.00	1.00	Dummy
MARRIED	0.933	0.251	26.9	0.00	1.00	Dummy
THAI	0.970	0.170	17.5	0.00	1.00	Dummy
BUDD	0.991	0.093	9.4	0.00	1.00	Dummy
CARESELF	0.992	0.090	9.0	0.00	1.00	Dummy

Source : From calculation

Table 2 Tobit Model Results

	Model 1			Model 2		
	FOOD	ALCOHOL	RELIGIOUS	RELIGIOUS	NGO	DIRECT
Economic variables						
INCOME	-0.0029 0.000***	0.0001 0.001***	-0.0001 0.000***	-0.0001 0.000***	0.0001 0.000***	0.0021 0.000***
INCOME2	0.0000 0.000***	0.0000 0.000***	0.0000 0.000***	0.0000 0.000***	0.0000 0.000***	0.0000 0.000***
NONMUK		0.0035 0.000***	0.0058 0.032**	0.0058 0.000***	0.0084 0.000***	-0.0599 0.000***
EWEC			0.0028 0.000***	0.0028 0.000***		-0.0158 0.032**
EMPLRATE	-0.0412 0.000***	0.0134 0.000***	0.0063 0.000***	0.0063 0.000***	0.0022 0.097*	0.2182 0.000***
PROFF		-0.0042 0.008***				
TECHN		0.0074 0.000***	-0.0030 0.024**	-0.0030 0.022**	-0.0112 0.000***	0.0226 0.163
CLERK			-0.0032 0.043**	-0.0032 0.042**		
LABOURER	0.0298 0.000***	-0.0023 0.000***	-0.0023 0.000***	-0.0023 0.000***	-0.0045 0.000***	-0.0506 0.000***
EMPLOYEE						0.0298 0.014**
Education						
SECONDARY	-0.0519 0.000***	0.0025 0.002***	-0.0012 0.018**	-0.0012 0.019**	0.0042 0.000***	0.0888 0.000***
UNIV	-0.0902 0.000***				0.0025 0.096*	0.1314 0.000***
MASTERS	-0.0982 0.000***				0.0064 0.022**	0.1094 0.000***
Socioeconomic variables						
AGE	-0.0007 0.000***	-0.0002 0.000***	0.0003 0.000***	0.0003 0.000***	0.0001 0.000***	-0.0012 0.000***
FEMALE	-0.0194 0.000***	-0.0101 0.000***			0.0019 0.025**	
MARRIED	-0.0087 0.064**		0.0036 0.000***	0.0036 0.000***		0.0596 0.000***
THAI		-0.0080 0.000***	-0.0185 0.000***	-0.0187 0.000***	0.0132 0.000***	
BUDD	-0.0219 0.090*	0.0063 0.091*	0.0046 0.034**	0.0046 0.030**	0.0000	0.0000
CARESELF			0.0046 0.033**	0.0045 0.049**		

Note ***sig. at 99 confidence level, ** sig. at 95 confidence level, * sig. at 90 confidence level

Table 3 A Policy Matrix Based on the Empirical Results of This Study

Preference rank	Policy-amenable variable	Pos or neg impacts on: Direct	NGO	Relig	Implied policies	Implementable by Govt	NGO	Third way
. 1	Employment rate	pos	pos	pos	Create jobs	x	x	
					Launch social programs to strengthen community relationships → direct giving	x	x	
2	Married	pos		pos	Social advertising to promote formal marriage	x	x	
3	Employee	pos			Convert laborers into employees through training and legislation, but do away with clerks	x	x	
4	Female		pos		Social advertising to attract males to donate to NGOs	x	x	
5	Buddhist			pos	Promote tourism campaigns as a Buddhism and heritage cultural	x		
6	Careself			pos	Help the infirm, elderly, handicapped, single-parent families to meet their food and housing needs		x	
7	Income (Engel pos, neg or neutral)	pos	pos	neg	Improve tax policies through reduced income tax rates and credits for donations	x		
8	Education	pos	pos	neg	Promote education at all levels through scholarships	x		
9	Technicians	pos	neg	neg	Create training programs, scholarships, junior college programs in engineering	x	x	x
10	Modernization	neg	pos	pos	Promote tourism campaigns as a Buddhism and heritage cultural			
11	Age	neg	pos	pos	Discourage alcohol consumption among young	x	x	
12	EWEC	neg		pos	Promote tourism campaigns as a Buddhism and heritage cultural	x		
13	Laborer	neg	neg	neg	Convert laborers into employees	x	x	

Table 4 Estimation of Poverty Elimination

Variable	Total households in Mukdahan munipality	Urban	Semi urban	Rural
Universe	101,420	14,635	66,107	20,678
Demand side				
Incidence of poverty	0.372	0.289	0.393	0.400
Depth of poverty	2,892	1,978	2,980	3,679
Total value to be made up to eliminate poverty	109,125,161	8,364,120	77,410,939	30,428,587
Supply side				
Income per household	242,277	279,649	217,011	279,637
Total income across all households	24,571,733,340	4,092,663,115	14,345,946,177	5,782,333,886
percent giving necessary to eliminate poverty (i.e. line 6 divided by line 9)	0.44 percent	0.20 percent	0.54 percent	0.53 percent
Current mean direct donations per household	672	407	787	510
Current mean NGO donations per household	485	634	542	147
Current mean religious donations per household	3,445	3,982	3,301	3,327
Current mean religious donations per household	3,445	3,982	3,301	3,327
Current mean donations assuming no administrative costs	4,601	5,023	4,629	3,984
Current mean donations assuming administrative costs	2,295	2,231	2,447	1,778
Total current donations assuming no administrative costs (m baht)	466.6	73.5	306.0	82.4
Total current donations assuming administrative costs (m baht)	232	32.7	161.7	36.7
Net positive or negative balance				
Gap 1 (m baht) between all donations and poverty gap with administrative costs	- 123.6	- 24.3	-84.3	- 6.3
percent	213	390	209	121
Gap 2 (m baht) between all donations and poverty gap with no administrative costs -357.5	-65.1	-228.6	-51.9	
percent	428	879	35	271
Gap 3 (m baht) between direct donations alone and the poverty gap	41.0	2.4	25.4	19.9
percent	62	71	67	35

References

1. Amemiya, T., et al.: Multivariate Regression and Simultaneous Equation Models When The dependent Variables are Truncated Normal. Econometrica 42, 999–1012 (1974)
2. Barslund, M.: MVTOBIT: Stata module to calculate multivariate tobit models by simulated maximum likelihood (SML), Boston College Department of Economics (2009)
3. Brown, S., Harris, M.N., Taylor, K.: Modelling charitable donations to an unexpected natural disaster: Evidence from the U.S. panel study of income dynamics. Sheffield Economics Research Paper Series SERP Number: 2009015, 1–23 (2009) ISSN: 1749-8368
4. Carroll, J., McCarthy, S., Newman, C.: The determinants of charitable donations in the republic of Ireland. Trinity Economics Papers tep12, Trinity College Dublin, Department of Economics (2005)
5. Cornick, J., Cox, T.L., Gould, B.W.: Fluid Milk Purchases: A Multivariate Tobit Analysis. Am. J. Agr. Econ. 76(1), 74–82 (1994)
6. Eiswerth, M.E., Shonkwilerb, J.S.: Examining post-wildfire reseeding on arid rangeland: A multivariate tobit modelling approach. Ecological Modelling 192(1-2), 286–298 (2006)
7. Lee, L.-F.: Multivariate Tobit Models in Econometrics. Handbook of Statistics 11, 145–173 (1993)
8. Showers, V.E., Showers, L.S., Begg, J.M.: Charitable Giving Expenditues and the Faith Factor. American Journal of Economics and Sociology 70(1), 152–186 (2011)
9. Sickles, et al.: The politics of nature. In: Smith, J. (ed.) The Rise of Modern Genomics, 3rd edn. Wiley, New York (1979)

Analysis of Volatility and Dependence between the Tourist Arrivals from China to Thailand and Singapore: A Copula-Based GARCH Approach

Jianxu Liu and Songsak Sriboonchitta

Abstract. This paper aims to estimate the dependency between the growth rates of tourist arrivals of Thailand and Singapore from China, and also analyze their conditional volatilities. Firstly, we assume that both margins are skewed-t distribution, and then make use of ARMA-GARCH model to fit monthly time series data. Secondly, fifteen types of static copulas are used to fit static dependence between tourist arrivals to Thailand and Singapore from China. We take the AIC, BIC and the two tests based on Kendall's transform as criterions for goodness of fit test. Moreover, we apply time-varying copulas that described the dynamic Kendall's tau process. Results show that each growth rate of tourist arrivals has a long-run persistence of volatility, and the time-varying Gaussian copula has the highest explanatory power of all the dependence structures between tourist arrivals to Thailand and Singapore from China in terms of AIC and BIC values.

1 Introduction

With the continuous development of China's economy and the continuous improvement of Chinese people's living standards, China has become the main tourist source market for ASEAN countries. In the past two years, the annual growth rate of the number of outbound tourism maintains more than 20%. In 2010, the number of mainland Chinese tourists abroad reached 57.39 million passengers, an increase of 20.4%, and the number of outbound tourists in 2011 was 70.25 million passengers, an increase of 22.4%. Thus, the market size of the outbound tourism continues to expand, and the study of outbound tourism should be a great interest to the relevant government departments and tourism enterprises.

Jianxu Liu · Songsak Sriboonchitta
Faculty of Economics, Chiang Mai University,
Chiang Mai 50200 Thailand
e-mail: liujianxu1984@163.com,
songsak@econ.cmu.ac.th

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 283–294.
DOI: 10.1007/978-3-642-35443-4_20

Travel to Southeast Asia has long been favored by Chinese tourists. Thailand, Singapore and Malaysia are the top destinations for travel. According to statistical results of the number of tourists in 2006-2010, the rank of Thailand and Singapore has always been maintained at second or third in Southeast Asia, and there are many tour routes that are designed for traveling to these two countries. We found that not only to maintain a trend of increase of the tourist arrivals to these two countries, Thailand and Singapore, but also there is a positive linear correlation between them. From the raw data, it is roughly seen to have the correlation between the two countries significantly higher than other countries. So it is interesting to investigate the rank and linear correlations econometrically of Chinese outbound to Thailand and Singapore, which is one of the problems of this study. In addition, the growth rates of outbound tourism have obvious volatility. It is then interesting to analyze the persistence of volatility and short-term shocks in the volatility of the Chinese tourist arrivals to Thailand and Singapore.

There are very few papers studying the travel-related volatility, linear correlation and rank correlation in tourism field. Zhenzhen Liu [16] studied the linear correlation of tourist flows and trade flows; Fengbo Wang [8] studied the volatility of international inbound tourism demand in China using ARIMA and GARCH model; Felix Chan [7], John T. Coshall [11], Johann du Preez [10] also used ARIMA and CCC-GARCH, as well as GJR models, to analyze volatilities and forecast demand. Recently, the copula based GARCH model has been very popular in financial field, as it can be used to analyze volatilities and dependence structure. Patton [1] used this model to analyze the dynamic dependence between exchange rates of Yen-USD and DM-USD. Chih-Chiang Wu [5] also researched the economic value of co-movement between oil prices and exchange rates using copula-based GARCH models. Kehluh Wang [12] studied the dynamic dependence between the Chinese market and other international stock markets using time-varying copula approach. We can see that the copula based GARCH model has reached maturity. Therefore, we apply this model to analyze the dependence structure and volatility in tourism field, which could be beneficial to relevant stakeholders in tourism industry.

The contributions in this paper are threefold: first, we bring the copula based GARCH model into tourism field. Second, the kinds of dynamic copulas are expanding. i.e, the dynamic copulas of BBX and Joe are invented. Last, for the marginal distribution in this paper, we use the skewed student-t distribution that is different from the one Hansen provided in 1994.

The remainder of this paper is organized as follows. In section 2, the econometrics model is reviewed, which contains skewed-t distribution, ARMA-GARCH model and copulas. In Section 3, the empirical study is described and results are presented. Section 4 proposes the policy implication aim at the development of tour industries in Thailand and Singapore. Section 5 offers our conclusions.

2 Econometrics Model

Copula approach is a useful tool for modeling joint distribution. It was proposed by Sklar [15], and Nelsen [12] extensively discussed the characteristics and theorem of copula. In this part, we introduce the skewed student-t distribution, ARMA-GARCH model and copulas.

2.1 *Skewed Student-t Distribution*

Carmen Fernandez and Mark F.J. Steel [3] generate a skewed student-t distribution, which displays both flexible tails and possible skewness, each entirely controlled by a separate scalar parameter. The formula of skewed-t distribution is shown as

$$P(x_i \mid v, \gamma) = \frac{2}{(\gamma + 1/\gamma)} \{ f_v(x_i/\gamma) I_{[0,\infty]}(x_i) + f_v(\gamma x_i) I_{[\infty,0]}(x_i) \} \tag{1}$$

where fv(.) is the density function of student t distribution. The parameter v represents the degree of freedom, and γ is the skewness parameter that is defined from 0 to ∞; I denotes the indicator function.

2.2 *ARMA-GARCH Model*

To investigate the volatility and co-movement of the Tourist Arrivals from China to Thailand and Singapore, we proposed copula based ARMA-GARCH model, which is defined that each variable is a process of an ARMA-GARCH for the marginal distribution and the standardized innovations submit to skewed student-t distribution, and one copula for the joint distribution. In the following, the ARMA-GARCH model is shown as:

$$r_t = c + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_i \tag{2}$$

$$\varepsilon_t = h_t \eta_t \tag{3}$$

$$h_t^2 = \omega + \sum_{i=1}^{k} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{l} \beta_i h_{t-i}^2 \tag{4}$$

where $\sum_{i=1}^{p} \phi_i < 1, \omega > 0, \alpha_i >= 0, \beta_i >= 0$, and $\sum_{i=1}^{k} \alpha_i + \sum_{i=1}^{l} \beta_i < 1$. η_t is the standardized residual, which can be assumed for any distribution. Here, we assume that it is skewed student-t distribution. Specially, skewed student-*t* distribution can capture characteristics of heavy tail and asymmetry anyway, and symmetric heavy tail for student-*t* distribution.

2.3 Copulas

It is necessary to understand the copula family and the characteristics of each copula. The copula family used in our work that includes Gaussian copula, T copula, Clayton copula, Gumbel copula, Frank copula, BB1, BB7, BB8 and rotate copulas. Different copulas have different characteristics. To accurately capture the dependency, we apply copulas as much as possible.

(1) Gaussian copula

The Gaussian copula takes the following form:

$$C_{Ga}(u_1,u_2 \mid \rho) = \int_{-\infty}^{\phi^{-1}(u_1)} \int_{-\infty}^{\phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\} dx_1\, dx_2. \tag{5}$$

where the ρ is the Pearson correlation, both u_1 and u_2 are submitted to uniform distribution, which are the CDFs of the standardized residuals from the marginal models. The Gaussian copula can reflect the positive and negative correlation, and the Pearson correlation ρ can be transformed to kendalltau that equals to 2/arcsin(ρ).

(2) T copula

T copula is the same with Gaussian copula that belongs to elliptical copula, but T copula can capture the tail dependency, and it is symmetric extreme dependence. The T copula is defined as

$$C_T(u_1,u_2 \mid \rho,\upsilon) = \int_{-\infty}^{T^{-1}(u_1)} dx_1 \int_{-\infty}^{T^{-1}(u_2)} dx_2 \frac{1}{2\pi\sqrt{1-\rho^2}} \{1+\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{\upsilon(1-\rho^2)}\}^{\lambda}. \tag{6}$$

where the $\lambda = -(\upsilon+2)/2$.The ρ is the Pearson correlation that is the same with Gaussian copula, and the υ is the degree of freedom that is related to the symmetric heavy tail. When the υ is large enough, for example, equals to 100 that means the T copula would be the same with Gaussian copula.

(3) Archimedean copulas

Archimedean copulas are nonelliptical copulas that include Clayton, Frank, Gumbel, Joe, BBX etc.Different copulas have different properties. Clayton copula can reflect the lower tail dependence. Frank copula has symmetric tail dependence, and can describe the positive and negative dependence. Gumbel copula is an asymmetric copula of the Archimedean family, which allows for upper tail dependence. Joe copula can help us capture the upper tail dependence.BBX copulas are two-parameter copula, and BB8 can capture the upper tail dependence. But both BB1 and BB7 copulas can reflect the upper tail and lower tail dependence.

(4) Time-varying copulas

Patton [1] proposed that it is very difficult to know what might influence the parameters to change in copula model, thus, he assumed that the parameters in Gaussian and SJC copula follows an ARMA (1,10) type process. Hans Manner [9] and Chih-Chiang Wu [5], Carlos Almeida [2], et al., further research time varying copulas. In our study, on one hand, we go on use time varying copulas that have been proposed by predecessors; on the other hand, we also present new time varying copulas for two parameters copula, namely, BB1, BB7 and BB8.

1. Time varying Gaussian copula

$$\rho_t = \tilde{\Lambda}(\omega + \beta_1\rho_{t-1} + \cdots + \beta_p\rho_{t-p} + \alpha\frac{1}{q}\sum_{j=1}^{q}\phi^{-1}(u_{1,t-j})\phi^{-1}(u_{2,t-j})) \tag{7}$$

$\tilde{\Lambda}$ is a logistic transformation which is defined as follows:

$$\tilde{\Lambda} = (1 - e^{-x})(1 + e^{-x})^{-1} \tag{8}$$

the purpose of using this logistic transformation is to keep the correlation coefficient ρ belonging to (-1, 1).

2. Time varying T copula

$$\rho_t = \tilde{\Lambda}(\omega + \beta_1\rho_{t-1} + \cdots + \beta_p\rho_{t-p} + \alpha\frac{1}{q}\sum_{j=1}^{q}\phi^{-1}(u_{1,t-j};\upsilon)\phi^{-1}(u_{2,t-j};\upsilon)) \tag{9}$$

T copula has two parameters that are Pearson correlation and degree of freedom υ. Obviously, assume that fixed the degree of freedom, just let the correlation be change with time.

3. Time varying (rotate) Gumbel copula

$$\tau_t = \Lambda(\omega + \beta_1\tau_{t-1} + \cdots + \beta_p\tau_{t-p} + \alpha\frac{1}{q}\sum_{j=1}^{q}\left|u_{1,t-j} - u_{2,t-j}\right|) \tag{10}$$

where$\Lambda = (1 + e^{-x})^{-1}$. This guarantees that the Kendall's tau will be between -1 and 1, and the time varying Joe copula employ the same form as it.

4. Time varying (rotate) Clayton copula

$$\tau_t = \Lambda(\omega + \beta_1\tau_{t-1} + \cdots + \beta_p\tau_{t-p} + \alpha\left|u_{1,t-j} - u_{2,t-j}\right|) \tag{11}$$

5. Time varying BBX copulas

BB1, BB7 and BB8 are two parameters copulas, and we assume that two parameters in each copula are changeable with time, and the time varying form of BB1, BB7 and BB8 employ the time varying Gumbel one, which follows an ARMA (p, q) type process, specially, our case is an ARMA (1, 20) type process that is better fit than ARMA (1, 10) proposed by Patton [1].

$$\theta_t = H(\omega + \beta_1 \theta_{t-1} + \cdots + \beta_p \theta_{t-p} + \alpha \frac{1}{q} \sum_{j=1}^{q} |u_{1,t-j} - u_{2,t-j}|) \tag{12}$$

$$\delta_t = H(\omega + \beta_1 \delta_{t-1} + \cdots + \beta_p \delta_{t-p} + \alpha \frac{1}{q} \sum_{j=1}^{q} |u_{1,t-j} - u_{2,t-j}|) \tag{13}$$

where the H is a logistic transformation, which is the same with Λ in time varying Gumbel copula, when the formulas of two parameters focus on time varying BB1 and BB7 copula. In time varying BB8 copula, the H in each parameter conveys the different logistic transformation. The H in function of the parameter δ is the same with time varying BB1 and BB7 copula, but the H equals to $\tilde{\Lambda}$in the function of the parameter θ.

3 Data and Empirical Results

In this part, we successively exhibit the results of ARMA-GARCH model, KS test, Box-Ljung Test, static copula, goodness of fit for static copulas and time-varying copulas.

3.1 The Data

To analyze the volatility and dependence of Chinese tourists' arrivals to Thailand and Singapore we selected the monthly log growth rate of Chinese tourists' arrivals to Thailand and Singapore from 01/1997 to 12/2011. Table 1 provides the summary statistics for each rate of growth. As previously found in other studies, both growth rates demonstrate excess kurtosis and negative skewness. In addition, from the results of J-B test, we may find that they do not exhibit Gaussian distribution.

Table 1 Data description and statistics

	Thai	Singapore
Mean	-0.000826	0.008495
Median	0.019730	0.043021
Maximum	0.897066	0.923611
Minimum	-1.281747	-1.750788
Std. Dev.	0.343906	0.315213
Skewness	-0.761786	-0.944484
Kurtosis	4.480269	8.013719
Jarque-Bera	33.46747	212.8997
Probability	0.000000	0.000000

3.2 Results for ARMA-GARCH Model

To appropriate analyze the volatility and find the marginal distribution, as mentioned earlier, ARMA (2, 0)-GARCH model and ARMA (5, 0)-GARCH model are applied for tourist arrivals to Thailand and tourist arrivals to Singapore, respectively. Moreover, we assume that the margins are skewed-t distribution, the parameters of which are all significant that are shown in table 2 and table 3. The asymmetry parameters, γ, are significant and less than 1, exhibiting that the growth rates of tourist arrivals from China to Thailand and Singapore are skewed to the left. The $\alpha+\beta = 0.822$ in the table 2 and the $\alpha+\beta = 0.78$ in the table 3 that illustrate each growth rate of tourist arrivals has a long-run persistence of volatility, and the impact of unexpected shock to volatility lasts longer in the one to Thailand. For the values of α equal 0.21 and 0.23 whose size decides the short-run effect of unexpected factors, we can see that they are nearly the same, and have a large impaction for volatility.

Table 2 The results of tourism from China to Thai using ARMA (2, 0)-GARCH (1, 1) model

	parameters	std error	T statistics	P value
ar1	-0.23078	0.08753	-2.637	0.00837 ***
ar2	-0.2345	0.07372	-3.181	0.00147 ***
ω	0.02412	0.01324	1.822	0.06848*
α	0.21016	0.11636	1.806	0.07089*
β	0.61215	0.15289	4.004	6.23e-05 ***
γ	0.74744	0.07863	9.505	<2e-16 ***
υ	4.86494	2.04899	2.374	0.01758 **

[a] Log likelihood: -38.55479

Table 3 The results of tourism from China to Singapore using ARMA (5, 0)-GARCH (1, 1) model

	parameters	std error	T statistics	P value
ar1	-0.435762	0.08301	-5.25	1.53e-07 ***
ar2	-0.51074	0.089365	-5.715	1.10e-08 ***
ar3	-0.184694	0.090175	-2.048	0.040544 **
ar4	-0.27211	0.075867	-3.587	0.000335 ***
ar5	-0.232279	0.068239	-3.404	0.000664 ***
ω	0.017024	0.006959	2.446	0.014432 **
α	0.236912	0.130727	1.812	0.069945*
β	0.543044	0.16991	3.196	0.001393 ***
γ	0.786549	0.097996	8.026	1.11e-15 ***
υ	6.656059	3.652893	1.822	0.068435*

[a] Log likelihood: 5.1938

3.3 Results for KS and Box-Ljung Test

There exists a precondition for using any copula, which is the marginal distribution that must be uniform (0, 1), if not, the misspecified model for the marginal distribution may cause incorrect fit copulas. Further details are shown in Patton [1]. Thus, testing for marginal distribution model misspecification is a critical step in constructing multivariate distribution models using copulas. Therefore, we present the Box-Ljung tests for serial independence of the probability integral transforms, u_1 and u_2, and the Kolmogorov-Smirnov (K-S) tests of the density specification. The results of KS test and Box-Ljung Test are shown in table 4. It is very clear that each series accepts null hypothesis, which means both margins are uniform distribution. The second part of table 4 shows the Box-Ljung Test that tests serial independence of the first four moments, and all of them accept the null hypothesis at the 0.10 level. Therefore, the margins that we assumed are satisfied the two preconditions, uniform and serial independence.

Table 4 KS Test for Uniform and Box-Ljung Test for Autocorrelation

KS Test			
	Statistic	P value	Hypothesis
Margins 1 (from China to Thai)	0.0056	1	0 (acceptance)
Margins 2 (from China to Singapore)	0.014	1	0 (acceptance)
Box-Ljung Test			
	Moments	X-squared	P-value
Margins 1	First moment	14.1482	0.1663
	Second moment	6.4853	0.773
	Third moment	10.63	0.3871
	Fourth moment	8.4774	0.5823
Margins 2	First moment	14.1975	0.1642
	Second moment	12.9168	0.2284
	Third moment	10.5897	0.3904
	Fourth moment	13.7387	0.1852

3.4 Results for Static Copulas and Goodness of Fit Test

Just as mentioned in section 2.3, we use Gaussian copula, T copula, Clayton copula, Gumbel copula, Frank copula, BB1, BB7, BB8 and rotate copulas to fit the tourism data, and the methods we select as a criterion to appraise which copula is the best fitness are AIC and BIC. The table 5 and 6 show the estimated values of one parameter copula and two parameters copula, respectively, and the two tables provide the Kendall's tau, AIC and BIC as well. Those tables show that all parameters in copulas are significant, in term of AIC and BIC, and the Gaussian dependence structure exhibits better explanatory ability than other dependence structures. The parameter of the Gaussian copula is Pearson correlation that equals 0.7841, explaining that there is a strong linear correlation between them. In addition, the nonlinear correlation coefficient, Kendall's tau, is also very high that equals 0.5737 in Gaussian copula. Furthermore, the BB1 copula is also fit very well, although the Gaussian copula is the best one. BB1 copula may capture lower tail and upper tail whose values are 0.46 and 0.51, respectively. So, it illustrates that there exists tailed correlation, and upper tail correlation is higher than lower tail correlation. The difference between the correlation coefficient of upper tail and lower tail is not so large, meaning that Thailand arrivals and Singapore arrivals from China have a big probability appearing extreme value at the same time, which means Thailand and Singapore have the same peak tourist season and tourist off-season.

Although we selected a copula of the best fitness to describe the dependence structure, one crucial problem for copulas is to determining whether the copula we selected using AIC and BIC appropriately models the dependency structure. Therefore, a kind of goodness of fit is needed. In this study, the two tests based on Kendall's transform are applied, which contain Cramer-von Mises and Kolmogorov Smirnov test. More details is shown in Genest and Rivest [4], Wang and Wells [17] and Genest and Quessy [6]. Table 7 shows the testing results of goodness of fit by providing the probabilities of CvM and KS. It is not difficult find that half of copulas have not rejected the null hypothesis, which means they can appropriately model the dependency structure, except for Clayton, Gumbel, R-Clayton 180°, R-Gumbel 180°, BB7, BB8 and R-BB8 180°, which all reject the null hypothesis at the 5% level. Therefore, on the one hand, we guarantee the Gaussian copula is the optimal choice. On the other hand, it appropriately models the dependency structure as well.

Table 5 The estimated results of one parameter copulas

Copulas	parameters	std error	Kendall'tau	AIC	BIC
Gaussian	0.7840656	0.0227466	0.5737149	-166.817	-163.635
Clayton	1.653462	0.1835102	0.452574	-126.358	-123.177
Gumbel	2.133061	0.128922	0.5311901	-148.228	-145.046
Frank	6.992071	0.6211994	0.5619072	-149.538	-146.357
Joe	2.436203	0.1809137	0.4384843	-114.715	-111.533
R-Clayton 180	1.600283	0.1819629	0.4444882	-120.221	-117.039
R-Gumbel 180	2.145593	0.1293038	0.5339283	-151.418	-148.236
R-Joe 180	2.479488	0.1826424	0.4455462	-120.473	-117.291

Table 6 The estimated results of two parameters copulas

Copulas	parameters	estimated	std error	Kendall'tau	AIC	BIC
T	ρ	0.7816263	0.024029	0.571219	-163.905	-157.542
	DoF	30	NA			
BB1	θ	0.5262624	0.1777935	0.546033	-157.673	-151.31
	δ	1.743922	0.1583411			
BB7	θ	1.930817	0.2025542	0.515444	-151.916	-145.553
	δ	1.234913	0.2190782			
BB8	θ	6	1.790987	0.538269	-143.088	-136.725
	δ	0.717612	0.122668			
R-BB1 180	θ	0.4293901	0.16759	0.546686	-157.517	-151.154
	δ	1.816074	0.1613621			
R-BB7 180	θ	2.024321	0.2022186	0.515509	-151.499	-145.136
	δ	1.124618	0.2172944			
R-BB8 180	θ	6	1.923104	0.54279	-145.216	-138.853
	δ	0.7236699	0.1323331			

Table 7 The estimated results of one parameter copulas

Copulas	CvM	KS	Copulas	CvM	KS
Gaussian	0.72	0.56	T	0.55	0.62
Clayton	0.01	0.01	BB1	0.25	0.11
Gumbel	0	0.02	BB7	0.07	0.01
Frank	0.1	0.11	BB8	0.01	0.01
Joe	1	1	R-BB1 180	0.64	0.62
R-Clayton 180	0	0	R-BB7 180	0.12	0.14
R-Gumbel 180	0.03	0.16	R-BB8 180	0.01	0
R-Joe 180	1	1			

3.5 Results for Time-Varying Copulas

We still select the AIC and BIC as the criterion for choosing the best time-varying copulas. Table 8 shows the results of time-varying copulas, and it displays the parameters' value of time-varying copulas, standard error, AIC and BIC. Firstly, all parameters of time-varying copulas are significant, and the time-varying Gaussian copula is the best fitted since the lowest AIC and BIC. Secondly, we can see the autoregressive parameter β in time-varying Gaussian copula equals to 0.36, implying a low degree of persistence pertaining to the dependence structure between Thailand arrivals and Singapore arrivals from China.

Table 8 The estimated results of dynamic copulas

Dynamic copulas	parameters	ω	α	β	AIC	BIC
Gaussian	ρ	1.9527106	-0.12643	0.360156	-167.707	-164.612
	standard error	0.0477424	0.0603176	0.028947		
BB1	θ	-1.650144	1.0221166	4.263453	-154.595	-151.413
	standard error	0.0500146	0.0156214	0.269174		
	δ	-0.195336	-0.337295	-0.28646		
	standard error	0.0811855	0.0498284	0.098308		
R-BB1 180	θ	1.0544764	-0.751023	-1.6388	-149.285	-142.921
	standard error	0.0979645	0.0528295	0.486668		
	δ	0.2258694	0.2717021	0.353206		
	standard error	0.0946822	0.0581762	0.110524		
Joe	θ	0.4837019	0.3181147	-0.39841	-115.122	-111.94
	standard error	0.0291362	0.0099199	0.055507		
R-Joe 180	θ	0.1575325	0.6898117	0.380169	-132.488	-129.306
	standard error	1.04E-06	6.02E-06	2.43E-08		

4 Policy Implication

There exist strong linear and nonlinear correlation between Thailand arrivals and Singapore arrivals from China, which implies that policymakers should strengthen communication and cooperation of tourism industry and encourage a series of favorable policies or classical tour routes, etc. During tourist off-seasons, the two countries further should enhance cooperation, getting rid of the stale and bringing forth the fresh, thereby, promoting the development of tourism.

Unexpected shocks have a long run impaction for Thailand and Singaporeparticularly for Thailand. The two governments should be aware that no matter which country was affected by the negative impact, it will cause the damage of tourism for the other country as well, which is a loss for both sides. Therefore, the two countries should depend on and help each other. The package tour promotion between these two countries must be implemented immediately to help reduce the negative impact from the crises. This could be a win-win strategy.

5 Conclusions

In this paper, for the analysis of volatility, ARMA-GARCH models were used by assuming that standardized innovation was skewed-t distribution. Meanwhile, we discuss how traditional tests for marginal distribution, using the Kolmogorov-Smirnov and Box-Ljung tests, can be implemented to see if the underlying assumptions are satisfied. In addition, fifteen kinds of static copulas were used to analyze dependence between tourist arrivals to Thailand and Singapore from China. Another point is that we applied time-varying copulas that described the dynamic Kendalls tau. Overall, our empirical results show that the time-varying Gaussian copula has the highest explanatory power of all the dependence structures between tourist arrivals to Thailand and Singapore from China in terms of AIC and BIC values.

References

1. Patton, A.J.: Modelling asymmetric exchange rate dependence. International Economics Review 47(2), 527–556 (2006)
2. Almeida, C., Czado, C.: Efficient Bayesian inference for stochastic time-varying copula models. Computational Statistics and Data Analysis 56, 1511–1527 (2012)
3. Fernandez, C., Steel, M.F.J.: On bayesian modeling of fat tails and skewness. J. Am. Statist. Assoc. 93, 359–371 (1998)
4. Genest, C., Rivest, L.-P.: Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88(423), 1034–1043 (1993)
5. Wu, C.-C., Chung, H., et al.: The economic value of co-movement between oil price and exchange rate using copula-based GARCH models. Energy Economics 34(1), 270–282 (2012)
6. Genest, C., Quessy, J.-F., et al.: Goodness-of-fit procedures for copula models based on the probability integral transformation. Scand. J. Statist. 33(2), 337–366 (2006)
7. Chan, F., et al.: Modelling multivariate international tourism demand and volatility. Tourism Management 26, 459–471 (2005)
8. Wang, F.: A study on the volatility of international inbound tourism demand in China based ARIMA and GARCH model. Economic Vision 7 (2009)
9. Manner, H., Reznikova, O.: A survey on time-varying copulas: Specification, simulations and application. Econometric Reviews 31(6), 654–687 (2012)
10. du Preez, J., et al.: Univariate versus multivariate time series forecasting: an application to international tourism demand. International Journal of Forecasting 19, 435–451 (2003)
11. Coshall, J.T.: Combining volatility and smoothing forecasts of UK demand for international tourism. Tourism Management 30, 495–511 (2009)
12. Wang, K., Chen, Y.-H., et al.: The dynamic dependence between the Chinese market and other international stock markets: A time-varying copula approach. International Review of Economics and Finance 20(4), 654–664 (2011)
13. Nelsen, R.: An introduction to copulas. Springer, New York (2006)
14. Sklar, A.: Fonctions de Rparitition n dimensions et Leurs Marges. Publ. l'Inst. Stat. l'Univ. Paris 8, 229–231 (1959)
15. Wang, W., Wells, M.T.: Model selection and semiparametric inference for bivariate failure-time data. Journal of the American Statistical Association 95(449), 62–72 (2000)
16. Liu, Z., et al.: A correlation analysis of the tourist flows and trade flows. Tourism Forum 2(6) (2009)

A Quantile Regression Analysis of Price Transmission in Thai Rice Markets

Aree Wiboonpongse, Yaovarate Chaovanapoonphol, and George E. Battese

Abstract. We analyze the price efficiency by investigating the backward and forward transmissions of prices from the wholesale markets for white rice to and from the farm level. We compare the price transmissions at different price levels. This study utilizes monthly price data at farm and wholesale markets of white rice during 1997 to 2012. The effect of the structural change due to the price insurance program is considered. The quantile cointegration regression model is used in the empirical analysis. The findings reveal the existence of asymmetry in most aspects of the price transmission of farm-white rice markets. In the backward transmission (from wholesale to farm), asymmetric adjustment did occur but at only low prices and it was in favor of farmers. Asymmetry persists in the price transmission in the forward direction (farm to wholesale markets) at all price levels. The price insurance program had statistically significant effects on elasticities of transmission at almost all price levels but their effects were numerically very small. As backward transmission is in favor of farmers, an appropriate credit program is strongly recommended to manage timely sale of paddy.

1 Introduction

Rice is both the staple food and a major export-earning crop of Thailand. As being the top rice export country, Thailand shares approximately 22% of the world

Aree Wiboonpongse · Yaovarate Chaovanapoonphol
Department of Agricultural Economics and Extension,
Faculty of Agriculture,
Chiang Mai University,
Chiang Mai, Thailand
e-mail: aree.w@cmu.ac.th, yao_chao@yahoo.com

George E. Battese
UNE Business School, University of New England,
Armidale, NSW 2351, Australia
e-mail: gbattese@une.edu.au

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 295–305.
DOI: 10.1007/978-3-642-35443-4_21

rice market in 2011/2012, and serves countries in Asia and Africa [1]. Together with other economic crops, livestock and fishery, agriculture has enabled Thailand to become a world-leading exporter in many food products. As a consequence, Thai governments under various administrations have determined to promote the country as the "Kitchen of the World". Although Thailand has so far performed well in agricultural exports, Thai farming households incurred greater average debt from 43,415 baht in 2001/02 to 63,901 baht in 2004/05 [2]. without a corresponding increase in productive assets. The deepening indebtedness was caused primarily by the price instability and low realized prices that left no profit margins for the farmers. Beside farmers, non-government organizations and policy makers dealing with farm income and farm indebtedness highlight the unsatisfactory patterns of farm prices, marketing margins and the price transmission along the supply chain. Regular protests by farmers (rice, maize, cassava and other farmers) have occurred whenever farm prices drop to unacceptable levels. Consequently, the Thai government has implemented price support policies and mortgage programs to assist farmers. The recent measures for rice are the paddy mortgage and the price insurance programs.

Theoretically, prices of commodities play a major role in resource allocation and output distribution. Misallocation due to asymmetric price transmission therefore has impact on welfare and, thus, has policy implications. With asymmetric price transmission, buyers are not benefited from a price reduction or sellers are not receiving benefits from a price increase. It leads to a different distribution of welfare due to changes in the timing and the size of welfare associated with the price changes [3]. Peltzman (2000) [4] reported that of 282 products, including 120 agricultural and food products, asymmetric price transmission was the rule. However, other economists [5, 6]. caution that the findings could be due to method used that lead to excessive rejection of symmetry of price transmission. Because asymmetric price transmission reflects market failure (price inefficiency) due to risk, collusion or market intervention, several studies have been carried out on this matter. For example, the European Parliament [7]. commissioned a study at the European level. Reports of the FAO and the OECD cover a wide range of studies related to price transmission in various countries. Meyer and Cramon-Taubadel (2004) indicate two dominant causes of asymmetric price transmission, namely, the existence of non-competitive markets and adjustment costs. Other possible causes include policies interventions, asymmetric information and inventory management.

Our study concentrates on price transmission for paddy and wholesale white rice in the Thai economy. We address two questions: (1) is the rice market efficient such that symmetric price transmission prevails? And (2) has the rice price insurance program affected the farm price of paddy? To answer the former, we investigate how local farm prices of paddy influence the wholesale prices of white rice (forward price transmission) at the same magnitude as the backward transmission (from wholesale to farm prices).

2 Rice Sector in Thailand

Rice is of primary economic importance for Thailand as a staple food, major agricultural income source (66% of farming households grow rice) and a predominant export item (40-45% of domestically produced rice is exported), Thailand being the world's largest exporter of rice since 1981. However, problems in the Thai rice sector have increased over the more than 30 years of efforts by the state to assist farmers through intervention in the rice market. The overly high floor prices for the loan or rice mortgaging program since 2000, particularly in 2004/5 when the floor prices were 20-30% above market levels, have led to the problem of non-redemption of the collaterized rice [8]. The accumulated problems in the quite lengthy time span have rendered the state a heavy burden to shoulder in terms of capital for running the rice mortgaging. Rice was used as collateral in the loan program during 1986–1996 to reduce effective supply but this was not followed by increases in the market price because, at the same time, the traders delayed their purchase schedules [9]. The Siam Intelligence Unit (2009) [10]. cited that the Thailand Development Research Institute found that about 40% of the benefit of the rice collateral credit policy was received by farmers, especially those better-off cultivators in the irrigated areas. Among middle agents, the exporters enjoyed the highest benefit, being about 24%, while rice millers and warehouse owners received about 14% and 4%, respectively.

Although the rice mortgaging scheme is a government measure intended for dealing with rice price fluctuations and increasing farm incomes, it has been regarded as market price intervention and action that distorts the market mechanism, so that farm prices do not reflect the actual functioning of the market. These policy actions are designed to ensure more reasonable prices for farmers. In international trade, the gap between the rice export price for Thailand and Vietnam in the last couple of years widened from US$10-20 to US$100-200 per ton. Furthermore, the compliance with the WTO agreements to open markets for free trade in rice in 2010 is a factor for Thailand to revise its rice market intervention measure because the high domestic price from the rice mortgaging program has invited enormous rice movements from its neighboring countries. The Thai government had to increase its budget substantially to meet the demand for rice mortgaging. The undesirable outcomes led to a major policy change when the cabinet passed a resolution in its meeting on 21 July 2009 to replace the rice mortgaging program with a Price Insurance Program that existed during July – October 2009 and August 2010 to July 2011. The underlying principles that the National Rice Policy Board has established to aid farmers by this price support measure are that farmers will be assured the minimum price or target price and they will be compensated for the difference between the market price and the target price when the former is lower. Along with the price insurance program, the government encourages trading in the agricultural futures exchange to promote competition and efficient pricing signals. However, the rice mortgage program replaced price insurance program after only two years of operation.

To enable a better understanding of the price system, marketing channels from the levels of local paddy (rice grain) procurers to rice millers, retailers and exporters are described. The two subsectors of paddy and rice markets are closely integrated. Local traders serve as assemblers as well as brokers (acting on behalf of millers). Prior to the rice mortgage scheme, both types of assemblers handled approximately 80% of paddy before delivery to rice mills. (Most rice millers are located in producing areas but large rice millers usually operate on a regional scale while some rice milling businesses also export their rice products.) The flow of paddy changed substantially after implementation of the mortgage scheme. For example, in 2009, 33.7% was mortgaged to the Bank of Agriculture and Agricultural Cooperatives [11]. Most paddy then was stored by millers under contracts with the government. Before the new harvest, the government opened an auction for the stored paddy. Rice millers, exporters and large traders were involved in the auction for the paddy. Apparently, the final auction prices were far below desired levels. The prices were determined by prevailing market conditions and the price expectations of traders. Once the mortgaged paddy was sold, it was milled mostly by large millers. In the milled rice subsystem, commission agents have an important role in matching supply from millers to the demands of wholesalers and exporters whereas the international brokers conducted the same function for exporters and overseas importers. At the final stage in the marketing system, approximately 55% of the rice was stored and consumed domestically, and 45% was for export [12, 13]. As described, the rice marketing system in Thailand is quite complex, the private sector having a key role in both export and domestic markets. The pricing system was inevitably influenced by export and wholesale markets. With the different price support measures, it is questionable how pricing systems transmit price changes from one section to another and how fast the price adjustment takes place.

3 Literature Review

Efficiency of the marketing of agricultural products has always been an issue of interest on the presumption that middle agents are primarily responsible for any inefficient market practice and depressed farm prices in a country due to their market power. One way to prove the concentration of market power is by analysis of the efficiency of the market system addressing market integration and price transmission among various sub-markets. Study of market integration is, therefore, useful for forecasting prices at different market levels; for example, for estimating the magnitude of the effect of urban market price movements on the rural market price patterns [14, 15].

Furthermore, an efficient market can help prevent problems in arbitrage trade from business losses. Knowledge about the state of integration in the domestic market can also help middle agents in their selection of markets by arbitrage measures to reduce business risk [16]. Market integration is fundamental for economic growth by enabling efficient resource allocation [15]. and leading to specialization in production because of the principle of comparative advantage and, eventually, to production expansion, economies of scale, market competitiveness, and market efficiency.

Price transmission is an important area of research in agricultural economics. Unexpected price movements affect welfare of producers and consumers which is a concern for policy makers. Inadequate market information may cause market distortion, inefficient pricing and, thus, welfare distribution. A large body of research on price transmission has recently been undertaken, especially in agricultural and food markets (e.g. [3]; [17]; [18]). Most of these studies focus on existence of asymmetric transmission. Meyer and Cramon-Taubadel (2004) report that, for 48% of cases, the hypothesis of symmetry of price transmission have been rejected using different analytical methods. The nature of price transmission seems to depend on the particular products and local circumstances.

Frey and Manera (2007) state that "asymmetry is very likely to occur in a wide range of situations and econometric models. Most research in this area did not explore causes of transmission performance". Bakucs, Falkowski and Fertö (2012), in their study on the agro-food sector, concluded that market power variables (entry barriers, size, regulation) farmers' bargaining power and manufacturing turnover are significant causes of asymmetry.

4 Data and Methodology

This study uses farm price of paddy rice (5% broken) and wholesale prices of white rice (5% broken) in Bangkok. Data are obtained from the Office of Agricultural Economics. Monthly data are available from January 1997 to June 2012. In our study, we estimated both error correction and quantile regression models to investigate the nature of the price transmission between the two rice markets. We suspect that our rice price data may contain some extreme values and exhibit non-normal distribution, so that the quantile regression model is particularly relevant for our study. Because we use time-series data, we adopt the quantile cointegration regression (QCR) model, proposed by Xiaov (2009) [19].[1] For backward price transmission, we estimate the QCR model that is defined by

$$p_{f,t}(\tau|\, F_t) = \alpha(\tau) + \beta_t(\tau)\, p_{w,t} + \lambda_t(\tau)\, G \times p_{w,t} + \sum_{j=-K}^{K} \pi_{j,t}(\tau)\, \Delta\, p_w + \sum_{j=-K}^{K} \gamma_{j,t}(\tau)\, \Delta\, p_{w,t-j} + \varepsilon_t$$

where p_f is the logarithm of the farm price of paddy; τ denotes the quantile value (where $0 < \tau \leq 1$); p_w is the logarithm of the wholesale price of white rice (5%); G is

[1] In our study, we estimated both error correction and quantile regression models but only the quantile Regression results are presented in this paper. The working paper, Wiboonpongse, Chaovanapoonphol and Battese (2012), presents and compares the results for both econometric models.

a dummy variable representing the period of the price insurance program[2]; K (=2 in our study) is the number of years for leads and lags to absorb endogeneity; and ε_t is the normal random error having mean zero and constant variance. For investigating forward price transmission, the two price variables, $p_f$and p_w, are reversed so that p_w is the dependent variable and p_f is the independent variable.

5 Empirical Results

Summary statistics for the data used in our analysis are presented in Table 1. These indicate that the mean of the paddy rice is almost half that for wholesale white rice but the variability is considerably less than for white rice. The skewness and kurtosis measures obtained indicate departure from normality. Using the Shapiro-Wilkstatistic, we the null hypothesis of normality would be rejected for both price series at the 5% level of significance.

These prices move in concert during the period of the study, although the wholesale prices are more variable during periods of higher prices. Increases in the prices are associated with rice export. For example the domestic price spike in 2008 followed sharp rise in the world price due to global rice crisis. Price margins are much higher in some periods than in others which indicate variation in price transmission or different regimes (due to market intervention).

Table 1 Net present value of all scenarios for the duration of 30 years.

Price Variable	Mean	Std.Dev	Min	Max	Skewness	Kurtosis	Shapiro-Wilk
Paddy rice (baht/100kg)	681.46	201.15	438.40	1,325.90	1.02	3.26	0.888
Wholesale white rice 5% (baht/100kg)	1,288.64	351.31	855.00	2,794.61	1.46	5.79	0.877

The empirical results for the QCR model, obtained using EViews 6, [20]. are reported in Table 2 for the paddy and wholesale white rice prices. For the backward and forward relationships between the farm prices of paddy rice and the wholesale white rice prices, we are particularly interested in whether there is equality of the slopes for the different quantiles and the directional effects of the price insurance program. On average, the farm price responds to the wholesale price with elasticity of 0.995 (coefficient of the OLS estimate of p_w in the first column of Table 2), which is not significantly different from one. From the quantile regression results,

[2] During July2009, the Thai government replaced the rice mortgage program with the price guarantee program. The dummy variable, G, is used to capture the effect of this structural change on the transmission.

Table 2 Quantile cointegration regression results for paddy and wholesale white rice prices

Variable	OLS	τ=0.10	τ=0.25	τ=0.50	τ=0.75	τ=0.90
Backward transmission						
Constant	-0.62**	1.8	-0.72**	-0.87**	-0.82***	-1.45***
	(0.26)	(1.20)	(0.33)	(0.33)	(0.27)	(0.31)
p_w	0.995***	0.63***	1.005***	1.032***	1.032***	1.129***
	(0.037)	(0.17)	(0.047)	(0.048)	(0.038)	(0.044)
$G \times p_w$	0.000080***	0.000170***	0.000090***	0.000060***	0.000040**	0.000030
	(0.000020)	(0.000040)	(0.000010)	(0.000020)	(0.00001)	(0.000020)
$\Delta p_{w,\, t-1}$	0.00002	0.00048	-0.00002	0.00007	-0.00015	-0.00007
	(0.00017)	(0.00043)	(0.00011)	(0.00040)	(0.00030)	(0.00040)
$\Delta p_{w,\, t-2}$	0.00002	-0.00004	0.000010	-0.00004	0.00018	0.00002
	(0.00011)	(0.00017)	(0.000060)	(0.00014)	(0.00022)	(0.00025)
$\Delta p_{w,\, t-1}$	0.00013	-0.00028	0.00002	0.00012	0.00027	0.00024
	(0.00017)	(0.00074)	(0.00010)	(0.00017)	(0.00021)	(0.00040)
$\Delta p_{w,\, t-2}$	-0.00008	-0.00007	0.000040	-0.00008	-0.00022*	-0.00012
	(0.00010)	(0.00039)	(0.000050)	(0.00034)	(0.00013)	(0.00021)
Pseudo R^2	0.876	0.449	0.604	0.691	0.744	0.740
Wald Test	-			40.71806**		
Forward transmission						
Constant	1.90***	1.65***	1.60***	1.47***	2.03***	2.67***
	(0.18)	(0.26)	(0.21)	(0.20)	(0.28)	(0.35)
p_f	0.802***	0.828***	0.841***	0.868***	0.789***	0.693***
	(0.028)	(0.043)	(0.033)	(0.031)	(0.043)	(0.054)
$G \times p_f$	0.00008***	-0.00002	-0.00004*	-	-	-0.00007
	(0.00003)	(0.00002)	(0.00002)	0.00008***	0.00009***	(0.00005)
				(0.00002)	(0.00003)	
$\Delta p_{f,\, t-1}$	0.00065**	0.00016	0.00052	0.00007	0.00078*	0.0018
	(0.00030)	(0.00093)	(0.00051)	(0.00037)	(0.00043)	(0.0033)
$\Delta p_{f,\, t-2}$	0.00019	0.00026	0.00002	0.00031*	0.00038*	0.0006
	(0.00017)	(0.00049)	(0.00034)	(0.00018)	(0.00018)	(0.0016)
$\Delta p_{f,\, t-1}$	-0.00072	-0.00027	-0.00075	-0.00026	-0.00060	-0.0013
	(0.00047)	(0.00078)	(0.00065)	(0.00051)	(0.00066)	(0.0021)
$\Delta p_{f,\, t-2}$	-0.00004	-0.00007	0.00006	-0.00022	-0.00017	-0.00066
	(0.00033)	(0.00073)	(0.00046)	(0.00035)	(0.00039)	(0.00082)
Pseudo R^2	0.884	0.655	0.664	0.689	0.680	0.667
Wald Test	-			37.69440*		

Notes: Coefficients that are significant at the 1%, 5% and 19% levels are denoted by ***, ** and *, respectively.

the Wald test statistic to test the null hypothesis of common slope parameters for the different quantiles is rejected at the 5% level of significance. The quantile slopes increase from 0.63 (for τ=0.10) to 1.13 (for τ=0.90), but both coefficients are significantly different from one. For quantile values between τ=0.25 and τ=0.75, the coefficients are not significantly different from one. Displayed equations or formulas are centered and set on a separate line (with an extra line or halfline space above and below). Displayed expressions should be numbered for reference. The numbers should be consecutive within each section or within the contribution, with numbers enclosed in parentheses and set on the right margin.

These results can be interpreted that changes in the wholesale white rice price at low prices result in less than proportional price changes for paddy rice, but, at higher prices, the changes in paddy prices to farmers are expected to be greater. The positive coefficients of the product of the dummy variable for the price insurance program and the wholesale price of white rice are generally significantly greater than zero but their numeric values are quite small. These results indicate that the price changes for paddy rice to changes in the wholesale white rice price are quite small during the period of the rice price insurance program.

On the forward transmission from the farm to the wholesale white rice market, the average transmission, estimated by OLS is 0.80, while the quantile regression estimates first increase from 0.83 (for τ=0.10) to 0.87 (for τ=0.50) and then decrease to 0.69 (for τ=0.90). The null hypothesis of common slope coefficients for the different quantile regressions in the forward transmission is rejected at the 10% level of significance using the Wald test statistic. The negative coefficients of the product term the dummy variable for the price insurance program and the farm price indicate significant effects of the program in lowering the prices for wholesalers, except when τ=0.10 and τ=0.90. Although these effects are statistically significantly they are quite small numerically for all quantile values. When the farm price increases, wholesale price adjusts to increases wholesale price less than 1% (0.83 to 0.87).

6 Conclusions

The QCR model indicates that there are mostly symmetric responses in the backward transmission (wholesale to farm prices) but there are significant asymmetric responses in forward transmission (farm to wholesale prices). Hence, we assert that wholesalers in the white rice market are more sensitive to price increases than price decreases. At only very low price levels (τ= 0.10) farm prices are inelastic to changes in the price of wholesale white rice. This means that when prices are too low (say, 4,834 baht/ton, as in 2000/01) local buyers were reluctant to alter their buying price from farmers but they were more willing to adjust prices at higher price levels.

We conclude from the results of the QCR model to conclude that at very low price, farm price adjusts only 63% to changes in the wholesale white rice market, regardless of upward or downward price changes, in order to maintain price stability. But 100% to 113% adjustments occur at higher prices. This means that the farm price would be more volatile than wholesale price since elasticities in the forward direction only 69% to 87%. At the highest price level, wholesalers of white rice become least responsive to farm price changes. This means that they try to maintain price stability in the wholesale market as expected that retailers are normally reluctant to increase retail price to the very high level. This does imply that wholesalers do not earn the highest margins when there are opportunities to do so. They acted to maintain more stable margins while local buyers (at farm level) take the riskier option (in backward direction with high elasticity). Comparing the price guarantee program to the rice mortgage program, the farmers enhance (wholesalers discourage) higher elasticity of transmission in the backward direction (forward direction) of the farm-white rice market. This means, farmers react stronger than wholesalers to price changes. However, the effects of the program are very small despite being statistically significant in almost all cases.

The results about the forward price transmission (farm to wholesale market of white rice) are important for policy planning. As local demand and supply have much effect on the price at the other end, storage of paddy to sell with good timing is needed for raising farm prices of paddy. Financial assistance through credit programs ought to be redesigned to replace existing programs that are apparently unsuccessful. This research supports the notion of perfect price transmission from the wholesale white rice market in favor of farmers. But there is no guarantee of absolute exploitation when prices are above 4,834.05 baht per ton. According to Dawe et al. (2008), marketing margins in the Thai rice market (central Thailand) were low due to marketing innovations. Under free trade, world prices and domestic prices are identical (adjusted to the same point in the marketing chain). This left zero return to management (of marketing agents in Thailand in 2003) when gross margin equals marketing cost. [21]. On the other hand, the latest statistics show that, on average, Thai farmers earned negative profits two out of three years during 2009/10 to 2011/12. The cost per ton rose from 8,349 to 9,359 and 10,399 baht. The corresponding prices were 9,029 baht 8,600 and 10,289 baht. [22]. This certainly requires updated marketing margins to ensure no excess profits in the Thai rice marketing system. If so, coupled with perfect price transmission aforementioned, a policy to lift technical efficiency in rice production is clearly of high priority. Allocative efficiency is also relevant for Thai farmers so as to adjust the use of fertilizer to the appropriate rate because the prices of fertilizer in Thailand are usually higher than most countries in ASEAN.

Acknowledgments. The project is jointly supported by the Thailand Research Fund and Chiang Mai University.

References

1. USDA, Grain: World markets and trade. Available via DIALOG (2011) http://www.fas.usda.gov/grain/circular/2011/05-11/grain.pdf (cited July 12 , 2012)
2. Office of Agricultural Economics, Changes in the structure of household debt and the size of loans for farmers (in Thai). Available via DIALOG (2007), http://www.oae.go.th (cited March 25, 2010)
3. Meyer, J., Cramon-Taubadel, S.V.: Asymmetric price transmission: a survey. Journal of Agricultural Economics 50, 581–611 (2004)
4. Peltzman, S.: Prices Rise Faster than They Fall. The Journal of Political Economy 108, 466–502 (2000)
5. Gauthier, W.M., Zapata, H.: Testing Symmetry in Price Transmission Models. Louisiana State University, Department of Agricultural Economics & Agribusiness, Working Paper (2001)
6. Cramon-Taubadel, S.V., Meyer, J.: Asymmetric price transmission: Fact or artefact? University of Gottingen, Institute of Agricultural Economics, Working Paper (2000)
7. Agra CEAS (2007), The gap between producer prices and the prices paid by the consumer. Available via DIALOG, http://www.ceasc.com/Images/Content/2326%20Report.pdf (cited November 9, 2011)
8. Agricultural Futures Exchange of Thailand, From the rice mortgage program to the rice strategies program (in Thai). Available via DIALOG (2009), http://www.afet.or.th/v081/thai/news/commodityShow.php?id=216 (cited August 7, 2009)
9. Sriboonchitta, S.: The effect of rice harvest collateral credit policy on rice price: Statistical evidence during 1979-1997. Chiangmai University Journal of Economics 4, 119–128 (2000)
10. Siam Intelligence Unit, TDRI's comments on the market distortion and reduction of competition of the rice mortgage program (in Thai). Available via DIALOG (2009), http://www.siamintelligence.com/tdri-to-criticise-rices-pawn-policy/ (cited August 7, 2009)
11. Department of Internal Trade, Report on mortgage program for major paddy of 2008-09 crop-year (in Thai). Available via DIALOG (2010), http://www.dit.go.th/uploads/51-52_061552.pdf (cited November 9, 2011)
12. Boonsomjai, S.: Logistics and supply chain of Thai rice (in Thai). Available via DIALOG (2009), http://logisticssharing.com (cited November 9, 2011)
13. Ruamtrakul, D.: Physical flow of rice (in Thai). Available via DIALOG (2008), http://www.riceproduct.org (cited August 7, 2009)
14. Okoh, R.N., Egbon, P.C.: A test of market integration: The case of Nigerian staple foodstuffs. African Journal Economic Policy 9, 117–139 (2002)
15. Khang, M.N., Ndifon, H.M., Odok, G.N.: Price transmission and integration of cocoa and palm oil market in Cross River State, Nigeria: Implications for rural development. Agricultural Journal 2, 457–463 (2007)
16. Falsafian, A., Moghaddasi, R.: Spatial integration and asymmetric price transmission in selected Iranian chicken markets. Paper Prepared for Presentation at the 12th European Agricultural Economics Congress, Gent, Belgium (2008)
17. Frey, G., Manera, M.: Econometric models of asymmetric price transmission. Journal of Economic Surveys 21, 349–415 (2007)

18. Boktcs, L.Z., Fackowski, J., Fertö, I.: What causes asymmetric price transmission in agro-food sector? Meta-analysis perspective. Available via DIALOG (2012), http://www.aes.ac.uk (cited July 12, 2012)
19. Xiao, Z.: Quantilecointegrating regression. Available via DIALOG (2009), http://www.bc.edu/ec-p/wp708.pdf (cited July 12, 2012)
20. EViews 6, Econometric Software Package. Quantitative Micro Software, LLC (2008)
21. Dawe, D.C., Moya, P.F., Casiwan, C.B., et al.: Rice marketing systems in the Philippines and Thailand: Do Large numbers of competitive traders ensure good performance? Food Policy 33, 455–463 (2008)
22. Office of Agricultural Economics, Basic Information in Agricultural Economy in 2012 (in Thai). Available via DIALOG (2012), http://www.oae.go.th/download/download_journal/fundamation-2554.pdf (cited July 7, 2012)

Analyzing Dependence Structure of Obesity and High Blood Pressure: A Copula Approach

Jing Dai, Cheng Zi, Songsak Sriboonchitta, and Zhanqiong He

Abstract. China's economy has experienced remarkable growth in past 20 years. With rapid economic growth, Chinese people have enjoyed significant nutritional improvements. Meanwhile, with the changes in lifestyle, dietary behavior and other aspects, the prevalence of obesity and high blood pressure has also increased quickly. The relationship between obesity, high blood pressure and risk of chronic non-communicable diseases is continuous and consistent. The higher BMI and blood pressure, the greater the chance of heart attack, stroke, kidney disease and etc. The objective of this paper is to analyze how socio-demographic and socioeconomic factors affect the prevelance of obesity and high blood pressure, and find the dependence structure between obesity and high blood pressure (HBP) with the help of copula functions. Computational results were obtained by R programme, and the results show that Frank copula model provides a better estimation than others. The empirical findings of this paper provide useful insights which can be expected to be of interest to public health sectors and local government in the formulation of health management policies especially on obesity, high blood pressure, and related chronic non-communicable disease.

Jing Dai · Zhanqiong He
Faculty of Economics, Chiang Mai University,
Chiang Mai, Thailand, and Kunming University of Science and Technology,
Yunnan, China
e-mail: jdai3126@gmail.com

Cheng Zi
Kunming University of Science and Technology,
Yunnan, China
e-mail: zicheng28@hotmail.com

Songsak Sriboonchitte
Faculty of Economics,
Chiang Mai, Thailand
e-mail: songsak@econ.cmu.ac.th

V.-N. Huynh et al. (Eds.): *Uncertainty Analysis in Econometrics with Appl.*, AISC 200, pp. 307–318.
DOI: 10.1007/978-3-642-35443-4_22

1 Introduction

China is currently undergoing a rapid economic development and demographic transformation. Since the reform and opening up in 1978, average living standards have experienced a sustained and rapid growth [11]. The quick increase in productivity has resulted in higher incomes and an ample food supply, and Chinese people have enjoyed significant nutritional improvements [3]. However, at the same time, some worrying trends have been detected. The prevalence of overweight, obesity and related chronic and non-communicable diseases has increased at an alarming rate in past 20 years [17, 18].

Many kinds of chronic non-communicable disease have close relations with obesity and high blood pressure, such heart disease, diabetes and stroke ect. According to WHO report, it is estimated that about 18.3 million people die every year from cardiovascular disease, for which obesity and high blood pressure are ranked as top two leading risky factors [19]. Obesity and high blood pressure burdens the health care system, strains economic resources, and has far reaching social consequences [7, 8]. Therefore, obesity and high blood pressure should be paid much more attention. It is urgent to identify those influencing factors related to obesity and high blood pressure, take some measures and make corresponding policies to control the prevalence of obesity and high blood pressure.

After reviewing literatures, we found that there are many papers concerned about obesity and HBP. He and Griffins investigated the influencing factors of obesity and HBP [8, 14], however, their studies tended to be qualitative and conceptual. Some of researchers analysed HBP and obesity only from the perspectives of medicine, and overlooked the influences from individual social demographic and socio-economic aspects [5]. In addition, there are some studies explored the relationship between obesity and HBP by using traditional regression method. For example, Wildman et al. examined the association of changes in hypertension diagnoses with changes in BMI among older Chinese adults by using logitic regression model [18]. However, we found there were few studies investigate dependence structure between HBP and obesity by copula approach. Therefore, more comprehensive studies should be conducted, and then much more useful information could be provided to policy maker to make effective intervention measures to prevent the prevalence of obesity and HBP, thus to reduce the possibility of the occurrence of chronic non-communicable diseases.

There are two objectives of this study, the first is to analyse how social demographic and socio-economic factors affect the prevalence of obesity and high blood pressure, and the second is to find the dependence structure between obesity and high blood pressure (HBP). Dependence models are constructed with the help of copula functions to explain the relationship between BMI and HBP.

The remainder of the paper is organized as follows. Section 2 describe the Archimedean copulas which are used in this study, and explained the model formulation in detail. Section 3 describes the data set we used. The estimation results are presented and discussed in Section 4. Finally, some concluding remarks are given in the last section.

2 Methods

Since statistical procedures based on normal distribution of the error terms can produce biased estimates when the normality assumption is violated. To conquer this problem, copula method is adopted in this study. Copula approach allows flexible marginal distribution, and separates the marginal distributions from the dependence structure, so the dependence structure is unaffected by the marginal distribution [10].

In 1959, the word ‘copula’ was first employed in a mathematical by Abe Sklar in describing the functions that ’join together’ one-dimensional distribution functions to form multivariate distribution functions [9]. Any joint distribution function has a copula representation in which dependence and marginals are separately specified.

Copula approach derives from Sklar’s theorem. Based on Sklar’s (1973) theorem [15], when F is a joint distribution function with marginal distribution functions $\mathrm{F_x}$ and $\mathrm{F_y}$, where exists a bivariate coupla C, that is:

$$\mathrm{F_{XY}}(\mathrm{x},\mathrm{y}) = \mathrm{C}\left(F_x(\mathrm{x}), F_y(\mathrm{y})\right) \tag{1}$$

If F_x and F_y are continuous, then C is unique. Otherwise, C is uniquely determined on $\mathrm{Ran}F_x \times \mathrm{Ran}F_y$. Conversely, if C is a copula and F_x and F_y are the cumulative distribution function, then the function F_{xy} defined by above equation is a joint distribution function with margins F_x and F_y . From Sklar’s theorem, it can be seen that a join distribution F_{xy} can be divided into its univariate marginal distribution F_x and F_y, and a copula C, which captures the dependence structure between the variables X and Y [12]. Therefore, copula models allow us to model the marginal distributions and the dependence structure of multivariate random variable separately.

There are many copula functions. In this study, Archimedean copulas are preferred because they are closed-form copulas that can be used to obtain the joint bivariate cumulative distribution function, and copulas in Archimedean copula family allows testing a variety of radially symmetric and asymmetric joint distributions. In addition, the closed-form nature of Archimedean copula family make itself easy to the implementation of a computationally straightforward maximum likelihood procedure for parameter estimation [16].

2.1 Archmedean Copulas and Their Properties

Archimedean copulas are a particular class of copula that includes several popular families. These copulas find a wide range of applications for a number of reasons. Firstly, the ease with which they can be constructed. Secondly, the great variety of families of copulas belongs to this class. Finally, the members of this class posses many nice properties.

Archimedean copulas are copulas whose form, in n dimensions, can be reduced to a single function, called a generator. The generator function is a strictly decreasing, convex and continuous function φ from $[0,1] \to [0-\infty]$ in a set [0,1], where

$\varphi(0) = \infty, \varphi(1) = 0$ and with inverse $\varphi^{-1}:[0,\infty] \rightarrow [0,1], \varphi^{-1}(0) = 1$ and $\varphi^{-1}(\infty) = 0$.Then we can generate bivariate Archimedean copulas as:

$$C(u,v,\theta) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{2}$$

Where the dependence parameter θ is embedded within the generator function. In Archemedean copulas families, different copulas are identified based on different forms of the generator function φ. In this paper, we will consider three types of Archimedean copulas which are Gumbel copula, Clayton copula and Frank copula.

Clayton copula was proposed by Clayton in 1978. It has the generator function $\varphi(t) = \frac{t^{-\theta_c}-1}{\theta_c}, \theta_c > 0$, then the form of Clayton Copula function is given by [2]:

$$C(u,v,\theta) = \left\{u^{-\theta} + v^{-\theta} - 1\right\}^{-1/\theta} \tag{3}$$

Where θ is the dependence parameter and $0 < \theta < \infty$. This copula is best suited for strong left tail dependence and weak right tail dependence. That is, it is best suited when HBP and BMI show strong tendencies to have low values but not high values together.

Gumbel copula is suited not only to random variables that are positively correlated, but to those in which right tail of each are more strongly correlated than left tail . It has a generator function given by $\varphi_\theta(\mathrm{t}) = (-\ln\mathrm{t})^\theta$. The form of the Gumbel copula is given by [6]:

$$C(u,v;\theta) = exp\left[-\left[(-lnu)^\theta + (-lnv)^\theta\right]^{\frac{1}{\theta}}\right] \tag{4}$$

Gumbel copula is well suited for the case when there is strong right tail or upper tail dependence. Thus, this copula would be applicable when individuals BMI and HBP show strong tendencies to have high values together.

The Frank copula is symmetric in its dependence structure [4]. The generator function is given by :

$$\varphi_\theta(t) = \ln\left[\frac{\exp(-\theta t) - 1}{\exp(\theta) - 1}\right] \tag{5}$$

and the corresponding coupla function is given by (Frank, 1979):

$$C(u,v;\theta) = -\frac{1}{\theta}ln\left(1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right) \tag{6}$$

Where the dependence parameter $\theta \in (-\infty, \infty)$. Franc copula is suitable for equal levels of dependency in the left and right tails, that is, when HBP and BMI show tendencies at both lower values and high values together.

The extent of concordance between HBP and BMI can be measured by Kendall's tau. According to Nelsen [12], let (X_1, Y_1) and (X_2, Y_2) be independent vectors of continuous random variables with joint distribution functions H_1 and H_2, with common margins F and G. Let C_1 and C_2 denote the copulas of (X_1, Y_1) and (X_2, Y_2). So that $H_1(x,y) = C_1(F(x), G(y))$ and $H_2(x,y) = C_2(F(x), G(y))$, then let tau denote the difference which is defined as the probability of concordence minus the probability of discordence:

$$\tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \tag{7}$$

Then we have (the derivation procedure refer to (Nelsen, 2006):

$$\tau = 4E(C(U,V)) - 1 \tag{8}$$

If X and Y be random variables with an Archimedean copula C generated by φ . The population version τ_c of Kendalls tau for X and Y is given by:

$$\tau_c = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \tag{9}$$

Then substitute sepcific generator function $\varphi(\cdot)$ of the different copula in Archimedean family, we can get the results of tau to measure the dependence between HBP and BMI.

2.2 Model Formulation

Since in this study, we only consider those who suffer from obesity and HBP. Then if BMI or blood pressure are over than normal range, they will be denoted as their true value, otherwise will be denoted as zero. Therefore, we see there are many zero values in health outcome. To model this kind of data, tobit model of Amemiya [1] will be applied.

Consider a set of latent dependent variables $(y_1^*, \ldots, y_n^*)$ with means $(x'\beta_1, \ldots, x'\beta_n)$, covariance matrix Σ, and probability density function (pdf) $f(y_1^*, \ldots, y_n^*)$, where x is a vector of conditioning variables, and $\beta_1, \ldots, \beta_n$ are parameters. Then Tobit model for observed dependent variables yi is as follows:

$$y_i = \begin{cases} y_i^* \ \text{if} \ y_i^* > 0 \\ 0 \ \text{if} \ y_i^* \leq 0, i = 1, 2, \ldots, n \end{cases} \tag{10}$$

The likelihood function for a sample observation with outcome $\left(0, \ldots, 0, y_{m+1}^*, \ldots, y_n^*\right)$, where the first m goods are zeros $(1 \leq m \leq n))$, is shown as follows:

$$L_c = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f(y_1^*, \ldots, y_m^*, y_{m+1}, \ldots, y_n) \prod_{j=1}^{m} dy_j^* \tag{11}$$

The likelihood function above is maximized by using conventional maximum likelihood procedures approach. To accommodate non-Gaussian error distributions, the bivariate distribution of $\left(y_i^*, y_j^*\right)$ is specified as a copula function which links the two corresponding marginal cumulative distribution functions (cdfs).

According to some literatures, for this type of data in my study, Burr distributions can be assumed [20]. Therefore, each variable y_i^* is assumed to follow the generalized log-Burr distribution with scale σ_i and location parameterized as $x'\beta_i$, with pdf as follows:

$$f\left(y_i^*;\kappa_i\right) = \sigma_i^{-1} e^{z_i}\left(1 + e^{z_i}/\kappa_i\right)^{-\kappa_i - 1}, -\infty < y_i^* < \infty \tag{12}$$

Where $\mathrm{z_i} = \left(\mathrm{y_i^*} - x'\beta_i\right)/\sigma_\mathrm{i}$ and κ_i is the skewness parameter. Variable y_j^* is also follow the same distribution with skewness parameter κ_i. However, these two variables $\left(y_i^*, y_j^*\right)$ need not have the same distribution and any continuous univariate distribution can be specified for the margins.

Next, we apply the copula approach to model the dependence among BMI and blood pressure. Let us denote the marginal distribution of two random variables $\left(y_1^*, y_2^*\right)$ as $F_1\left(y_1^*\right) = \Pr\left(Y_1^* \leq y_1^*\right)$ and $F_2\left(y_2^*\right) = \Pr\left(Y_2^* \leq y_2^*\right)$ and their joint cdf as $F\left(y_1^*, y_2^*\right) = \Pr\left(Y_1^* \leq y_1^*, Y_2^* \leq y_2^*\right)$. Then, the joint cdf has the copula representation is as follows:

$$H\left(y_1^*, y_2^*\right) = C\left[F_1\left(y_1^*\right), F_2\left(y_2^*\right); \theta_{12}\right] \tag{13}$$

The dependence between y_i^* and y_j^* is described as follows:

$$C\left[F_i\left(y_i^*\right), F_j\left(y_j^*\right); \theta_{ij}\right] = C\left[F_i\left(y_i^*|x;\beta_i\right), F_j\left(y_j^*|x;\beta_j\right); \theta_{ij}\right] \tag{14}$$

Then the copula density function is:

$$\begin{aligned} c_{ij}\left[F_i\left(y_i^*\right), F_j\left(y_j^*\right); \theta_{ij}\right] &= \frac{\partial^2}{\partial y_i^* \partial y_j^*} C\left[F_i\left(y_i^*|x;\beta_i\right), F_j\left(y_j^*|x;\beta_j\right); \theta_{ij}\right] \\ &= C\left[F_i\left(y_i^*\right), F_j\left(y_j^*\right); \theta_{ij}\right] f_i\left(y_i^*\right) f_j\left(y_j^*\right) \end{aligned} \tag{15}$$

Where$f_i\left(y_i^*\right) = \partial F_i\left(y_i^*\right)/\partial y_i^*$, and $f_j\left(y_j^*\right) = \partial F_j\left(y_j^*\right)/\partial y_j^*$.

Then we denote $F_i\left(y_i\right) = F\left[\left(y_i - x'\beta_i\right)/\sigma_i\right]$,$F_i\left(0\right) = F_i\left[\left(0 - x'\beta_i\right)/\sigma_i\right]$, $F_j\left(y_j\right) = F_j\left[\left(y_j - x'\beta_j\right)/\sigma_j\right]$ and $F_j\left(0\right) = F_j\left[\left(0 - x'\beta_j\right)/\sigma_j\right]$ and define dichotomous indicator 1 which takes the value 1 if event holds and 0 otherwise. Then the likelihood function for the bivariate Tobit for $y_i^*$and y_j^* is:

$$\begin{aligned} L_{ij} = &\left\{C\left[F_i(0), F_j(0); \theta_{ij}\right]\right\}^{1(y_i=0)1\left(y_j=0\right)} \\ &\times\left\{\mathrm{C}\left[\mathrm{F_i}\left(y_i\right), \mathrm{F_j}\left(y_j\right); \theta_\mathrm{ij}\right] f_i\left(y_i\right) f_j\left(y_j\right)\right\}^{1(y_i>0)1\left(y_j>0\right)} \\ &\times\left\{\mathrm{C}\left[\mathrm{F_i}\left(y_i\right), \mathrm{F_j}(0); \theta_\mathrm{ij}\right] f_i\left(y_i\right)\right\}^{1(y_i>0)1\left(y_j=0\right)} \\ &\times\left\{\mathrm{C}\left[\mathrm{F_i}(0), \mathrm{F_j}\left(y_j\right); \theta_\mathrm{ij}\right] f_j\left(y_j\right)\right\}^{1(y_i=0)1\left(y_j>0\right)} \end{aligned} \tag{16}$$

The specification is complete with the choice of copula functions. In this study, we consider three forms of Archimedean copula which are Gumbel copula, Clayton copula, and Frank copula, with two different univariate distribution assumptions (normal and Log-Burr distribution) for the random error term. Thus a total of six copula-based models were estimated: (1) Clayton-Normal, (2) Gumbel-Normal, (3) Frank-Normal, (4) Clayton- LogBurr, (5) Gumbel-LogBurr, and (6) Frank-LogBurr.

3 Data

The data we used is from China Health and Nutrition Survey (CHNS). This survey collected detailed information on individual and household nutritional status, health status, lifestyle habits including dietary and physical activities, socio-demographic, and physical examination results. The main objective of the survey is to see how the social and economic transformation of Chinese society in recent 20 years affect the health and nutritional status of population.

CHNS is a longitudinal survey which includes nine provinces: Henan, Jiangsu, Hubei, Liaoning, Shandong, Guizhou, Hunan and Guangxi Zhuang Autonomous Region. All provinces vary substantially in geography, economic development, public resources, and health indicators.

A multi-stage, random cluster process was used to draw the samples surveyed in each of the province. Counties in these nine provinces were stratified by income (low, middle, and high), and a weighted sampling scheme was used to randomly select four counties in each province. Villages and townships within the counties and urban/suburban neighborhoods within the cities were selected randomly. In 2009, there were 216 primary sampling units: 36 urban neighbourhoods, 36 suburban neighborhoods, 36 towns, and 108 villages.

In this study, we only use the data drawn from CHNS 2009. This dataset covers 6490 households and 94812 individuals. In our sample, we restricted the people whose age are from 30 to 65 years old, leaving a final sample of 5469 respondents from 9 provinces, of these 671 from Liaoning province, 604 from Heilongjiang province, 695 from Jiangsu province, 605 from Shandong province, 537 from Henan province, 614 from Hubei province, 522 from Hunan province, 642 from Guangxi province, and 579 Guizhou.

The dependent variables are BMI value, diastolic blood pressure and systolic blood pressure. Since in this study, we only consider those whose BMI value and blood value greater than normal standard. Then if BMI and blood pressure are over than normal range, they will be denoted as their true value, otherwise will be denoted as zero.

BMI was calculated from the respondents weight (in kilograms) divided by their height in square meters. There are two standards for BMI cutoff points, one is proposed by WHO which defines obesity if BMI is greater than 30, and another is recommended by Chinese center of disease control and prevention which defines obesity if BMI is over 28. In this paper, we will follow Chinese standard.

Blood pressure is the force of the blood pushing against the walls of the arteries. The normal range of systolic and diastolic blood pressure are less than 140mmHg and 90mmHg respectively. When either systolic value is greater than 140 mmHg or

Table 1 Diagnostic standard for High Blood Pressure

Category	Systolic		Diastolic
Normal	Less than 120	and	Less than 80
Prehypertention	120-139	or	80-89
	High Blood Pressure		
Stage 1	140-159	or	90-99
Stage 2	160 or higher	or	100 or higher

diastolic value is more than 90 mmHg, or both value beyond the critical value, the high blood pressure can be diagnosed (see Table.1).

The explanatory variables include age, number of cigarettes per day, quantity of liquors per week, gender, and dummy variables indicating education level, individual income level and nine provincial dummies. The definition and descriptive statistics of all variables are shown in Table 2.

4 Estimation Results

For the consideration of space, we only provide the data fit results for the best copula model. Among all coupla models, our results indicate that the Frank copula model provides the best data fit. Then the results of Frank copula with two marginal distributions (Normal margins and generalized log-Burr margins) are presented in following Table.3. For the rest of paper we focus on comparing the results from these two models.

Table 3 presented estimation results for Frank copula models with both normal margins and log-Burr margins. For Frank-Burr model, it can be seen that the estimates for all three skewness parameters are significant at the 1% level. For diastolic blood pressure, the skewness parameter estimates are 3.29 which makes the estimation results produced with the normal margins are doubtful.

In addition, we find there are some notable differences in the parameter estimates between Frank-Normal and Frank-Burr models. For example, Frank-Normal model shows that number of cigarettes per day, amount of liquor per week, education level, and residence are statistical insignificant. However, all of these variables are significant in Frank-Burr model. These differences demonstrate how a misspecified error distribution can obscure the effects of explanatory variables on the dependent variables and highlight the importance of accommodating skewed error distributions in censored equation systems. For the dependence structure, in both models, two of the error correlation (diastolic-systolic and BMI-diastolic) estimates are positive and significant at the 1% level, and one (BMI and systolic) is significant at 5% level.

Focusing on the results of the preferred Frank-Burr model, number of cigarettes per day has positive effects on both diastolic and systolic blood pressure and slight negative effects on body mass index, this means that smoking may lead to the increase of blood pressure. It is also can be seen that more amount of liquor drinks per

Table 2 Variables Definition and Sample Statistics

Variable	Definition	Mean
Dependent Variables		
BMI	Body Mass Index	23.64(3.37)
DBP	Diastolic blood pressure	83.46(8.14)
SBP	Systolic blood pressure	128.18(14.58)
Continuous independent variables		
Age	Calculated age in years to 2 decimal points	56.13(10.63)
Cigarettesperday	Number of cigarettes smokes per day	17.15(10.1)
Liquorperwk	Liquor: Number of Liang(50gms) drinks per week	9.11(12.68)
Binary independent variables (1=Yes; 0=no)		
Gender		
Male	Male	0.51
Education Level		
<primary	Education years are less than 9 years	0.46
Middle-school	Graduated from Middle School	0.32
High-school	Graduated from high School	0.19
>college	Education years are more than 15 years	0.05
Household Income Level		
Level-1	Annual individual income: $<$ 14000 RMB	0.22
Level-2	Annual individual income: 14000 28000 RMB	0.25
Level-3	Annual individual income: 28000 53000 RMB	0.39
Level-4	Annual individual income: $>$53000 RMB	0.14
Province		
Liaoning	Resides in Liaoning	0.12
Heilongjiang	Resides in Heilongjiang	0.11
Jiangsu	Resides in Jiangsu	0.13
Shandong	Resides in Shandong	0.11
Henan	Resides in Henan	0.10
Hubei	Resides in Hubei	0.11
Hunan	Resides in Hunan	0.09
Guangxi	Resides in Guangxi	0.12
Guizhou	Resides in Guizhou	0.11
Residence		
Urban	Resides in central city	0.39

Note. standard deviations are in parentheses

week, the higher value of diastolic, systolic blood pressure and BMI. We also find that household income has some impacts on blood pressure, the people with household income between 28000–53000RMB has higher risk in suffering high blood pressure. People residing in urban areas are more likely to suffer from high blood pressure and obesity. Provincial factors are also affect blood pressure. Compared with other provinces, Liaoning and Jiangsu has higher prevalence of high blood pressure.

Table 3 Estimation Results: Frank copula with two margins

	Frank (Normal margins)			Frank (Generalized Log-Burr)		
	Diastolic	Systolic	BMI	Diastolic	Systolic	BMI
Constant	17.28***	24.38***	3.06***	15.17***	21.16***	2.92***
	(13.49)	(13.18)	(8.721)	(2.51)	(12.48)	(7.74)
Age	12.91	29.37	0.52	12.69	29.37	0.51
	(1.92)	(1.12)	(0.32)	(1.98)	(1.76)	(0.31)
Age2	-1.151	-1.849	-0.0764	-1.152	-1.872	-0.0762
	(-0.22)	(-1.42)	(-0.54)	(-0.11)	(-1.30)	(-0.51)
Cigarettesperday	0.172	0.252	-0.0141	0.212**	0.263**	-0.012
	(-1.67)	(-1.68)	(-0.01)	(-3.64)	(-3.62)	(-0.01)
Liquorperwk	0.104*	0.156	0.00347	0.189**	0.356**	0.00647
	(2.23)	(1.63)	(-0.36)	(6.71)	(2.38)	(-0.49)
Household income						
14000-28000RMB	-1.235***	-2.64***	0.177	-1.975***	2.57***	0.177
	(-3.44)	(-3.43)	(0.42)	(-3.42)	(-3.51)	(0.43)
28000-53000RMB	1.037*	1.94**	0.517	1.017*	2.04***	1.17*
	(2.17)	(2.60)	(0.12)	(2.36)	(3.65)	(1.91)
>53000RMB	-2.10*	-2.968*	0.52	-3.10***	-3.968**	0.42**
	(-2.28)	(-2.13)	(1.13)	(-4.03)	(-2.18)	(2.12)
Education Level						
middle school	1.199	-6.793	0.742	1.099	-5.793	0.742
	(0.41)	(-1.53)	(1.81)	(0.39)	(-1.23)	(1.79)
High school	0.899	-3.793	0.942	2.109**	-2.793	0.632
	(0.41)	(-1.53)	(1.81)	(2.39)	(-1.23)	(1.79)
>College	1.276	2.713	0.792	2.099***	3.793**	0.912**
	(0.41)	(1.53)	(1.81)	(3.39)	(2.23)	(2.39)
Province						
Heilongjiang	-5.79	-5.312	-0.108	-6.71**	-7.31**	-0.168*
	(1.12)	(1.47)	(-0.19)	(2.33)	(2.45)	(-1.91)
Jiangsu	2.655**	5.838**	-0.169	4.655	7.512***	-0.214
	(2.07)	(2.37)	(-0.3)	(-0.45)	(5.91)	1.31
Shandong	-4.391	-7.979	1.732**	-5.161**	-8.172**	1.032**
	(-1.29)	(-1.58)	2.09	(-2.22)	(-2.26)	2.32
Henan	-2.79	-1.07**	1.147	-3.79	-1.87**	1.217**
	(-1.30)	(-2.07)	1.28	(-1.24)	(-2.88)	2.26
Hubei	-3.415*	-2.294	-0.665	-3.615	-2.04*	-0.635
	(-2.02)	(-1.21)	(-1.21)	(-1.96)	(-1.91)	(-1.13)
Hunan	-4.929	-9.76*	-0.784	-5.227	-10.76***	-0.924*
	(-1.37)	(-2.10)	(-1.43)	(-1.25)	(-6.97)	(-1.91)
Guangxi	-7.640*	-12.04*	-0.752	-7.64	-12.04*	-0.552
	(-2.06)	(-2.07)	(-1.25)	(-1.92)	(-2.09)	(-1.17)
Guizhou	-7.991*	-11.06*	-1.215*	-8.291**	-11.76***	-1.432*
	(-2.20)	(-2.38)	(-2.16)	(-2.31)	(-4.35)	(-2.27)
urban	2.071	5.096	0.38*	4.003***	6.096**	0.678**
	1.86	1.74	(1.21)	(4.17)	(2.51)	(2.21)
σ_i	4.72***	31.28***	12.06***	2.67***	29.12***	2.97***
	(11.23)	(28.17)	(15.27)	(12.76)	(5.92)	(11.96)
κ_i^*				3.29***	0.419***	3.13***
				(9.45)	(3.12)	(9.17)
Concordance(θ_{ij})						
Systolic BP	4.13***			3.97***		
	(6.17)			(6.07)		
BMI	2.36***	2.63**		2.06***	2.21*	
	(6.92)	(2.01)		(6.12)	(1.95)	
Kendall's tau						
Systolic BP	0.41***			0.46***		
	(8.21)			(7.21)		
BMI	0.24***	0.26**		0.28***	0.24**	
	(18.49)	(2.27)		(14.91)	(2.06)	
Log Likelihood	-1176.39				-997.29	

5 Concluding Remarks

Such quick increase in the prevalence of obesity and high blood pressure has gradually drawn much attentions from many aspects in recent years. If the increase in BMI and blood pressure could be diminished, much of corresponding chronic non-communicable disease, such as diabetes, cardiovascular, kidney disease, and stroke could be reduced.

In this study, since for the dependent variables HBP and obesity contain zeros in some items, this data feature must be accommodated to obtain consistent empirical estimates. Many modelling approaches are based on the assumption of normality, if violated, the estimates results may inconsistent. To conquer this problem, we use censored equation system and draw on recent developments in copula methods. The copula approach allows the use of more flexible error distributions than normal distribution, and can also overcome the computational difficulty with multiple probability integrals in larger systems.

In this paper, we considered three forms of Archimedean copula (Gumbel copula, Clayton coupla, and Frank copula) with two different univariate distribution assumptions (normal and Log-Burr distribution) for the random error term. Thus a total of six copula-based models were estimated: (1) Clayton-Normal, (2) Gumbel-Normal, (3) Frank-Normal, (4) Clayton- LogBurr, (5) Gumbel-LogBurr, and (6) Frank-LogBurr. Through comparing log-likeligood and BIC, we found, among all copula models, Frank copula with Log-Burr marginal distribution model provides the best data fit.

This study provides a detailed econometric analysis of the role of socio-economic factors in the prevalence of obesity and HBP in China. In addition, this analysis also provide more messages to policy maker in making effective policies and take useful intervention measures for reducing the prevalence of HBP and obesity.

This study represents one of the attempts at economically determine the socio-demographic and social-economic factors affecting body mass index and blood pressure among Chinese people. With several rounds data of CHNS, future studies will be conducted by applying longitudinal approaches.

Acknowledgement. This research uses data from China Health and Nutrition Survey (CHNS). We thank the National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700) and the Fogarty International Center, NIH for financial support for the CHNS data collection and analysis files from 1989 to 2009 and both parties plus the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009 and future surveys.

References

1. Amemiya, T.: Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. Econometrica 42(6), 999–1012 (1974)
2. Clayton, D.G.: A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. Biometrika 65, 141–152 (1978)
3. Du, S., Mroz, T.A., Zhai, F., Popkin, B.M.: Rapid income growth adversely affects diet quality in China? particularly for the poor. Social Science Medicine 59, 1505–1515 (2004)
4. Frank, M.J.: On the Simultaneous Associativity of F(x, y) and x + y - F(x, y). Aequationes Mathematicae 19, 194–226 (1979)
5. Turnbull, F., Woodward, M., Anna, V.: Effectiveness of blood pressure lowering: evidence-based comparisons between men and women. Expert Review of Cardiovascular Therapy 8(2), 199–209 (2010)
6. Gumbel, E.J.: Bivariate Exponential Distributions. Journal of the American Statistical Association 55, 698–707 (1960)
7. Hajjar, I., Kotchen, T.: Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988-2000. JAMA 290, 199–206 (2003)
8. He, J., Gu, D., Chen, J., et al.: Premature deaths attributable to blood pressure in China: a prospective cohort study. Lancet. 374, 1765–1772 (2009)
9. Joe, H.: Multivariate Dependence and Dependence Models. Chapman & Hall, New York (1997)
10. Joe, H.: Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis 94(2), 401–419 (2005)
11. National Bureau of Statistics of China (NBSC), Statistical data. National Bureau of Statistics of China, Beijing (2009, 2010), `http://www.stats.gov.cn/english/`
12. Nelsen, R.B.: An Introduction to Copulas. Springer (1998)
13. Oruc Ege, O., Ucer, B.: A new method for local dependence map and its applications. Turkiye Klinikleri J. Biosta 1, 1–8 (2009)
14. Griffiths, S.M.: Leading a Healthy Lifestyle:The Challenges for China. Asian-Pacific Journal of Public Health 22(3), 110–116 (2010)
15. Schweizer, B., Sklar, A.: Operations on distribution functions not derivable from operations on random variables. Studia Mathematica 52, 43–52 (1974)
16. Sener, I.N., Eluru, N., Bhat, C.R.: On jointly analyzing the physical activity participation levels of individuals in a family unit using a multivariate copula framework. Journal of Choice Modelling 3(3), 1–38 (2010)
17. Wang, Y., Mi, J., Shan, X.Y., Wang, Q.J., Ge, K.Y.: Is China facing an obesity epidemic and the consequences? The trends in obesity and chronic disease in China. International Journal of Obesity 31(1), 177–188 (2006)
18. Wildman, R.P., Gu, D., Munter, P., Wu, X., Reynolds, K., Duan, X., et al.: Trends in overweight and obesity in Chinese adults: between 1991 and 2000. Obesity 16(6), 1448–1453 (2008)
19. World Health Report, Reducing risks, promoting healthy life. World Health Organization, Geneva (2002), `http://www.who.int/whr/2002/`
20. Yen, S.T., Yuan, Y., Liu, X.: Alcohol consumption by men in China: A non-Gaussian censored system approach. Journal of China Economic Review 20, 162–173 (2009)

Author Index