International Series in Operations Research & Management Science

Alain Bensoussan
Dominique Guegan
Charles S. Tapiero *Editors*

Future Perspectives in Risk Models and Finance

Springer

International Series in Operations Research & Management Science

Volume 211

The book series *International Series in Operations Research and Management Science* encompasses the various areas of operations research and management science. Both theoretical and applied books are included. It describes current advances anywhere in the world that are at the cutting edge of the field. The series is aimed especially at researchers, doctoral students, and sophisticated practitioners.
The series features three types of books:

- *Advanced expository books* that extend and unify our understanding of particular areas.
- *Research monographs* that make substantial contributions to knowledge.
- *Handbooks* that define the new state of the art in particular areas. They will be entitled *Recent Advances in (name of the area)*. Each handbook will be edited by a leading authority in the area who will organize a team of experts on various aspects of the topic to write individual chapters. A handbook may emphasize expository surveys or completely new advances (either research or applications) or a combination of both.

The series emphasizes the following four areas:
Mathematical Programming: Including linear programming, integer programming, nonlinear programming, interior point methods, game theory, network optimization models, combinatorics, equilibrium programming, complementarity theory, multiobjective optimization, dynamic programming, stochastic programming, complexity theory, etc.
Applied Probability: Including queuing theory, simulation, renewal theory, Brownian motion and diffusion processes, decision analysis, Markov decision processes, reliability theory, forecasting, other stochastic processes motivated by applications, etc.
Production and Operations Management: Including inventory theory, production scheduling, capacity planning, facility location, supply chain management, distribution systems, materials requirements planning, just-in-time systems, flexible manufacturing systems, design of production lines, logistical planning, strategic issues, etc.
Applications of Operations Research and Management Science: Including telecommunications, health care, capital budgeting and finance, marketing, public policy, military operations research, service operations, transportation systems, etc.

More information about this series at http://www.springer.com/series/6161

Alain Bensoussan • Dominique Guegan
Charles S. Tapiero
Editors

Future Perspectives in Risk Models and Finance

Springer

Editors
Alain Bensoussan
Naveen Jindal School of Management
University of Texas at Dallas
Richardson
Texas
USA

Charles S. Tapiero
Department of Finance & Risk Engineering
Polytechnic Institute of New York University
Brooklyn
New York
USA

Dominique Guegan
Economic Doctorate School EPS
University Paris 1 Panthéon-Sorbonne
Paris
France

ISSN 0884-8289 ISSN 2214-7934 (electronic)
ISBN 978-3-319-07523-5 ISBN 978-3-319-07524-2 (eBook)
DOI 10.1007/978-3-319-07524-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014944809

Printed on acid-free paper

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

Editors Introduction

Risk models are models of uncertainty, and therefore all risk models are an expression of perceptions, priorities, needs and the information we have. In this sense, all risk models are complex hypotheses we have constructed and are based on "what we have or believe". Risk models are then challenged by their definition. Are risk definitions defining in fact prospective risks? How are risks estimated, what data can we apply to estimate their parameters and how can we do so and use them to useful and constructive ends? The purpose of this book is to provide a perspective on a number of financial analytics models and estimation techniques. The papers assembled for this special issue arise from inter-University collaborations between the New York University School of Engineering, the Chinese University of Hong Kong and the University of Paris I-Pantheon-Sorbonne. The first part of this book includes an outline by Alain Bensoussan et al. of GLM estimation techniques combined with the proof of fundamental results. Applications to static and dynamic models provide a unified approach to the estimation of nonlinear risk models.

A second paper is concerned with the definition of risks and their management. In particular, Guegan and Hassani review a number of risk models, emphasizing the importance of bi-modal distributions for financial regulation. An additional paper (Stress Testing Engineering: the real risk measurement?) provides a review of stress testing and their implications. "The Skin In The Game as a Risk Filter" by Taleb, and Sandis provide an anti-fragility approach based on "skin in the game". To conclude these papers Raphael Douady and Taleb (Capital Adequacy, Pro-cyclicality and Systemic Risk) provide a new Fragility Theorem providing a quantitative foundation to Fragility and Antifragility.

The third part of this book is concerned with financial modelling and in particular a variety of financial systems seeking to model when markets are incomplete. Tapiero and Vallois (Financial Modelling and Memory: Mathematical Systems) provide an overview of mathematical systems and their uses in financial modelling. The intent of this chapter is to identify a number of mathematical approaches that include as special cases underlying financial models for complete markets. In particular, exponential memory models as well as the lognormal financial pricing model. The intent of the paper is to motivate a pricing measurement of financial assets when markets are incomplete relative to models that define complete markets. Memory based financial models, spanning non-memory models, long run-fractional models and short (persistent) memory models are considered. Applications of these models are used to highlight their variety and their importance to Financial Analytics. Subsequently Bianchi and Pianese (Asset price modeling: From Fractional to Multifractional Processes) provide an extensive overview of multi-fractional models and their important applications to Asset price modeling: from Fractional to Multi-fractional Processes. The latter models assume that the fractional parameter in a stochastic system is non-stationary. Finally, Tapiero and Jiangyi Qi (Financial Analytics and A Binomial Pricing Model) consider the binomial pricing model by discussing the effects of memory on the pricing of asset prices.

The papers in this book are concerned with both theoretical and practical issues. Theoretically, financial risk models are rationally bounded models, based on information and rules that are both available and agreed to by their user. Empirical and data finance however, has provided a bridge between theoretical constructs of risk models and the empirical evidence that these models entail. Numerous approaches are then used to model financial risk models, emphasizing mathematical and stochastic models based on the fundamental theoretical tenets of finance and others departing from the fundamental assumptions of finance. The underlying mathematical foundations of these risks models provide a future guideline for risk modeling. Both static and dynamic risk models are then considered. The papers in this book provide selective insights and developments that can contribute to a greater understanding of the complexity of financial modelling and its ability to bridge financial theories and their practice. An extended outline of the papers include is given below.

Generalized Linear Models (GLM) have been introduced by J. A. Nelder and R. W. M. Wedderburn, They describe random observations depending on unobservable variables of interest, generalizing the standard Gaussian error model. Many estimation results can be obtained in this context, which generalize with some approximation procedures from the Normal-Gaussian case. These results are important in risk models in general and in finance in particular. Bensoussan, Bertrand and Brouste revisit these results, providing proofs and extending the results of GLM. In particular, they prove the Central Limit theorem for the MLE, maximum likelihood estimator, in a general setting. They also provide a recursive estimator, similar to the Kalman filter, thus providing a statistical approach to the estimation of dynamic problems (many of which recur in financial modeling). Examples are used profusely.

Dominique Guégan and Bertrand K. Hassani provide two papers. The first paper, on New Distorsion Risk measure based on Bimodal Distributions, reviews and compares a number of approaches to risk management and the relevance to VaR (Value at Risk), ES (Expected shortfall) quantile measures, spectral risk measure and distortion risk measure. Knowing that quantile based risk measure cannot capture correctly the risk aversion of risk manager and spectral risk measure can be inconsistent to risk aversion, they propose a new distortion risk measure extending the work of Wang (2000) and Seresda et al. (2012). In particular they demonstrate how we may construct bi-modal risk distributions which are deemed appropriate for the kind of risks finance is confronted by. Finally they provide a comprehensive analysis of the feasibility of this approach using S&P500 data from 1999 to 2011.

Guégan and Hassan's second paper, on Stress Testing Engineering: Risk Vs Incident, first discusses the importance, and banks' and regulators' need and use, of stress testing. Such procedures have assumed a particular importance following the 2008 financial crisis and regulators to limit systemic risks. Stress testing is used to determine the stability or the resilience of a given financial institution by deliberately submitting it to intense scenario of adverse conditions, which is not considered a priori. These scenarios involve testing beyond the traditional capabilities – usually to determine limits – to confirm and comply to specifications that are both accurate and provide a greater understanding of the financial processes that underlie potential and future failures. Stress testing is therefore quintessential to financial risk management. This paper's focus is on two families of triggers: a first assesses the impact of external (and/or extreme) events, the second considers the choice of models and their fault in their predictions. Specifically, models increasing inadequacy over time due to their inflexibility to adapt to dynamical changes. The first trigger accounts for macro-economic data measurements or massive operational risks while the second focuses on the limits of quantitative models for forecasting, pricing, evaluating capital or managing risks. Of course, if banks' internal controls were to identify their limitations, pitfalls and other models' drawbacks, they could be prevented better.

Nassim N. Taleb and Constantine Sandis's paper, The *Skin In The Game* Heuristic for Protection Against Tail Events, addresses the standard economic theory. This theory makes an allowance for the agency problem, but not the compounding of moral hazard in the presence of informational opacity, particularly in what concerns high-impact events in fat tailed domains (under slow convergence for the law of large numbers). Skin in the game In the language of probability, skin in the game creates an absorbing state for an agent and not just its principal. In their paper, they propose a global and morally mandatory heuristic that anyone involved in an action which can possibly generate harm for others, should be required to be exposed to some damage, regardless of context. It is supposed to counter voluntary and involuntary risk hiding—and risk transfer—in the tails. Finally, they link the rule to various philosophical approaches to ethics and moral luck.

Raphael Douady's and Taleb paper on "A Fragility Theorem" integrates model error (and biases) into a fragile or antifragile context. Unlike risk, which is linked to psychological notions such as subjective preferences, Douady and Taleb offer a measure that is universal and concerns any object that has a probability distribution (whether such distribution is known or, critically, unknown). The notions of fragility and antifragility were introduced in Taleb (2012). In short, fragility is related to how a system *suffers* from the variability of its environment beyond a certain preset threshold (when threshold is K, it is called K-fragility), while antifragility refers to when it *benefits* from this variability—in a similar way to "vega" of an option or a nonlinear payoff, that is, its sensitivity to volatility or some similar measure of scale of a distribution. To these purposes, they use measures in L^2 such as standard deviations, which restrict the choice of probability distributions. The broader measure of absolute deviation, cut into two parts: lower and upper semi-deviation above the distribution center Ω. The paper's contributions provide a mathematical definition of fragility and antifragility as negative or positive sensitivity to a semi-measure of dispersion and volatility (a variant of negative or positive "vega") and examine the link to nonlinear effects. They also construct a measure of "vega" in the tails of the distribution that depends on the variations of s, the semi-deviation below a certain level Ω, chosen in the L^1 norm in order to insure its existence under "fat tailed" distributions with finite first semi-moment. In fact s would exist as a measure even in the case of infinite moments to the right side of Ω. Finally, the paper proposes a detection of fragility using a single "fast-and-frugal", model-free, probability free heuristic that also picks up exposure to model error. The heuristic lends itself to immediate implementation, and uncovers hidden risks related to company size, forecasting problems, and bank tail exposures (it explains the forecasting biases). While simple to implement, it improves on stress testing and bypasses the common flaws in Value-at-Risk.

Charles S. Tapiero and Pierre Vallois' paper consider a number of approaches to memory-based stochastic models. Essentially, stochastic mean reverting models, long run and short memory models are introduced and their underlying assumptions specifically outlined. In particular, a family of models based on the exponential memory Ornstein Uhlenbeck process is presented and generalized to a broad set of memory models including extreme distribution models (such as the Weibull) as well as long run memory models. The intent of these models is to assess the constructs and the implications of memory within stochastic models based on previous disturbances. Long run memory models are also presented to emphasize their differences from the Brownian motion based models. The essential contribution of the paper however is a review and an application of short run (persistent) stochastic models. These models, although not commonly used in financial modeling have a potential contribution, providing "a recurrent bifurcation" of states evolutions as a function of the information observed (e.g., a shift in the Federal Reserve policy or its support to financial markets). It may alter traders' and investors' expectations and future probabilities of market prices. Similarly, in insurance, events such as an "accident" may alter the probabilities that insured will claim, etc. Although, discrete random walk and counting processes are considered, a summary of recent results based on Hermann and Vallois' paper are given.

The subsequent paper by Tapiero and Jiangyi Qi, consider a binomial memory model addressing questions such as "do stock prices have memory"? How does it affect financial models and our decisions? The binomial memory-less price model is extended to memory prone models including Short run and Bayesian Learning models. The implication of such models to financial pricing is also outlined.

Sergio Bianchi and Augusto Pianese address the financial modeling of Asset prices, spanning approaches from Fractional to Multifractional Processes. Motivation for these models arose following the 2007–2009 crisis increasing the awareness that the standard financial models used were not describing real world data. The use of Brownian motion as paradigm and its variants are also difficult to justify in light a tumultuous behavior of financial markets. Multifractional processes are presented as a more general approach that may account for the complexity of global financial markets with nonlinear autocorrelations in their log-price variations and their slow decay in absolute (or squared) returns; the asymmetric behavior of stock prices that produces large and sudden drawdowns, but only slow upward movements; volatility clustering; heavy tails in the unconditional distributions of returns and the presence of conditional heavy tails even for residuals obtained by correcting returns for volatility clustering; "Gaussianity", as one increases the time scale used to calculate the returns; correlation between volatility and traded volumes; an asymmetry in time scales, i.e. that fine-scale measures of volatility predict coarse-grained volatility worse than the way round, etc. The considerations above suggest that a broad set of new stochastic models may be considered that may be able to better explain theoretically and practically the complexity of observed financial markets behaviors. Long run memory models and Multi-fractional processes are presented as providing an avenue for current and future research in stochastic financial modeling.

As a whole, these papers provide a future and analytical perspective to both risk management, estimation and future financial modelling embedded in traditional financial models and yet expanded to account to the effects of markets incompleteness.

April 1, 2014

Contents

Contributors

Alain Bensoussan International Center for Decision and Risk Analysis, School of Management, University of Texas at Dallas, Dallas, Texas, USA

Pierre Bertrand Laboratoire de Mathématiques, UMR CNRS 6620, Université Clermont-Ferrand 2, Clermont-Ferrand, France

Sergio Bianchi Department of Economics and Law, University of Cassino, Cassino, Italy

Alexandre Brouste Laboratoire Manceau de Mathématiques, Université du Maine, Le Mans, France

Raphael Douady Labex ReFi, University Paris 1, Paris Cedex 13, France

Dominique Guégan University Paris 1 Panthéon-Sorbonne et New York University Polytechnic School of Engineering, Brooklyn, New York, USA

CES UMR 8174, 106 boulevard de l'Hopital, 75647 Paris Cedex 13, France

Bertrand Hassani University Paris 1 Panthéon-Sorbonne, Paris Cedex 13, France

Bertrand K. Hassani Université Paris 1 Panthéon-Sorbonne CES UMR 8174 and Grupo Santander, Paris Cedex 13, France

Augusto Pianese Department of Finance and Risk Engineering, New York University, New York, USA

Jiangyi Qi Department of Finance and Risk Engineering, New York University Polytechnic School of Engineering, Brooklyn, New York, USA

Constantine Sandis School of Philosophy, Oxford Brooks University, Oxford, U.K.

Nassim N. Taleb New York University School of Engineering, Department of Finance and Risk Engineering, Brooklyn, New York, USA

Charles S. Tapiero Department of Finance and Risk Engineering, New York University Polytechnic School of Engineering, Brooklyn, New York, USA

Pierre Vallois Institut de Mathematiques Elie Cartan—UMR7502, University of Lorraine, Vandoeuvre-les-Nancy Cedex, France

Estimation Theory for Generalized Linear Models

Alain Bensoussan, Pierre Bertrand and Alexandre Brouste

Also with the College of Science and Engineering, Systems Engineering and Engineering Management, City University Hong Kong. Research supported by the National Science Foundation under grant DMS-1303775 and the Research Grants Council of the Hong Kong Special Administrative Region (CityU 500113).

1 Introduction

The GLM are a generalization of the classical linear gaussian model

$$z = Hx + h + \epsilon, \; x \in R^n, \; z \in R^d \tag{1}$$

in which x are the variables of interest, H is a given matrix $\in \mathcal{L}(R^n; R^d)$ and ϵ is a gaussian variable in R^d with mean 0 variance Σ. The major generalization is in giving up the Gaussian assumption. However, this prevents us also to write the observation as in (1). We have to work with the conditional probability density $f(z|x)$. We shall in the introduction assume $d = 1$, to simplify the presentation.

We consider a real random variable, still denoted by z, whose range is not necessarily $(-\infty, +\infty)$ and assume that it has a probability density defined on the range, depending on a parameter θ, also a scalar. This parameter is called the ***canonical parameter.*** We recall that we use the same notation for the random variable z and the argument of the probability density, called also the likelihood. So we call $f(z, \theta)$ the likelihood of the random variable z. In the case z is a discrete probability distribution, we keep the notation to represent the probabilities of specific values of z. We consider

A. Bensoussan (✉)
International Center for Decision and Risk Analysis, School of Management,
University of Texas at Dallas, Dallas, Texas, USA
e-mail: axb046100@utdallas.edu

P. Bertrand
Laboratoire de Mathématiques, UMR CNRS 6620,
Université Clermont-Ferrand 2, Clermont-Ferrand, France

A. Brouste
Laboratoire Manceau de Mathématiques,
Université du Maine, Le Mans, France

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,*
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_1,

the loglikelihood $L(z,\theta) = \log\ f(z,\theta)$, defined on the range of values of the random variable z.

The basic assumption is that this function has the form

$$L(z,\theta) = z\theta - b(\theta) + c(z) \tag{2}$$

Because the function $L(z,\theta)$ is a loglikelihood, we shall see that the function $b(\theta)$ cannot be arbitrary. If it is smooth, we shall show that it is necessarily convex. First of all, we check the relation

$$\mu = b'(\theta) \tag{3}$$

in which μ is the mean of z. Indeed

$$E\,\frac{\partial L(z,\theta)}{\partial\theta} = 0$$

from which, (3) follows immediately. Assume that the function $b'(\theta)$ is invertible, on the range of values of z. Therefore, the canonical parameter can be expressed in terms of the mean. Next, from the relation

$$\frac{\partial^2 L(z,\theta)}{\partial\theta^2} + \left(\frac{\partial L(z,\theta)}{\partial\theta}\right)^2 = \frac{1}{f(z,\theta)}\frac{\partial^2 f(z,\theta)}{\partial\theta^2} \tag{4}$$

we have

$$E\left[\frac{\partial^2 L(z,\theta)}{\partial\theta^2} + \left(\frac{\partial L(z,\theta)}{\partial\theta}\right)^2\right] = 0$$

and

$$V = E(z-\mu)^2 = b"(\theta) \tag{5}$$

which proves the convexity of $b(\theta)$.

Also

$$\frac{\partial^3 L}{\partial\theta^3} + \left(\frac{\partial L}{\partial\theta}\right)^3 + 3\frac{\partial L}{\partial\theta}\frac{\partial^2 L}{\partial\theta^2} = \frac{1}{f}\frac{\partial^3 f(z,\theta)}{\partial\theta^3}$$

from which we deduce

$$E\left[\frac{\partial^3 L}{\partial\theta^3} + \left(\frac{\partial L}{\partial\theta}\right)^3\right] = 0$$

hence

$$E(z-\mu)^3 = b'''(\theta) \tag{6}$$

The next ingredient of GLM models is the link function. It connects the canonical parameter to the variables of interest $x \in R^n$, and uses the mean as an intermediary. We express the link by the relation

$$h^* x = g(\mu) \tag{7}$$

where $h^* \in \mathcal{L}(R^n; R)$ is the equivalent of matrix H in formula (1), since $d = 1$, with the constant term taken as 0. The function g is a link function, defined on the range of values of z. If g is invertible on its domain, we can express the mean as a function of $h^* x$, and also the canonical parameter as a function of the variables of interest, by inverting the relation

$$h^* x = g(b'(\theta)) \tag{8}$$

These considerations form the core of the GLM models, up to some extensions. We will review the case of vector observations in Sect. 5.

2 Examples

2.1 *Gaussian Distribution*

Let

$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(z-\mu)^2}{2\sigma^2}$$

and

$$\log f(z) = -\frac{(z-\mu)^2}{2\sigma^2} - \log \sqrt{2\pi}\sigma$$

If we set $\theta = \dfrac{\mu}{\sigma^2}$ and $b(\theta) = \dfrac{1}{2}\sigma^2\theta^2$, considering σ as a fixed constant, we can write

$$\log f(z) = L(z, \theta)$$

with $L(z, \theta)$ defined by (2) in which

$$c(z) = -\frac{1}{2}\frac{z^2}{\sigma^2} - \log \sqrt{2\pi}\sigma$$

2.2 *Exponential Distribution*

Let

$$f(z) = \frac{1}{\mu} \exp - \frac{z}{\mu}$$

for $z \in R^+$. On the range we have

$$\log f(z) = -\log \mu - \frac{z}{\mu}$$

which can be written as $L(z, \theta)$ with

$$\theta = -\frac{1}{\mu} \quad b(\theta) = -\log -\theta \quad c(z) = 0 \tag{9}$$

The function $b(\theta)$ is defined on R^-.

2.3 Poisson Distribution

We have

$$f(z) = \exp -\mu \, \frac{\mu^z}{z!}$$

with z integer. Of course $f(z)$ is not a density. We keep this notation for convenience. Therefore

$$\begin{aligned} \log f(z) &= z \log \mu - \mu - \log z! \\ &= L(z, \theta) \end{aligned}$$

with $\theta = \log \mu$, $b(\theta) = \exp \theta$, $c(z) = -\log z!$.

2.4 Binomial Distribution

Let

$$f(z) = C_q^z \pi^z (1-\pi)^{q-z}$$

and z runs from 0 to q. So

$$\begin{aligned} \log f(z) &= z \, \log \frac{\pi}{1-\pi} + q \, \log(1-\pi) + \log C_q^z \\ &= L(z, \theta) \end{aligned}$$

with

$$\theta = \log \frac{\pi}{1-\pi}, \; b(\theta) = q \log(1 + \exp \theta), \; c(z) = \log C_q^z$$

We have

$$\mu = q\pi = q \frac{\exp \theta}{1 + \exp \theta}$$

2.5 Gamma Distribution

We have

$$f(z) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} z^{\alpha-1} \exp -\frac{z}{\beta}$$

The range is R^{+} and α, β are positive parameters. We have

$$\begin{aligned} \log f(z) &= -\frac{z}{\beta} - \alpha \log \beta + (\alpha - 1) \log z - \log \Gamma(\alpha) \\ &= L(z, \theta) \end{aligned}$$

with

$$\theta = -\frac{1}{\beta},\ b(\theta) = -\alpha \log -\theta,\ c(z) = (\alpha - 1) \log z - \log \Gamma(\alpha)$$

The function $b(\theta)$ is defined on R^{-}.

We consider α as a given number. Note that $\mu = \alpha\beta = -\frac{\alpha}{\theta}$. For $\alpha = 1$, we recover the exponential distribution.

2.6 Weibull Distribution

The Weibull distribution is defined by

$$\varphi(y) = \frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} \exp - \left(\frac{y}{\lambda}\right)^{k}$$

over the range $y > 0$. We have used different notation $\varphi(y)$ instead of $f(z)$ intentionally, because this distribution does not satisfy the assumptions of GLM.

If $k = 1$ it reduces to the exponential distribution with mean λ. The mean is given by

$$\mu = \lambda\Gamma\left(1 + \frac{1}{k}\right)$$

The parameter $k \geq 1$ is called the shape parameter and λ is called the scale parameter. Suppose λ is linked to the variables of interest x by the relation

$$\lambda = h^{*}x \tag{10}$$

Can we estimate x by observing the random variable y? The answer is yes, because we can associate to y a random variable, observable when y is observed, which belongs to the GLM family. This is done by defining

$$z = y^{k}$$

for which the density is given by

$$f(z) = \frac{1}{\lambda^k} \exp -\frac{z}{\lambda^k}$$

defined on R^+. Thus it is an exponential distribution, that belongs to the GLM family with

$$\theta = -\frac{1}{\lambda^k},\ b(\theta) = -\log -\theta$$

The function $b(\theta)$ is defined on R^-. If now we introduce the link function

$$g(\mu) = \mu^{\frac{1}{k}}$$

we can write

$$h^* x = g(b'(\theta)) = \left(-\frac{1}{\theta}\right)^{\frac{1}{k}}$$

and we are in the general framework described in the introduction.

2.7 Nonlinear Gaussian Model

Consider the following model

$$z = \varphi(h^* x) + \epsilon \tag{11}$$

in which ϵ is gaussian with mean 0 and variance σ^2. We assume φ invertible. It belongs to the GLM family with

$$b(\theta) = \frac{\sigma^2 \theta^2}{2},\ g(\mu) = \varphi^{-1}(\mu) \tag{12}$$

2.8 Canonical Links

A link function $g(.)$ is $\boldsymbol{canonical}$ if

$$g(b'(\theta)) = c\theta \tag{13}$$

where c is a constant. Therefore the following link functions are canonical for the GLM models indicated in parenthesis

$$g(\mu) = \mu, \quad \text{Gaussian}$$

$$g(\mu) = -\frac{1}{\mu}, \quad \text{Exponential}$$

$$g(\mu) = \log \mu, \quad \text{Poisson}$$

$$g(\mu) = \log \frac{\mu}{q-\mu}, \quad \text{Binomial}$$

$$g(\mu) = -\frac{\alpha}{\mu}, \quad \text{Gamma}$$

Note that in the first case the constant is σ^2. In the other cases, the constant is 1.

For the Weibull distribution, discussed in Sect. 2.6, the link function for the exponential variable $z = y^k$ is not canonical since

$$g(b'(\theta)) = \left(-\frac{1}{\theta}\right)^{\frac{1}{k}}$$

Similarly for the nonlinear Gaussian case the link function is not canonical. For canonical link functions, we have simply

$$c\theta = h^* x \tag{14}$$

and also

$$g'(b'(\theta))b"(\theta) = c \tag{15}$$

Remark By changing h into $\frac{h}{c}$ it is always possible to take $c = 1$, which is the more common definition of canonical link.

3 Maximum Likelihood Estimator (MLE)

3.1 Preliminaries

We begin with the general theory of MLE, which is often presented in a heuristic way (Sorenson 1980). We follow (Ibramovic and Has'minskii 1981) with some modifications. We shall consider a probability density $f(z, \theta)$, in which the parameter $\theta \in R^k$. We assume necessary smoothness, without stating it explicitly. Since the number of parameters will play a role, we do not take just $k = 1$. The random variable z is in R^d. We call θ_0 the true value, which we want to estimate. In fact the probability density of z is $f(z, \theta_0)$. We proceed with a sample of independent random variables

$z^1, \cdots, z^M$. The joint probability density of this sample, is of course

$$\prod_{j=1}^{M} f\left(z^j, \theta_0\right)$$

Assume, to simplify technicalities, that we have an open bounded subset of R^k, denoted Θ, such that $\theta_0 \in \Theta$. We set

$$Z_M(u) = \prod_{j=1}^{M} \frac{f\left(z^j, \theta_0 + u\right)}{f(z^j, \theta_0)} \tag{16}$$

A maximum likelihood estimator (MLE) $\hat{\theta}_M = \theta_0 + \hat{u}_M$ satisfies

$$Z_M\left(\hat{u}_M\right) = \sup_{\{u|\theta_0+u\in\Theta\}} Z_M(u) \tag{17}$$

Since Θ is open, we cannot guarantee the existence of a maximum. Thus, we will postulate its existence, and derive properties of a MLE.

In the following, we shall omit to write explicitly the constraint $\theta_0 + u \in \Theta$, unless useful.

The *consistency* of $\hat{\theta}_M$ is the property that $\hat{\theta}_M \to \theta_0$ as $M \to \infty$. We may have consistency a.s. or in probability. We shall use the observation

$$\begin{aligned}
\{|\hat{\theta}_M - \theta_0| > \gamma\} &= \{|\hat{u}_M| > \gamma\} \\
&= \left\{ \sup_{|u|>\gamma} Z_M(u)| \geq \sup_{|u|\leq\gamma} Z_M(u) \right\} \\
&\subset \left\{ \sup_{|u|>\gamma} Z_M(u)| \geq Z_M(0) \right\} \\
&= \left\{ \sup_{|u|>\gamma} Z_M(u)| \geq 1 \right\}
\end{aligned} \tag{18}$$

which will be very useful in obtaining estimates.

3.2 Consistency

Consider the quantities

$$\pi_\theta(\gamma) = \inf_{\substack{\theta' \in \Theta \\ |\theta' - \theta| \geq \gamma}} \int_{R^d} (f^{\frac{1}{2}}(z, \theta') - f^{\frac{1}{2}}(z, \theta)^2 dz \tag{19}$$

$$\varpi_\theta^2(\delta) = \int_{R^d} \sup_{|\theta'-\theta|\leq\delta} (f^{\frac{1}{2}}(z, \theta') - f^{\frac{1}{2}}(z, \theta)^2 dz \tag{20}$$

We have the following

Theorem 1 *Assume*

$$\forall \theta \in \Theta, \gamma > 0, \ \pi_\theta(\gamma) > 0 \tag{21}$$

$$\forall \theta \in \bar{\Theta} \ \lim_{\delta \to 0} \varpi_\theta(\delta) = 0 \tag{22}$$

then

$$\hat{\theta}_M \to \theta_0 \ as \ M \to \infty, \ a.s. \tag{23}$$

Remark 2 The first assumption simply means that, for two elements $\theta, \theta' \in \Theta$ such that $|\theta - \theta'| \geq \gamma$, then necessarily

$$f(z, \theta) \neq f(z, \theta'), \text{ on a set of positive measure.}$$

We begin with a Lemma

Lemma 3 *The property* (22) *implies the stronger property*

$$\lim_{\delta \to 0} \sup_{\theta \in \bar{\Theta}} \varpi_\theta(\delta) = 0 \tag{24}$$

Proof If the property is not true, there exists a sequence $\theta_n \in \bar{\Theta}$, $\delta_n \to 0$ such that

$$\varpi_{\theta_n}(\delta_n) \geq \beta > 0$$

We can assume that, for a subsequence

$$\theta_n \to \theta^*$$

but then, using the inequality

$$\varpi_{\theta_n}(\delta_n) \leq \varpi_{\theta^*}\left(|\theta^* - \theta_n|\right) + \varpi_{\theta^*}\left(|\theta^* - \theta_n| + \delta_n\right)$$

and from (22) we get necessarily $\varpi_{\theta_n}(\delta_n) \to 0$, which contradicts the assumption. ■ □

Proof of Theorem 1 Consider u_0 such that $|u_0| \geq \gamma$ and $\theta_0 + u_0 \in \bar{\Theta}$. Such points exist for γ sufficiently small, since Θ is open and $\theta_0 \in \Theta$. Let Γ_0 be the sphere of center $\theta_0 + u_0$ and radius δ. We estimate

$$E \sup_{\{u | \theta_0 + u \in \Gamma_0\}} Z_M^{\frac{1}{2}}(u) = E \sup_{\Gamma_0} Z_M^{\frac{1}{2}}(u)$$

and write

$$Z_M^{\frac{1}{2}}(u) = \prod_{j=1}^{M} f^{-\frac{1}{2}}\left(z^j, \theta_0\right) \prod_{j=1}^{M} f^{\frac{1}{2}}\left(z^j, \theta_0 + u\right)$$

$$\leq \prod_{j=1}^{M} f^{-\frac{1}{2}}(z^j, \theta_0) \prod_{j=1}^{M} \left(f^{\frac{1}{2}}(z^j, \theta_0 + u_0) + |f^{\frac{1}{2}}(z^j, \theta_0 + u) - f^{\frac{1}{2}}(z^j, \theta_0 + u_0)| \right)$$

hence

$$\sup_{\Gamma_0} Z_M^{\frac{1}{2}}(u) \leq \prod_{j=1}^{M} f^{-\frac{1}{2}}\left(z^j,\theta_0\right) \prod_{j=1}^{M} \left(f^{\frac{1}{2}}\left(z^j,\theta_0+u_0\right)\right.$$
$$\left. + \sup_{\Gamma_0} |f^{\frac{1}{2}}\left(z^j,\theta_0+u\right) - f^{\frac{1}{2}}\left(z^j,\theta_0+u_0\right)|\right)$$

Therefore

$$E \sup_{\Gamma_0} Z_M^{\frac{1}{2}}(u) \leq \left(X_{\Gamma_0}\right)^M$$

with

$$X_{\Gamma_0} = \int_{R^d} f^{\frac{1}{2}}(z,\theta_0)\, f^{\frac{1}{2}}(z,\theta_0+u_0)dz +$$
$$+ \int_{R^d} f^{\frac{1}{2}}(z,\theta_0) \sup_{\Gamma_0} |f^{\frac{1}{2}}(z,\theta_0+u) - f^{\frac{1}{2}}(z,\theta_0+u_0)\,|dz$$

We then use

$$\int_{R^d} f^{\frac{1}{2}}(z,\theta_0) f^{\frac{1}{2}}(z,\theta_0+u_0)dz = 1 - \frac{1}{2}\int_{R^d}\left(f^{\frac{1}{2}}(z,\theta_0) - f^{\frac{1}{2}}(z,\theta_0+u_0)\right)^2 dz$$
$$\leq 1 - \frac{1}{2}\pi_{\theta_0}(\gamma)$$

Next

$$\int_{R^d} f^{\frac{1}{2}}(z,\theta_0) \sup_{\Gamma_0} |f^{\frac{1}{2}}(z,\theta_0+u) - f^{\frac{1}{2}}(z,\theta_0+u_0)|dz \leq$$
$$\sqrt{\int_{R^d} \sup_{\Gamma_0} |f^{\frac{1}{2}}(z,\theta_0+u) - f^{\frac{1}{2}}(z,\theta_0+u_0)|^2 dz} = \varpi_{\theta_0+u_0}(\delta)$$

Collecting results we can write

$$X_{\Gamma_0} \leq 1 - \frac{1}{2}\pi_{\theta_0}(\gamma) + \varpi_{\theta_0+u_0}(\delta)$$

Therefore

$$E \sup_{\Gamma_0} Z_M^{\frac{1}{2}}(u) \leq \left(1 - \frac{1}{2}\pi_{\theta_0}(\gamma) + \varpi_{\theta_0+u_0}(\delta)\right)^M$$
$$\leq \exp -M\left(\frac{1}{2}\pi_{\theta_0}(\gamma) - \varpi_{\theta_0+u_0}(\delta)\right) \tag{25}$$

where we have used the elementary inequality

$$a + 1 \leq \exp a \ , \forall a$$

applied with $1 + a > 0$.

For any vector u such that $|u| \geq \gamma$ and $\theta_0 + u \in \bar{\Theta}$, we consider the ball of center $\theta_0 + u$ with radius δ. We obtain a covering of the set $\{\theta_0 + u \in \bar{\Theta}, |u| \geq \gamma\}$. Since this set is compact, we obtain a finite covering of this set by balls Γ_j with center $\theta_0 + u_j$ and radius δ, with $j = 1, \cdots J$. Hence

$$\sup_{|u|>\gamma} Z_M^{\frac{1}{2}}(u)| \leq \sum_{j=1}^{J} \sup_{\Gamma_j} Z_M^{\frac{1}{2}}(u)$$

and

$$\begin{aligned} E \sup_{|u|>\gamma} Z_M^{\frac{1}{2}}(u)| &\leq \sum_{j=1}^{J} \exp -M \left(\frac{1}{2} \pi_{\theta_0}(\gamma) - \varpi_{\theta_0+u_j}(\delta) \right) \\ &\leq J \exp -M \left(\frac{1}{2} \pi_{\theta_0}(\gamma) - \sup_{\theta \in \bar{\Theta}} \varpi_\theta(\delta) \right) \end{aligned}$$

From the property (24) we can choose δ sufficiently small so that

$$\sup_{\theta \in \bar{\Theta}} \varpi_\theta(\delta) \leq \frac{1}{4} \pi_{\theta_0}(\gamma)$$

hence

$$E \sup_{|u|>\gamma} Z_M^{\frac{1}{2}}(u)| \leq J \ \exp -M \left(\frac{1}{4} \pi_{\theta_0}(\gamma) \right)$$

Now, from (18) we have

$$\begin{aligned} P\left(\left\{ |\hat{\theta}_M - \theta_0| > \gamma \right\} \right) &\leq P\left(\left\{ \sup_{|u|>\gamma} Z_M(u)| \geq 1 \right\} \right) \\ &= P\left(\left\{ \sup_{|u|>\gamma} Z_M^{\frac{1}{2}}(u)| \geq 1 \right\} \right) \\ &\leq E \sup_{|u|>\gamma} Z_M^{\frac{1}{2}}(u)| \\ &\leq J \ \exp -M \left(\frac{1}{4} \pi_{\theta_0}(\gamma) \right) \end{aligned}$$

It follows that

$$P\left(\cup_{M\geq M_0}\left\{|\hat{\theta}_M-\theta_0|>\gamma\right\}\right)\leq\sum_{M\geq M_0}P\left(\left\{|\hat{\theta}_M-\theta_0|>\gamma\right\}\right)$$

$$\leq J\sum_{M\geq M_0}\exp -M\left(\frac{1}{4}\pi_{\theta_0}(\gamma)\right)$$

$$=J\,\frac{\exp -M_0\left(\frac{1}{4}\pi_{\theta_0}(\gamma)\right)}{1-\exp -(\frac{1}{4}\pi_{\theta_0}(\gamma))}\rightarrow 0,\text{ as } M_0\rightarrow+\infty$$

Therefore

$$P\left(\cap_{M_0=1}^{\infty}\cup_{M\geq M_0}\left\{|\hat{\theta}_M-\theta_0|>\gamma\right\}\right)=0$$

Since γ is arbitrary, a.s. the sequence $\hat{\theta}_M$ cannot have an accumulation point different from θ_0, which implies (23), and completes the proof. ■

3.3 Asymptotic Normality

Let us consider the log likelihood

$$L(z,\theta)=\log f(z,\theta)$$

The MLE maximizes in θ

$$\sum_{j=1}^{M}L\left(z^j,\theta\right)$$

in the open domain Θ. Therefore, we can write

$$\sum_{j=1}^{M}D_\theta L\left(z^j,\hat{\theta}_M\right)=0 \tag{26}$$

We assume some regularity on the derivatives of the log likelihood function. Namely, there exists δ, $\frac{1}{2}<\delta<1$ such that

$$E|D_\theta L(z,\theta_0)|^{1+\delta}<+\infty \tag{27}$$

$$E||\,D_\theta^2L(z,\theta_0)||^{1+\delta}<+\infty \tag{28}$$

Next define

$$R(z,\theta)=\sup_{\theta'\in\Theta}\frac{||\,D_\theta^2L(z,\theta')-D^2L(z,\theta)||}{|\theta'-\theta|^\delta} \tag{29}$$

and assume

$$ER(z,\theta_0) < +\infty \tag{30}$$

Consider the Fisher information matrix

$$I(\theta_0) = -ED_\theta^2 L(z,\theta_0) \tag{31}$$

It is well known that

$$I(\theta_0) = E\left(D_\theta L(z,\theta_0)(D_\theta L(z,\theta_0))^*\right) \geq 0$$

We assume

$$I(\theta_0) \text{ invertible} \tag{32}$$

and state the main result

Theorem 4 *We use the assumptions of Theorem 1 and (27), (28), (30), (32). We then have the property*

$$\sqrt{M}(\hat{\theta}_M - \theta_0) \to N(0, I^{-1}(\theta_0)) \tag{33}$$

the convergence being in law, and the limit is Gaussian, with mean 0 *and covariance matrix* $I^{-1}(\theta_0)$.

Since Theorem 1 holds, we know that $\hat{\theta}_M - \theta_0 \to 0$ a.s. We will prove a stronger result

Lemma 5 *We have*

$$M^{\frac{\delta}{1+\delta}}(\hat{\theta}_M - \theta_0) \to 0, \ a.s. \tag{34}$$

Proof We note that $\frac{\delta}{1+\delta} < \frac{1}{2}$, since $\delta < 1$. We recall the property

$$ED_\theta L(z,\theta_0) = 0 \tag{35}$$

From (26) we write

$$\begin{aligned}
\sum_{j=1}^M D_\theta L(z^j,\theta_0) &= \sum_{j=1}^M (D_\theta L(z^j,\theta_0) - D_\theta L(z^j,\hat{\theta}_M)) \\
&= -\sum_{j=1}^M \int_0^1 D_\theta^2 L(z^j, \theta_0 + \lambda(\hat{\theta}_M - \theta_0))d\lambda\,(\hat{\theta}_M - \theta_0) \\
&= -\sum_{j=1}^M \int_0^1 D_\theta^2 L(z^j,\theta_0)(\hat{\theta}_M - \theta_0)
\end{aligned}$$

$$
-\sum_{j=1}^{M}\int_0^1 \Big(D_\theta^2 L(z^j,\theta_0+\lambda(\hat\theta_M-\theta_0))-D_\theta^2 L(z^j,\theta_0)\Big)
$$
$$
d\lambda\,(\hat\theta_M-\theta_0)
$$

It follows

$$
\sum_{j=1}^{M} D_\theta L(z^j,\theta_0)=M\,I(\theta_0)(\hat\theta_M-\theta_0)
$$
$$
-\sum_{j=1}^{M}\left(D_\theta^2 L(z^j,\theta_0)-E D_\theta^2 L(z^j,\theta_0)\right)(\hat\theta_M-\theta_0)
$$
$$
-\sum_{j=1}^{M}\int_0^1 \Big(D_\theta^2 L(z^j,\theta_0+\lambda(\hat\theta_M-\theta_0))-D_\theta^2 L(z^j,\theta_0)\Big)
$$
$$
d\lambda\,(\hat\theta_M-\theta_0)
$$

This is an equality between vectors in R^k. We multiply to the left by the line vector

$$
\frac{(\hat\theta_M-\theta_0)^*}{|\hat\theta_M-\theta_0|}
$$

and obtain

$$
\frac{M}{|\hat\theta_M-\theta_0|}(\hat\theta_M-\theta_0)^* I(\theta_0)(\hat\theta_M-\theta_0)=\frac{(\hat\theta_M-\theta_0)^*\sum_{j=1}^M D_\theta L(z^j,\theta_0)}{|\hat\theta_M-\theta_0|}+
$$

$$
+\frac{(\hat\theta_M-\theta_0)^*\sum_{j=1}^M\left(D_\theta^2 L(z^j,\theta_0)-E D_\theta^2 L(z^j,\theta_0)\right)(\hat\theta_M-\theta_0)}{|\hat\theta_M-\theta_0|}+
$$

$$
+\frac{(\hat\theta_M-\theta_0)^*\sum_{j=1}^M\int_0^1\Big(D_\theta^2 L(z^j,\theta_0+\lambda(\hat\theta_M-\theta_0))-D_\theta^2 L(z^j,\theta_0)\Big)d\lambda\,(\hat\theta_M-\theta_0)}{|\hat\theta_M-\theta_0|}
$$

Since $I(\theta_0)$ is invertible, we deduce

$$
\alpha M|\hat\theta_M-\theta_0|\le|\sum_{j=1}^M D_\theta L(z^j,\theta_0)|+ \tag{36}
$$
$$
+||\sum_{j=1}^M\left(D_\theta^2 L(z^j,\theta_0)-E D_\theta^2 L(z^j,\theta_0)\right)|||\hat\theta_M-\theta_0|+(\sum_{j=1}^M R(z^j,\theta_0))|\hat\theta_M-\theta_0|^{1+\delta}
$$

where $\alpha>0$.

We set

$$\mathcal{R}_M = \frac{\sum_{j=1}^{M} R(z^j,\theta_0)}{M}$$

Since the random variables $R(z^j,\theta_0)$ are independent i.i.d and $ER(z^j,\theta_0) = ER(z,\theta_0) < +\infty$, we can refer to Kolmogorov strong law of large numbers to claim that

$$\mathcal{R}_M \to ER(z,\theta_0), \text{ a.s.} \tag{37}$$

Therefore

$$\mathcal{R}_M|\hat{\theta}_M - \theta_0|^{\delta} \to 0, \text{ a.s.} \tag{38}$$

From the assumptions (27), (28), we can assert that

$$\frac{\sum_{j=1}^{M} D_\theta L(z^j,\theta_0)}{M^{\frac{1}{1+\delta}}} \to 0, \; \frac{\sum_{j=1}^{M} \left(D_\theta^2 L(z^j,\theta_0) - ED_\theta^2 L(z^j,\theta_0)\right)}{M^{\frac{1}{1+\delta}}} \to 0, \text{ a.s.} \tag{39}$$

This follows from a result of Marcinkewicz, whose proof can be found in Loeve (1978). The result is the following: let $\xi_1,\cdots\xi_n,\cdots$ be independent identically distributed random variables, such that $E|\xi_n|^{1+\delta} < \infty$, $0 \le \delta < 1$ then

$$\frac{\sum_{j=1}^{n} (\xi_j - E\xi_j)}{n^{\frac{1}{1+\delta}}} \to 0, \text{ a.s.}$$

From (36), we can write

$$M^{\frac{\delta}{1+\delta}}|\hat{\theta}_M - \theta_0|(\alpha - \mathcal{R}_M|\hat{\theta}_M - \theta_0|^{\delta}) \le \left|\frac{\sum_{j=1}^{M} D_\theta L(z^j,\theta_0)}{M^{\frac{1}{1+\delta}}}\right| + \left|\frac{\sum_{j=1}^{M} \left(D_\theta^2 L(z^j,\theta_0) - ED_\theta^2 L(z^j,\theta_0)\right)}{M^{\frac{1}{1+\delta}}}\right|$$

□

and the result (34) follows immediately. ■

We turn now to the

Proof of Theorem 4 We write now

$$\sqrt{M}\, I(\theta_0)(\hat{\theta}_M - \theta_0) = \frac{1}{\sqrt{M}} \sum_{j=1}^{M} D_\theta L(z^j,\theta_0) + \Gamma_1^M + \Gamma_2^M$$

with

$$\Gamma_1^M = \frac{1}{\sqrt{M}} \sum_{j=1}^{M} \left(D_\theta^2 L(z^j,\theta_0) - ED_\theta^2 L(z^j,\theta_0)\right)(\hat{\theta}_M - \theta_0)$$

and

$$\Gamma_2^M = \frac{1}{\sqrt{M}} \sum_{j=1}^{M} \int_0^1 \left(D_\theta^2 L(z^j, \theta_0 + \lambda(\hat{\theta}_M - \theta_0)) - D_\theta^2 L(z^j, \theta_0) \right) d\lambda \, (\hat{\theta}_M - \theta_0)$$

Then

$$\begin{aligned} |\Gamma_2^M| &\leq \frac{1}{\sqrt{M}} \sum_{j=1}^{M} R(z^j, \theta_0) |\hat{\theta}_M - \theta_0|^{1+\delta} \\ &= \mathcal{R}_M \sqrt{M} |\hat{\theta}_M - \theta_0|^{1+\delta} \end{aligned}$$

Since $\delta > \frac{1}{2}$, $M^{\frac{1}{2(1+\delta)}} (\hat{\theta}_M - \theta_0) \to 0$, a.s. Since $\mathcal{R}_M$ is a.s. bounded we get $\Gamma_2^M \to 0$ a.s. Next we write

$$\Gamma_1^M = \frac{\sum_{j=1}^{M} \left(D_\theta^2 L(z^j, \theta_0) - E D_\theta^2 L(z^j, \theta_0) \right)}{M^{\frac{1}{1+\delta}}} M^{\frac{1-\delta}{2(1+\delta)}} (\hat{\theta}_M - \theta_0)$$

Since $\frac{1-\delta}{2} < \delta$, we can assert that $M^{\frac{1-\delta}{2(1+\delta)}} (\hat{\theta}_M - \theta_0) \to 0$,a.s. Thanks also to the second part of (39) we can conclude that $\Gamma_1^M \to 0$, a.s.

Furthermore, it is standard that

$$\frac{1}{\sqrt{M}} \sum_{j=1}^{M} D_\theta L(z^j, \theta_0) \to N(0, I(\theta_0))$$

in law. Hence

$$\frac{1}{\sqrt{M}} I(\theta_0)^{-1} \sum_{j=1}^{M} D_\theta L(z^j, \theta_0) \to N(0, (I(\theta_0))^{-1})$$

in law. This implies the result (33). ■

4 MLE for Generalized Linear Models

4.1 Statement of the Problem and Notation

Consider now a sequence of independent random variables $z^1, \cdots, z^M$ which follow GLM distributions with canonical parameters $\theta^1, \cdots, \theta^M$. We continue to assume these variables scalar, to simplify. So the canonical parameters are also scalar. These canonical parameters are linked to the variables of interest x by the relations

$$(h^j)^* x = g(b'(\theta^j)) \tag{40}$$

So the link function $g(.)$ and the function $b(\theta)$ are identical for all the variables. We define the functions $\mu^j(x)$ by solving

$$(h^j)^* x = g(\mu^j(x)) \tag{41}$$

which is possible, since g is invertible. Similarly we define the functions $\theta^j(x)$ by solving

$$\mu^j(x) = b'(\theta^j(x)) \tag{42}$$

and

$$V^j(x) = b"(\theta^j(x)) \tag{43}$$

Recalling the function

$$f(z,\theta) = \exp\left(z\theta - b(\theta) + c(z)\right) \tag{44}$$

then the probability density of the variable z^j is $f(z, \theta^j(x_0))$, in which we have denoted by x_0 the true value of the parameter. We note that the variables z^j are not identically distributed, which introduces a slight difficulty with respect to the MLE developped in the previous section. The loglikelihood function is $L(z, \theta^j(x))$, where $L(z,\theta) = \log\ f(z,\theta)$. The joint probability density of the sample $z^1, \cdots, z^M$ is

$$\prod_{j=1}^{M} f(z^j, \theta^j(x_0)) \tag{45}$$

The MLE is obtained by maximizing the function of x

$$\prod_{j=1}^{M} f(z^j, \theta^j(x)) \tag{46}$$

As we have done for the MLE in general, we shall assume, to simplify technicalities that we know a bounded convex open domain denoted $\mathcal{X}$, and

$$x_0 \in \mathcal{X}$$

So we maximize the function (46) on $\mathcal{X}$ and we assume that such a maximum exists denoted by $\hat{x}_M$. To prove the asymptotic properties of $\hat{x}_M$, we shall adapt the methods used for the MLE in general.

We first introduce a notation. From formula (40) we can write

$$\theta^j(x) = \varphi((h^j)^* x) \tag{47}$$

with

$$\varphi(\eta) = b'^{-1}(g^{-1}(\eta)) \tag{48}$$

hence

$$D_x \theta^j(x) = \varphi'((h^j)^* x) h^j$$

then

$$\begin{aligned} D_x L(z, \theta^j(x)) &= (z - b'(\theta^j(x)))D_x\theta^j(x) \\ &= (z - \mu^j(x))\varphi'((h^j)^* x)h^j \\ &= (z - b'(\varphi((h^j)^* x)))\varphi'((h^j)^* x)h^j \\ &= [z\varphi'((h^j)^* x) - (b\text{o}\varphi)'((h^j)^* x)]h^j \end{aligned} \tag{49}$$

therefore

$$D_x^2 L(z, \theta^j(x)) = \left[z\varphi"((h^j)^* x) - (b\text{o}\varphi)"((h^j)^* x)\right] h^j(h^j)^* \tag{50}$$

Note that

$$\varphi'(\eta) = \frac{1}{g'(b'(\varphi(\eta)))b"(\varphi(\eta))}$$

therefore

$$\varphi'((h^j)^* x) = \frac{1}{g'(b'(\theta^j(x)))b"(\theta^j(x))}$$

Recalling

$$\mu^j(x) = b'(\theta^j(x)),\ V^j(x) = b"(\theta^j(x))$$

then

$$\varphi'((h^j)^* x) = \frac{1}{g'(\mu^j(x))V^j(x)}$$

It is convenient to introduce the weights

$$W^j(x) = \frac{1}{(g')^2(\mu^j(x))V^j(x)}$$

therefore

$$\varphi'((h^j)^* x) = g'(\mu^j(x))W^j(x) \tag{51}$$

Also, we have

$$\mu^j(x) = b'\text{o}\varphi((h^j)^* x)$$

Next, we note that

$$b'\text{o}\varphi(\eta)\varphi"(\eta) - (b\text{o}\varphi)"(\eta) = -b"(\varphi(\eta))(\varphi')^2(\eta)$$

Therefore, from (50) we get

$$\begin{aligned} D_x^2 L(z, \theta^j(x)) &= \left[(z - \mu^j(x))\varphi"((h^j)^* x) - V^j(x)(\varphi')^2((h^j)^* x) \right] h^j (h^j)^* \\ &= \left[(z - \mu^j(x))\varphi"((h^j)^* x) - W^j(x) \right] h^j (h^j)^* \end{aligned} \tag{52}$$

Considering the true value of the parameter x_0, we note that

$$E D_x L \left(z^j, \theta^j(x_0) \right) = 0$$

hence

$$E z^j = \mu^j(x_0) \tag{53}$$

and

$$E D_x^2 L \left(z^j, \theta^j(x_0) \right) = -W^j(x_0) h^j (h^j)^* \tag{54}$$

Also

$$\begin{aligned} E D_x^2 L(z^j, \theta^j(x_0)) &= -E \left(D_x L(z^j, \theta^j(x_0)) \right) \left(D_x L(z^j, \theta^j(x_0)) \right)^* \\ &= -E \left((z - \mu^j(x_0)) \right)^2 \left(\varphi'((h^j)^* x_0) \right)^2 h^j \left(h^j \right)^* \end{aligned}$$

It follows

$$\begin{aligned} W^j(x_0) &= E \left((z - \mu^j(x_0)) \right)^2 \left(\varphi'((h^j)^* x_0) \right)^2 \\ &= E \left((z - \mu^j(x_0)) \right)^2 \frac{1}{(g'(\mu^j(x_0)) V^j(x_0))^2} \\ \frac{1}{(g')^2(\mu^j(x_0)) V^j(x_0)} &= E((z - \mu^j(x_0)))^2 \frac{1}{(g'(\mu^j(x_0)) V^j(x_0))^2} \end{aligned}$$

which implies the interpretation

$$E((z - \mu^j(x_0)))^2 = V^j(x_0) \tag{55}$$

We recover of course (5).

4.2 Examples

In the Gaussian case, $g(\mu) = \mu$ and $b(\theta) = \frac{1}{2}\sigma\theta^2$. We get easily

$$D_x L(Z; x) = \frac{1}{\sigma} \sum_{j=1}^{M} \left(z^j - (h^j)^* x \right) h^j \tag{56}$$

and thus $\hat{x}_M$ satisfies

$$\sum_{j=1}^{M} z^j h^j = \sum_{j=1}^{M} h^j \left(h^j\right)^* \hat{x}_M \tag{57}$$

and this system has one and only one solution provided the matrix $\sum_{j=1}^{M} h^j (h^j)^*$ which belongs to $\mathcal{L}(R^n; R^n)$ is invertible.

Let us consider the Weibull distribution case, with known shape k, see Sect. 2.6. We have for the variables $z^j = (y^j)^k$

$g(\mu) = \mu^{\frac{1}{k}}, b(\theta) = -\log -\theta$. Hence

$$\mu^j(x) = ((h^j)^* x)^k, \quad \theta^j(x) = -\frac{1}{\mu^j(x)} \tag{58}$$

$$V^j(x) = (\mu^j(x))^2, \quad W^j(x) = \frac{k^2}{(\mu^j(x))^{\frac{2}{k}}}$$

We obtain the system

$$\sum_{j=1}^{M} \left(\frac{z^j}{\mu^j(\hat{x}_M)} - 1 \right) \frac{h^j}{(h^i)^* \hat{x}_M} = 0 \tag{59}$$

Let us finally consider the nonlinear Gaussian case, see Sect. 2.7. We have

$$g(\mu) = \varphi^{-1}(\mu), \quad b(\theta) = \frac{\sigma^2 \theta^2}{2}$$

hence

$$\mu^j(x) = \varphi((h^j)^* x), \quad \theta^j(x) = \frac{\mu^j(x)}{\sigma^2}$$

$$V^j(x) = \sigma^2, \quad W^j(x) = \frac{(\varphi'((h^j)^* x))^2}{\sigma^2} \tag{60}$$

We obtain the system

$$\sum_{j=1}^{M} (z^j - \varphi((h^j)^* \hat{x}_M)) \varphi'((h^j)^* \hat{x}_M) h^j = 0 \tag{61}$$

4.3 Consistency

We begin with consistency. We will need an assumption of linear independence of the vectors h^j. More precisely, let us consider for $i = 1, \cdots$ the $n \times n$ matrix

$$H^i = \begin{pmatrix} (h^{(i-1)n+1})^* \\ \vdots \\ (h^{in})^* \end{pmatrix}$$

We assume

$$H^i \text{ is invertible } \forall i \tag{62}$$

and

$$||H^i||,\ ||(H^i)^{-1}|| \leq C \tag{63}$$

We shall need the following assumption

$$\int_R \sup_{\theta \in \Theta} f(z,\theta)dz < \infty,\ \int_R z^2 \sup_{\theta \in \Theta} f(z,\theta)dz < \infty\ \forall \Theta \text{ compact interval} \tag{64}$$

Theorem 6 *We consider the GLM defined by* (44) *with $b(\theta)$ $C^2$and strictly convex. We assume that the link function $g(\mu)$ has an inverse and is C^1. We also assume* (62), (63) *and* (64)*. Then we have the consistency property*

$$\hat{x}_M \rightarrow x_0\ a.s. \tag{65}$$

Proof We will operate as in Theorem 1. We need to obtain properties similar to (19) and (20). Consider the functions $\theta^j(x)$ and define vector functions (with values in R^n)

$$\bar{\theta}^i(x) = \begin{pmatrix} \theta^{(i-1)n+1}(x) \\ \vdots \\ \theta^{in}(x) \end{pmatrix}$$

We next define a sequence of probability densities in R^n, depending on the vector $\bar{\theta}^i(x)$, given by the formula

$$\bar{f}\left(\bar{z}, \bar{\theta}^i(x)\right) = \prod_{l=1}^{n} f\left(\bar{z}_l, \theta^{(i-1)n+l}(x)\right) \tag{66}$$

where the argument $\bar{z} \in R^n$ and $\bar{z}_l, l = 1 \cdots n$. We define the random vector $\bar{z}^i \in R^n$, by

$$\bar{z}^i = \begin{pmatrix} z^{(i-1)n+1} \\ \vdots \\ z^{in} \end{pmatrix}$$

We notice that the sequence of scalar random variables $z^1, \cdots, z^{nM}$ is equivalent to the sequence of vector random variables $\bar{z}^1, \cdots, \bar{z}^M$.

We first consider the random function

$$Z_{nM}(u) = \prod_{j=1}^{nM} \frac{f\left(z^j, \theta^j(x_0+u)\right)}{f(z^j, \theta^j(x_0)}$$

and we can write

$$\begin{aligned} Z_{nM}(u) &= \bar{Z}_M(u) \\ &= \prod_{i=1}^{M} \frac{\bar{f}\left(\bar{z}^i, \bar{\theta}^i(x_0+u)\right)}{\bar{f}(\bar{z}^i, \bar{\theta}^i(x_0))} \end{aligned}$$

Let u_0 with $|u_0| \geq \gamma$ and $x_0 + u_0 \in \bar{\mathcal{X}}$. We consider the sphere of center $x_0 + u_0$ and of radius δ. We call it Γ_0. We shall estimate

$$E \sup_{\Gamma_0} \bar{Z}_M^{\frac{1}{2}}(u) = E \sup_{\{u | x_0+u \in \Gamma_0\}} \bar{Z}_M^{\frac{1}{2}}(u)$$

Writing

$$\begin{aligned} \sup_{\Gamma_0} \bar{Z}_M^{\frac{1}{2}}(u) \leq \prod_{i=1}^{M} [&\bar{f}^{-\frac{1}{2}}\left(\bar{z}^i, \bar{\theta}^i(x_0)\right) \bar{f}^{\frac{1}{2}}(\bar{z}^i, \bar{\theta}^i(x_0+u_0)) \\ &+ \bar{f}^{-\frac{1}{2}}\left(\bar{z}^i, \bar{\theta}^i(x_0)\right) \sup_{\Gamma_0} |\bar{f}^{\frac{1}{2}}\left(\bar{z}^i, \bar{\theta}^i(x_0+u)\right) - \bar{f}^{\frac{1}{2}}\left(\bar{z}^i, \bar{\theta}^i(x_0+u_0)\right)|] \end{aligned}$$

we get

$$E \sup_{\Gamma_0} \bar{Z}_M^{\frac{1}{2}}(u) \leq \prod_{i=1}^{M} \bar{X}^i_{\Gamma_0}$$

with

$$\begin{aligned} \bar{X}^i_{\Gamma_0} = \int_{R^n} &\bar{f}^{\frac{1}{2}}\left(\bar{z}, \bar{\theta}^i(x_0)\right) [\bar{f}^{\frac{1}{2}}\left(\bar{z}, \bar{\theta}^i(x_0+u_0)\right) \\ &+ \sup_{\Gamma_0} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0+u)) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0+u_0))|] d\bar{z} \end{aligned}$$

We first have

$$\begin{aligned} \int_{R^n} &\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0)) \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0+u_0)) d\bar{z} = 1 - \\ &- \frac{1}{2} \int_{R^n} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0+u_0)) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0))|^2 d\bar{z} \end{aligned}$$

Next

$$\int_{R^n} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0+u_0)) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0))|^2 d\bar{z} \geq \tag{67}$$

$$\inf_{\left\{\begin{array}{c} x \in \mathcal{X} \\ |x-x_0| \geq \gamma \end{array}\right\}} \int_{R^n} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x)) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0))|^2 d\bar{z}$$

We recall the relations

$$(h^{(i-1)n+l})^* x = g(b'(\theta^{(i-1)n+l}(x)))$$

For $x \in \mathcal{X}$, it follows from the properties of the functions g and b, and from the assumption (63) that $\theta^{(i-1)n+l}(x)$ is bounded, so $\bar{\theta}^i(x)$ lies in a compact set $\bar{\Theta}^n$of R^n, where $\bar{\Theta}$ is a compact interval. Since

$$\left(h^{(i-1)n+l}\right)^* (x-x_0) = g\left(b'(\theta^{(i-1)n+l}(x))\right) - g\left(b'(\theta^{(i-1)n+l}(x_0))\right)$$

we deduce easily, using the fact that g is C^1, b is C^2 and the bounds on the arguments

$$|\left(h^{(i-1)n+l}\right)^* (x-x_0)| \leq c|\theta^{(i-1)n+l}(x) - \theta^{(i-1)n+l}(x_0)|$$

This can also be written as

$$|H^i(x-x_0)| \leq c|\bar{\theta}^i(x) - \bar{\theta}^i(x_0)|$$

From the assumptions (62), (63) we obtain also

$$|x-x_0| \leq \rho|H^i(x-x_0)| \leq c\rho|\bar{\theta}^i(x) - \bar{\theta}^i(x_0)|$$

Therefore for $x \in \mathcal{X}$ and $|x-x_0| \geq \gamma$ we get

$$|\bar{\theta}^i(x) - \bar{\theta}^i(x_0)| \geq \beta = \frac{\gamma}{c\rho}$$

Collecting results we obtain

$$\int_{R^n} |\bar{f}^{\frac{1}{2}}\left(\bar{z}, \bar{\theta}^i(x_0+u_0)\right) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0))|^2 d\bar{z} \geq$$

$$\inf_{\left\{\begin{array}{c} \bar{\theta}, \bar{\theta}' \in \bar{\Theta}^n \\ |\bar{\theta}-\bar{\theta}'| \geq \beta \end{array}\right\}} \int_{R^n} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}')|^2 d\bar{z} = \pi(\beta)$$

And we claim that $\pi(\beta) > 0$,for $\beta > 0$. Since $\bar{\Theta}^n$ is compact, it is easy to check that if $\pi(\beta) = 0$, then there would exist $\bar{\theta}$ and $\bar{\theta}'$ in $\bar{\Theta}^n$ such that

$$|\bar{\theta} - \bar{\theta}'| \geq \beta, \ \bar{f}(\bar{z}, \bar{\theta}) = \bar{f}(\bar{z}, \bar{\theta}'), \ \forall \bar{z} \in R^n$$

Since

$$\bar{f}(\bar{z}, \bar{\theta}) = \exp \sum_{l=1}^{n} (\bar{z}_l \bar{\theta}_l - b(\bar{\theta}_l) + c(\bar{z}_l))$$

we need to have

$$\sum_{l=1}^{n} (\bar{z}_l \bar{\theta}_l - b(\bar{\theta}_l)) = \sum_{l=1}^{n} \left(\bar{z}_l \bar{\theta}'_l - b(\bar{\theta}'_l)\right)$$

for any real $\bar{z}_l$, $l = 1, \cdots n$. Suppose there is l_0 such that $\bar{\theta}_{l_0} \neq \bar{\theta}'_{l_0}$. We take

$$\bar{z}_l = \frac{b(\bar{\theta}_l) - b(\bar{\theta}'_l)}{\bar{\theta}_l - \bar{\theta}'_l}, \text{ if } \bar{\theta}_l - \bar{\theta}'_l \neq 0$$

$$\bar{z}_{l_0} = 0, \ \bar{z}_l \text{ arbitrary if } \bar{\theta}_l - \bar{\theta}'_l = 0$$

It clearly follows that $b(\bar{\theta}_{l_0}) = b\left(\bar{\theta}'_{l_0}\right)$ and from the invertibility of the function b, $\bar{\theta}_{l_0} = \bar{\theta}'_{l_0}$, which is a contradiction. Hence $\bar{\theta}_l = \bar{\theta}_{l'}$ which is impossible. Finally we have obtained

$$\int_{R^n} \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0)) \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0 + u_0)) d\bar{z} \leq 1 - \frac{1}{2}\pi(\beta) \tag{68}$$

We next write

$$\int_{R^n} \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0)) \sup_{\Gamma_0} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0 + u)) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0 + u_0))| d\bar{z} \leq$$

$$\sqrt{\int_{R^n} \sup_{\Gamma_0} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0 + u)) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0 + u_0))|^2 d\bar{z}} \leq$$

$$\sqrt{\int_{R^n} \sup_{\left\{\begin{array}{c} x \in \mathcal{X} \\ |x' - x| \leq \delta \end{array}\right\}} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x')) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x))|^2 d\bar{z}} \leq$$

$$\sqrt{\int_{R^n} \sup_{\left\{\begin{array}{c} \bar{\theta} \in \bar{x}^n \\ |\bar{\theta}' - \bar{\theta}| \leq c(\delta) \end{array}\right\}} |\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}') - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta})|^2 d\bar{z}} = \varpi(\delta)$$

and, we see easily using (64) that $\varpi(\delta) \to 0$, as $\delta \to 0$. Therefore, we have obtained, recalling (68) that

$$\bar{X}^i_{\Gamma_0} \le 1 - \frac{1}{2}\pi(\beta) + \varpi(\delta)$$

and

$$E \sup_{\Gamma_0} Z^{\frac{1}{2}}_{nM}(u) \le \exp - M\left(\frac{1}{2}\pi(\beta) - \varpi(\delta)\right) \tag{69}$$

We next write

$$Z^{\frac{1}{2}}_M(u) = \bar{Z}^{\frac{1}{2}}_{n\left[\frac{M}{n}\right]}(u) \prod_{j=n\left[\frac{M}{n}\right]+1}^{M} \frac{f(z^j, \theta^j(x_0+u))}{f(z^j, \theta^j(x_0)}$$

Therefore

$$E \sup_{\Gamma_0} Z^{\frac{1}{2}}_M(u) \le E \sup_{\Gamma_0} \bar{Z}^{\frac{1}{2}}_{n\left[\frac{M}{n}\right]}(u) \prod_{i=n\left[\frac{M}{n}\right]+1}^{M} \int_R \sup_{\Gamma_0} f^{\frac{1}{2}}(z, \theta^j(x_0) f^{\frac{1}{2}}(z, \theta^j(x_0+u)dz$$

$$\le E \sup_{\Gamma_0} \bar{Z}^{\frac{1}{2}}_{n\left[\frac{M}{n}\right]}(u) \left(\sqrt{\int_R \sup_{\theta\in\bar{\Theta}} f(z,\theta)dz}\right)^n$$

where we have used the fact

$$\sqrt{\int_R \sup_{\theta\in\bar{\Theta}} f(z,\theta)dz} > 1$$

Thanks to (64) we can assert that

$$E \sup_{\Gamma_0} Z^{\frac{1}{2}}_M(u) \le C_n \exp - \left[\frac{M}{n}\right]\left(\frac{1}{2}\pi(\beta) - \varpi(\delta)\right)$$

We are then exactly in the situation of the MLE, see Theorem 1. We cover the set $\{x = x_0 + u|\ |u| \ge \gamma$ and $x_0 + u \in \bar{\mathcal{X}}\}$ with a finite number J of balls similar to Γ_0, with δ chosen such that $\varpi(\delta) \le \frac{1}{4}\pi(\beta)$. We obtain

$$E \sup_{|u|\ge\gamma} Z^{\frac{1}{2}}_M(u) \le JC_n \exp - \left[\frac{M}{n}\right]\left(\frac{1}{4}\pi(\beta)\right)$$

and

$$P\left(\{\hat{x}_M - x_0| > \gamma\}\right) \le J \exp - \left[\frac{M}{n}\right]\frac{1}{4}\pi\left(\frac{\gamma}{c\rho}\right)$$

As in Theorem 1we deduce

$$\hat{x}_M \to x_0 \text{ a.s. as} M \to +\infty$$

which concludes the proof. ∎

4.4 Further Consistency Estimates

Our objective is to prove convergence results as follows

$$M^{\beta}(\hat{x}_M - x_0) \to 0, \text{a.s. and in } L^q, \forall 1 \le q < \infty, \forall \beta < \frac{1}{2} \tag{70}$$

We cannot use the method of Lemma 5. This is because the Marcinkiewicz theorem used in this Lemma, necessitates that the variables are independent, which is not the case. We shall proceed differently, following ideas of (West et al. 1981).

It is convenient to also introduce the following notation. We know that the true value of the parameter is x_0. However, we may define the probability for which the true value is any value x. We call P_x. So the true probability is $P = P_{x_0}$.With the probability P the variables z^j are independent and have a marginal density $f(z, \theta^j(x))$. We shall assume that

$$\sup_{x \in \mathcal{X}} \sup_j E_x |z^j - \mu^j(x)|^m < +\infty, \; m > n \tag{71}$$

We have the

Proposition 7 *We make the assumptions of Theorem 6 and (1). Then the property (70) holds.*

Proof We begin with preliminary estimates. We define

$$Z_M(u) = \frac{\prod_{j=1}^{M} f(z^j, \theta^j(x_0 + \frac{u}{M^{\beta}}))}{\prod_{j=1}^{M} f(z^j, \theta^j(x_0))} \tag{72}$$

and we will consider vectors u such that $x_0 + \frac{u}{M^{\beta}} \in \mathcal{X}$. We have used for $Z_M(u)$ the same notation as in Theorem 6, but there is no risk of confusion. We recover the notation of Theorem 6 by taking $\beta = 0$. Let u, v such that $x_0 + \frac{u}{M^{\beta}}, x_0 + \frac{v}{M^{\beta}} \in \mathcal{X}$. We want to estimate

$$E|Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)|^m = E|\sum_{i=1}^{n} \int_0^1 (v_i - u_i) \frac{\partial}{\partial u_i} Z_M^{\frac{1}{m}}(u + \lambda(v-u)) d\lambda|^m$$

From Minkowsky inequality

$$E^{\frac{1}{m}} |\sum_{i=1}^{n} \int_0^1 (v_i - u_i) \frac{\partial}{\partial u_i} Z_M^{\frac{1}{m}}(u + \lambda(v-u)) d\lambda|^m \le$$

$$\sum_{i=1}^{n} E^{\frac{1}{m}} | \int_0^1 (v_i - u_i) \frac{\partial}{\partial u_i} Z_M^{\frac{1}{m}}(u + \lambda(v-u)) d\lambda|^m \le$$

$$\sum_{i=1}^{n} |v_i - u_i| E^{\frac{1}{m}} \int_0^1 |\frac{\partial}{\partial u_i} Z_M^{\frac{1}{m}}(u + \lambda(v-u))|^m d\lambda \le$$

$$|v-u|\int_0^1 \sum_{i=1}^n E^{\frac{1}{m}}|\frac{\partial}{\partial u_i}Z_M^{\frac{1}{m}}(u+\lambda(v-u))|^m d\lambda$$

Therefore

$$E|Z_M^{\frac{1}{m}}(u)-Z_M^{\frac{1}{m}}(v)|^m \leq |v-u|^m \int_0^1 \left(\sum_{i=1}^n E^{\frac{1}{m}}|\frac{\partial}{\partial u_i}Z_M^{\frac{1}{m}}(u+\lambda(v-u))|^m\right)^m d\lambda \tag{73}$$

Using the inequality

$$\left(\sum_{i=1}^n a_i\right)^m \leq n^{m-1}\sum_{i=1}^n a_i^m \tag{74}$$

for numbers $a_i > 0$,we can assert finally

$$E|Z_M^{\frac{1}{m}}(u)-Z_M^{\frac{1}{m}}(v)|^m \leq |v-u|^m n^{m-1}\sum_{i=1}^n \int_0^1 E|\frac{\partial}{\partial u_i}Z_M^{\frac{1}{m}}(u+\lambda(v-u))|^m d\lambda \tag{75}$$

To pursue the estimation, we consider

$$\frac{\partial}{\partial u_i}Z_M^{\frac{1}{m}}(u)=\frac{1}{mM^\beta}Z_M^{\frac{1}{m}}(u)\sum_{j=1}^M \frac{\partial}{\partial x_i}\log f\left(z^j,\theta^j\left(x_0+\frac{u}{M^\beta}\right)\right)$$

Therefore

$$E|\frac{\partial}{\partial u_i}Z_M^{\frac{1}{m}}(u)|^m=\frac{1}{m^m M^{m\beta}}E_{x_0+\frac{u}{M^\beta}}|\sum_{j=1}^M \frac{\partial}{\partial x_i}\log f\left(z^j,\theta^j\left(x_0+\frac{u}{M^\beta}\right)\right)|^m \tag{76}$$

Note that in (76), we take in the right hand side the expected value with respect to the probability P_x, with $x=x_0+\frac{u}{M^\beta}$. We note also that

$$E_x\frac{\partial}{\partial x_i}\log f\left(z^j,\theta^j(x)\right)=0$$

We then use the Marcinkiewicz-Zygmund inequality. Let $\xi_1,\cdots\xi_M$ be independent random variables with 0 mean, then

$$E|\sum_{j=1}^M \xi_j|^m \leq C_m E\left(\sum_{j=1}^M |\xi_j|^2\right)^{\frac{m}{2}}$$

and from (74) we deduce

$$E|\sum_{j=1}^M \xi_j|^m \leq C_m M^{\frac{m}{2}-1}E\sum_{j=1}^M |\xi_j|^m$$

Applying this inequality to (76), we get

$$E|\frac{\partial}{\partial u_i} Z_M^{\frac{1}{m}}(u)|^m \leq \frac{C_m}{m^m} \frac{1}{M^{1-m(\frac{1}{2}-\beta)}} \sum_{j=1}^{M} E_{x_0+\frac{u}{M^\beta}} |\frac{\partial}{\partial x_i} \log f\left(z^j, \theta^j\left(x_0 + \frac{u}{M^\beta}\right)\right)|^m \tag{77}$$

But, as easily seen

$$E_x|\frac{\partial}{\partial x_i} \log f\left(z^j, \theta^j(x)\right)|^m = E_x|z^j - \mu^j(x)|^m |W^j(x) g'\left(\mu^j(x)\right) h_i^j|^m$$

From the assumptions, in particular (71), we get, using C_m as a generic constant, depending only of m and of the compact $\bar{\mathcal{X}}$

$$\sup_{j, x \in \mathcal{X}} E_x|\frac{\partial}{\partial x_i} \log f\left(z^j, \theta^j(x)\right)|^m \leq C_m$$

Hence also

$$\sup_{\left\{u,v \left| \begin{array}{l} x_0 + \frac{u}{M^\beta} \in \mathcal{X} \\ x_0 + \frac{v}{M^\beta} \in \mathcal{X} \end{array} \right.\right\}} \frac{E|Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)|^m}{|v-u|^m} \leq C_m n^{m-1} M^{m(\frac{1}{2}-\beta)} \tag{78}$$

Note that

$$E|Z_M^{\frac{1}{m}}(u)|^m = EZ_M(u) = 1 \leq C_m n^{m-1} M^{m(\frac{1}{2}-\beta)}$$

We can then use a result on the uniform continuity of stochastic processes, see (1981), appendix I, Theorem 19, to claim that for $m > n$

$$E \sup_{\left\{u,v \left| \begin{array}{l} x_0 + \frac{u}{M^\beta}, x_0 + \frac{v}{M^\beta} \in \mathcal{X} \\ |u|,\ |v| \leq L,\ |u-v| \leq h \end{array} \right.\right\}} |Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)| \leq B_{n,m}|L|^{\frac{n}{m}} h^{\frac{m-n}{m}} M^{\frac{1}{2}-\beta} \tag{79}$$

where $B_{n,m}$ is a constant, depending only on n, m and the compact $\overline{\mathcal{X}}$. We proceed with another estimate. Consider

$$EZ_M^{\frac{1}{2}}(u) = \prod_{j=1}^{M} \int f^{\frac{1}{2}}(z, \theta^j(x_0)) f^{\frac{1}{2}}(z, \theta^j(\,))dz$$

We want to prove the estimate

$$EZ_M^{\frac{1}{2}}(u) \leq \exp(-|u|^2 \alpha_n M^{1-2\beta}),\ \forall u \text{ such that } x_0 + \frac{u}{M^\beta} \in \mathcal{X} \tag{80}$$

where the constant α_n is strictly positive, and depends only on n and on the compact $\bar{\mathcal{X}}$. We introduce the probability $\bar{f}(\bar{z}, \bar{\theta}^i(x))$ defined in (66) and the random function

$$\bar{Z}_M(u) = \frac{\prod_{i=1}^{M} \bar{f}\left(\bar{z}^i, \bar{\theta}^i(x_0 + \frac{u}{M^\beta})\right)}{\prod_{i=1}^{M} \bar{f}\left(\bar{z}^i, \bar{\theta}^i(x_0)\right)}$$

with the notation of Theorem 6. We recall that

$$\bar{Z}_M(u) = Z_{nm}(u)$$

Since

$$EZ_M^{\frac{1}{2}}(u) \leq E\bar{Z}^{\frac{1}{2}}_{n\left[\frac{M}{n}\right]}(u) \left(\sqrt{\int_R \sup_{\theta \in \bar{\Theta}} f(z, \theta) dz}\right)^n$$

It is sufficient to estimate $E\bar{Z}_M(u)$. We recall that $\bar{\theta}^i(x) \in \bar{\Theta}^n$, $\forall x \in \mathcal{X}$, where $\bar{\Theta}$ is a compact interval. We have

$$E\bar{Z}_M^{\frac{1}{2}}(u) = \prod_{i=1}^{m} [1 - \frac{1}{2} \int_{R^n} \left(\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i \left(x_0 + \frac{u}{M^\beta}\right)\right) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0))^2 d\bar{z}] \tag{81}$$

We are going to check that

$$\inf_{\bar{\theta}, \bar{\theta}' \in \bar{\Theta}^n} \frac{\int_{R^n} (\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}'))^2 d\bar{z}}{|\bar{\theta} - \bar{\theta}'|^2} = \beta > 0 \tag{82}$$

Suppose that (82) is not true, then considering a minimizing sequence $\bar{\theta}_k, \bar{\theta}'_k$ we must have $\bar{\theta}_k$-$\bar{\theta}'_k \to 0$. Indeed, if $\bar{\theta}_k - \bar{\theta}'_k$ has an accumulation point which is not 0, we may assume, since $\bar{\Theta}^n$ is compact, by taking a subsequence, that $\bar{\theta}_k \to \bar{\theta}$, $\bar{\theta}'_k \to \bar{\theta}'$ and $|\bar{\theta}$-$\bar{\theta}'| \neq 0$. By continuity

$$\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}_k) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}'_k) \to \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}')$$

and

$$\int_{R^n} \left(\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}')\right)^2 d\bar{z} > 0$$

as we have seen in the proof of Theorem 6. That will lead to a contradiction since

$$\frac{\int_{R^n} \left(\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}_k) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}'_k)\right)^2 d\bar{z}}{|\bar{\theta}_k - \bar{\theta}'_k|^2} \to 0 \tag{83}$$

Now we can write

$$\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}'_k) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}_k) = \frac{1}{2} \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}_k) D_{\bar{\theta}} \bar{L}(\bar{z}, \bar{\theta}_k)(\bar{\theta}'_k - \bar{\theta}_k)$$

$$+\frac{1}{2}\int_0^1\int_0^1 \lambda \bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k+\lambda\mu(\bar{\theta}_k^{'}-\bar{\theta}_k))$$
$$\Big[(\bar{\theta}_k^{'}-\bar{\theta}_k)^* D^2_{\bar{\theta}}\bar{L}\,(\bar{z},\bar{\theta}_k+\lambda\mu(\bar{\theta}_k^{'}-\bar{\theta}_k))(\bar{\theta}_k^{'}-\bar{\theta}_k)+$$
$$+\;\frac{1}{2}((\bar{\theta}_k^{'}-\bar{\theta}_k)^* D_{\bar{\theta}}\bar{L}\,(\bar{z},\bar{\theta}_k+\lambda\mu(\bar{\theta}_k^{'}-\bar{\theta}_k)))^2\Big]\,d\mu d\lambda$$

Recalling

$$D_{\bar{\theta}}\bar{L}\,(\bar{z},\bar{\theta})=\begin{pmatrix}\bar{z}_1-b'(\bar{\theta}_1)\\ \vdots \\ \bar{z}_n-b'(\bar{\theta}_n)\end{pmatrix}$$

$$\left(D^2_{\bar{\theta}}\bar{L}\,(\bar{z},\bar{\theta})\right)_{ll'}=-b"\left(\bar{\theta}_l\right)\delta_{ll'}$$

we get

$$\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k^{'})-\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k)-\frac{1}{2}\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k)\sum_{l=1}^{n}(\bar{z}_l-b'(\bar{\theta}_{kl}))(\bar{\theta}_{kl}^{'}-\bar{\theta}_{kl})=$$
$$+\frac{1}{2}\int_0^1\int_0^1 \lambda \bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k+\lambda\mu(\bar{\theta}_k^{'}-\bar{\theta}_k))$$
$$\Bigg[-\sum_{l=1}^{n} b"(\bar{\theta}_{kl}+\lambda\mu(\bar{\theta}_{kl}^{'}-\bar{\theta}_{kl}))(\bar{\theta}_{kl}^{'}-\bar{\theta}_{kl})^2+$$
$$+\frac{1}{2}\left(\sum_{l=1}^{n}(\bar{z}_l-b'(\bar{\theta}_{kl}))\left(\bar{\theta}_{kl}^{'}-\bar{\theta}_{kl}\right)\right)^2\Bigg]$$

and thus

$$\frac{\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k^{'})-\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k)-\frac{1}{2}\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k)\sum_{l=1}^{n}(\bar{z}_l-b'(\bar{\theta}_{kl}))(\bar{\theta}_{kl}^{'}-\bar{\theta}_{kl})}{|\bar{\theta}_k-\bar{\theta}_k^{'}|}\to 0,\forall\bar{z}.$$

We can also bound this function by a fixed function, which is square integrable. From Lebesgue's theorem we obtain easily

$$\frac{\int_{R^n}\left(\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k)-\bar{f}^{\frac{1}{2}}(\bar{z},\bar{\theta}_k^{'})\right)^2 d\bar{z}}{|\bar{\theta}_k-\bar{\theta}_k^{'}|^2}$$
$$-\frac{1}{4}\frac{\int_{R^n}\bar{f}(\bar{z},\bar{\theta}_k)\left(\sum_{l=1}^{n}(\bar{z}_l-b'(\bar{\theta}_{kl})\right)(\bar{\theta}_{kl}^{'}-\bar{\theta}_{kl}))^2 d\bar{z}}{|\bar{\theta}_k-\bar{\theta}_k^{'}|^2}\to 0,\text{ as }k\to\infty$$

However

$$\int_{R^n} \bar{f}(\bar{z}, \bar{\theta}_k) \left(\sum_{l=1}^{n} (\bar{z}_l - b'(\bar{\theta}_{kl}))(\bar{\theta}'_{kl} - \bar{\theta}_{kl}) \right)^2 d\bar{z}$$
$$= \sum_{l=1}^{n} \int_{R^n} \bar{f}(\bar{z}, \bar{\theta}_k)(\bar{z}_l - b'(\bar{\theta}_{kl}))^2 d\bar{z} (\bar{\theta}'_{kl} - \bar{\theta}_{kl}))^2$$

Since, by (5)

$$\int_R (z - b'(\theta))^2 f(z, \theta) dz = b"(\theta)$$

and

$$\int_{R^n} \bar{f}(\bar{z}, \bar{\theta}_k)(\bar{z}_l - b'(\bar{\theta}_{kl}))^2 d\bar{z} = \int_R (z - b'(\theta_{kl}))^2 f(z, \theta_{kl}) dz = b"(\theta_{kl}) \geq c > 0$$

we deduce

$$\int_{R^n} \bar{f}(\bar{z}, \bar{\theta}_k) \left(\sum_{l=1}^{n} (\bar{z}_l - b'(\bar{\theta}_{kl}))(\bar{\theta}'_{kl} - \bar{\theta}_{kl}) \right)^2 d\bar{z} \geq c |\bar{\theta}_k - \bar{\theta}'_k|^2$$

We obtain again a contradiction with (83). Therefore (82) is established. Now from the property (see the proof of Theorem 6) we have

$$|\bar{\theta}^i \left(x_0 + \frac{u}{M^\beta} \right) - \bar{\theta}^i (x_0)| \geq \frac{1}{c\rho} \frac{|u|}{M^\beta}$$

Combining this inequality with (82) yields

$$\int_{R^n} \left(\bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i \left(x_0 + \frac{u}{M^\beta} \right) \right) - \bar{f}^{\frac{1}{2}}(\bar{z}, \bar{\theta}^i(x_0))^2 d\bar{z} \geq \frac{\beta}{c^2 \rho^2} \frac{|u|^2}{M^{2\beta}}$$

Hence from (81) we obtain

$$E \bar{Z}_M^{\frac{1}{2}}(u) \leq \left(1 - \delta \frac{|u|^2}{M^{2\beta}} \right)^M$$

As in Theorem 6, we conclude that (80) holds for an appropriate constant $\alpha_n > 0$.

We next estimate

$$P\left(\{ |\hat{x}_M - x_0| M^\beta > \gamma \right) \leq P \left(\sup_{|u| > \gamma} Z_M(u) \geq 1 \right) \tag{84}$$

Let

$$\Gamma_r = \{ u |\ x_0 + \frac{u}{M^\beta} \in \mathcal{X},\ \gamma + r \leq |u| < \gamma + r + 1 \}$$

so

$$P(\sup_{|u|>\gamma} Z_M(u) \geq 1) \leq \sum_{r=0}^{+\infty} P(\sup_{u\in\Gamma_r} Z_M(u) \geq 1) \tag{85}$$

We next estimate $P(\sup_{u\in\Gamma_r} Z_M(u) \geq 1)$. We note that $\Gamma_r \subset [-(\gamma+r+1), \gamma+r+1]^n$. We subdivise this cube in cubes of diameter h. The number of such cubes is

$$N = B_n \frac{(\gamma+r+1)^n}{h^n}$$

where B_n is a generic constant depending only on n. We consider all the cubes of diameter h, which have a non empty interesection with Γ_r. The number of such cubes is $N' < N$. We obtain a covering of Γ_r with non overlapping cubes of diameter h. We call $\Gamma_r^j,\ j = 1, \cdots N'$ these cubes. Let u_r^j a point in Γ_r^j, which also belongs to Γ_r.

Let $\hat{u}_r$ be the point which maximizes $Z_M(u)$ over Γ_r. This point belongs to one and only one of the small cubes, say $\Gamma_r^{j_r}$. We can assert that

$$\begin{aligned}\{\sup_{u\in\Gamma_r} Z_M(u) \geq 1\} &= \left\{\sup_{u\in\Gamma_r} Z_M^{\frac{1}{m}}(u) \geq 1\right\} = \left\{Z_M^{\frac{1}{m}}(\hat{u}_r) \geq 1\right\} \\ &\subset \left\{Z_M^{\frac{1}{m}}\left(u_r^{j_r}\right) \geq \frac{1}{2}\right\} \cup \left\{|Z_M^{\frac{1}{m}}(\hat{u}_r) - Z_M^{\frac{1}{m}}\left(u_r^{j_r}\right)| \geq \frac{1}{2}\right\} \\ &\subset \left\{Z_M^{\frac{1}{m}}\left(u_r^{j_r}\right) \geq \frac{1}{2}\right\} \cup \left\{\sup_{\substack{|u-v|\leq h \\ u,v\in\Gamma_r}} |Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)| \geq \tfrac{1}{2}\right\} \\ &\subset \cup_{j=1}^{N'} \left\{Z_M^{\frac{1}{m}}\left(u_r^{j}\right) \geq \frac{1}{2}\right\} \cup \left\{\sup_{\substack{|u-v|\leq h \\ u,v\in\Gamma_r}} |Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)| \geq \tfrac{1}{2}\right\}\end{aligned}$$

Therefore

$$\begin{aligned}P\left(\left\{\sup_{u\in\Gamma_r} Z_M(u) \geq 1\right\} \leq \sum_{j=1}^{N'} P(\left\{Z_M^{\frac{1}{m}}(u_r^j) \geq \frac{1}{2}\right\}\right) \\ +P\left(\left\{\sup_{\substack{|u-v|\leq h \\ u,v\in\Gamma_r}} |Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)| \geq \tfrac{1}{2}\right\}\right)\end{aligned} \tag{86}$$

Now

$$\left\{Z_M^{\frac{1}{m}}(u_r^j) \geq \frac{1}{2}\right\} = \left\{Z_M^{\frac{1}{2}}(u_r^j) \geq (\frac{1}{2})^{\frac{m}{2}}\right\}$$

hence

$$P\left(\left\{Z_M^{\frac{1}{m}}(u_r^j) \geq \frac{1}{2}\right\}\right) \leq 2^{\frac{m}{2}}\, E Z_M^{\frac{1}{2}}(u_r^j)$$

and from (80)

$$\begin{aligned} E Z_M^{\frac{1}{2}}\left(u_r^j\right) &\leq \exp\left(-|u_r^j|^2 \alpha_n M^{1-2\beta}\right) \\ &\leq \exp\left(-(\gamma+r)^2 \alpha_n M^{1-2\beta}\right) \end{aligned}$$

Next, from (79) we have

$$\begin{aligned} P(\{\sup_{\substack{|u-v|\leq h \\ u,v\in\Gamma_r}} |Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)| \geq \frac{1}{2}\}) &\leq 2E \sup_{\substack{|u-v|\leq h \\ u,v\in\Gamma_r}} |Z_M^{\frac{1}{m}}(u) - Z_M^{\frac{1}{m}}(v)| \\ &\leq B_{n,m}|\gamma+r+1|^{\frac{n}{m}} h^{\frac{m-n}{m}} M^{\frac{1}{2}-\beta} \end{aligned}$$

Therefore, from (86) we obtain

$$\begin{aligned} P(\{\sup_{u\in\Gamma_r} Z_M(u) \geq 1\} &\leq N \exp\left(-(\gamma+r)^2 \alpha_n M^{1-2\beta}\right) \\ &\quad + B_{n,m}|\gamma+r+1|^{\frac{n}{m}} h^{\frac{m-n}{m}} M^{\frac{1}{2}-\beta} \\ &\leq B_n \left(\frac{\gamma+r+1}{h}\right)^n \exp\left(-(\gamma+r)^2 \alpha_n M^{1-2\beta}\right) \\ &\quad + B_{n,m}|\gamma+r+1|^{\frac{n}{m}} h^{\frac{m-n}{m}} M^{\frac{1}{2}-\beta} \end{aligned}$$

where $B_n, B_{n,m}$ are generic constants. So far h was not fixed. We choose h such that

$$\frac{\exp\left(-(\gamma+r)^2 \alpha_n M^{1-2\beta}\right)}{h^n} = h^{\frac{m-n}{m}} M^{\frac{1}{2}-\beta}$$

which means

$$h = \frac{\exp\left(-\frac{(\gamma+r)^2 \alpha_n M^{1-2\beta}}{1+n(1-\frac{1}{m})}\right)}{M^{\frac{\frac{1}{2}-\beta}{1+n(1-\frac{1}{m})}}}$$

With this choice we can state

$$P(\{\sup_{u\in\Gamma_r} Z_M(u) \geq 1\} \leq B_{n,m}(\gamma+r+1)^n M^{\frac{nm(\frac{1}{2}-\beta)}{nm+m-n}} \exp - \frac{(\gamma+r)^2 \alpha_n M^{1-2\beta}(m-n)}{m-n+mn}$$

By changing the constant α_n we have also

$$P(\{\sup_{u\in\Gamma_r} Z_M(u) \geq 1\} \leq B_{n,m} M^{\frac{nm(\frac{1}{2}-\beta)}{nm+m-n}} \exp - \frac{(\gamma + r)^2 \alpha_n M^{1-2\beta}(m-n)}{m-n+mn}$$

Hence, from (85) it follows easily

$$P(\{|\hat{x}_M - x_0|M^\beta > \gamma) \leq B_{n,m} M^{\frac{nm(\frac{1}{2}-\beta)}{nm+m-n}} \exp - \frac{\gamma^2 \alpha_n M^{1-2\beta}(m-n)}{m-n+mn}$$

and by changing α_n again we get

$$P(\{|\hat{x}_M - x_0|M^\beta > \gamma) \leq B_{n,m} \exp - \frac{\gamma^2 \alpha_n M^{1-2\beta}(m-n)}{m-n+mn} \tag{87}$$

As in Theorem 1, we deduce

$$|\hat{x}_M - x_0|M^\beta \to 0, \text{ a.s.as } M \to +\infty \tag{88}$$

To show that the variable tends to 0 in L^q, we can write

$$\begin{aligned}
E(|\hat{x}_M - x_0|M^\beta)^q &= \sum_{r=0}^{\infty} E[(|\hat{x}_M - x_0|M^\beta)^q 1_{r\leq|\hat{x}_M - x_0|M^\beta < r+1}] \\
&\leq \sum_{r=1}^{\infty} (r+1)^q P(\{|\hat{x}_M - x_0|M^\beta > r) \\
&+ E[(|\hat{x}_M - x_0|M^\beta)^q 1_{|\hat{x}_M - x_0|M^\beta < 1}] \\
&\leq B \sum_{r=1}^{\infty} (r+1)^q \exp - \frac{r^2 \alpha_n M^{1-2\beta}(m-n)}{m-n+mn} \\
&+ E[(|\hat{x}_M - x_0|M^\beta)^q 1_{|\hat{x}_M - x_0|M^\beta < 1}]
\end{aligned}$$

and both terms tend to 0, as $M \to +\infty$. Therefore

$$E(|\hat{x}_M - x_0|M^\beta)^q \to 0, \text{ as } M \to \infty$$

□

The proof has been completed. ■

4.5 Asymptotic Normality

Our objective now is to prove a result similar to that of Theorem 4, namely asymptotic normality. Unfortunately, the method does not work, since we cannot have Lemma 5, due to the fact that the Marcinkiewicz theorem used in this Lemma, necessitates that the variables are independent. We shall proceed differently, making use of Proposition 7.

Define the matrix

$$\Pi_M(x) = \sum_{j=1}^{M} W^j(x) h^j (h^j)^* \tag{89}$$

From the assumptions of Theorem 6, it is easy to check that, $W^j(x_0) \geq \alpha > 0$, hence for $M > n$

$$\sum_{j=1}^{M} W^j(x_0) h^j (h^j)^* \geq \alpha \sum_{j=1}^{M} h^j (h^j)^* \geq \alpha \sum_{j=1}^{\left[\frac{M}{n}\right]n} h^j (h^j)^*$$
$$= \alpha \sum_{i=1}^{\left[\frac{M}{n}\right]} (H^i)^* H^i \geq \alpha c \left[\frac{M}{n}\right] I \geq \alpha c \left(\frac{M}{n} - 1\right) I$$

therefore

$$\frac{\Pi_M(x_0)}{M} \geq \alpha c \left(\frac{1}{n} - \frac{1}{M}\right) I$$

hence

$$\frac{\Pi_M(x_0)}{M} \geq \alpha c \frac{1}{2n} I \;, \forall M \geq 2n \tag{90}$$

With these preliminaries our objective is to prove the following asymptotic normality result

Theorem 8 *We make the assumptions of Theorem 6, (71) and*

$$\varphi(\eta) \text{ is } C^2 \tag{91}$$

$$|W^j(x') - W^j(x)| \leq c|x' - x|^\beta 0 < \beta \leq 1, \forall x, x' \in \mathcal{X} \tag{92}$$
$$|\varphi"(x') - \varphi"(x)| \leq c|x' - x|^\beta 0 < \beta \leq 1, \forall x, x' \in \mathcal{X}$$

we then have

$$(\Pi_M(x_0))^{\frac{1}{2}} (\hat{x}_M - x_0) \to N(0, I) \tag{93}$$

the convergence being in law, in which $N(0, I)$ represents the Gaussian law in R^n, with mean 0 *and covariance matrix Identity.*

Proof Since $\hat{x}_M$ maximizes the likelihood in an open domain, we have

$$\sum_{j=1}^{M} D_x L \left(z^j, \theta^j(\hat{x}_M)\right) = 0$$

We can then write

$$\sum_{j=1}^{M} D_x L(z^j, \theta^j(x_0)) = \sum_{j=1}^{M} (D_x L(z^j, \theta^j(x_0)) - D_x L(z^j, \theta^j(\hat{x}_M)))$$

$$= -\sum_{j=1}^{M} \int_0^1 D_x^2 L(z^j, \theta^j(x_0 + \lambda(\hat{x}_M - x_0)))d\lambda(\hat{x}_M - x_0)$$

$$= -\sum_{j=1}^{M} D_x^2 L(z^j, \theta^j(x_0))(\hat{x}_M - x_0)$$

$$-[\sum_{j=1}^{M} \int_0^1 (D_x^2 L(z^j, \theta^j(x_0 + \lambda(\hat{x}_M - x_0)))$$

$$- D_x^2 L(z^j, \theta^j(x_0)))d\lambda](\hat{x}_M - x_0)$$

Recalling (54) we get

$$\Pi_M(x_0)(\hat{x}_M - x_0) = \sum_{j=1}^{M} D_x L(z^j, \theta^j(x_0)) +$$

$$+[\sum_{j=1}^{M} (D_x^2 L(z^j, \theta^j(x_0)) - E(D_x^2 L(z^j, \theta^j(x_0))))](\hat{x}_M - x_0)$$

$$+[\sum_{j=1}^{M} \int_0^1 (D_x^2 L(z^j, \theta^j(x_0 + \lambda(\hat{x}_M - x_0)))$$

$$- D_x^2 L(z^j, \theta^j(x_0)))d\lambda](\hat{x}_M - x_0)$$

We can then write, recalling the formulas for $D_x L(z^j, \theta^j(x_0))$ and $(D_x^2 L(z^j, \theta^j(x_0))$

$$\Pi_M^{\frac{1}{2}}(x_0)(\hat{x}_M - x_0) = \Pi_M^{-\frac{1}{2}}(x_0) \sum_{j=1}^{M} W^j(x_0) g'(\mu^j(x_0))(z^j - \mu^j(x_0))h^j +$$

$$+ \Pi_M^{-\frac{1}{2}}(x_0) \sum_{j=1}^{M} (z^j - \mu^j(x_0))\varphi''((h^j)^*(x_0))d\lambda h^j (h^j)^*(\hat{x}_M - x_0) \tag{94}$$

$$+\Pi_M^{-\frac{1}{2}}(x_0) \left(\sum_{j=1}^{M} \{z^j \int_0^1 (\varphi''((h^j)^*(x_0 + \lambda(\hat{x}_M - x_0))) - \varphi''((h^j)^* x_0))d\lambda \right.$$

$$- \int_0^1 (\mu^j(x_0 + \lambda(\hat{x}_M - x_0))\varphi''((h^j)^* \tag{95}$$

$$(x_0 + \lambda(\hat{x}_M - x_0))) - \mu^j(x_0)\varphi"((h^j)^* x_0))d\lambda$$
$$- \int_0^1 (W^j(x_0 + \lambda(\hat{x}_M - x_0)) - W^j(x_0))d\lambda\}h^j(h^j)^*\Bigg)$$
$$\Pi_M^{-\frac{1}{2}}(x_0)\Pi_M^{\frac{1}{2}}(x_0)(\hat{x}_M - x_0)$$

Define

$$\Gamma_M = \Pi_M^{-\frac{1}{2}}(x_0)\left(\sum_{j=1}^{M}\{z^j \int_0^1 (\varphi"((h^j)^*(x_0 + \lambda(\hat{x}_M - x_0))) - \varphi"((h^j)^* x_0))d\lambda + \right.$$
$$- \int_0^1 (\mu^j(x_0 + \lambda(\hat{x}_M - x_0))\varphi"((h^j)^*(x_0 + \lambda(\hat{x}_M - x_0))) - \mu^j(x_0)\varphi"((h^j)^* x_0))d\lambda$$
$$\left. - \int_0^1 (W^j(x_0 + \lambda(\hat{x}_M - x_0)) - W^j(x_0))d\lambda\}h^j(h^j)^*\right)\Pi_M^{-\frac{1}{2}}(x_0)$$

Noting that

$$D_x\mu^j(x) = \frac{1}{g'(\mu^j(x))}h^j$$

hence $\mu^j(x)$ is C^1. From the assumption (92) we can state

$$||\Gamma_M|| \leq C\left(\sum_{j=1}^{M}\frac{|z^j|}{M} + 1\right)|\hat{x}_M - x_0|^\beta$$

Set

$$X_{3M} = \Gamma_M \Pi_M^{\frac{1}{2}}(x_0)(\hat{x}_M - x_0)$$

which is the third term on the right hand side of (94). Using (90) we have

$$|X_{3M}| \leq C\left(\sum_{j=1}^{M}\frac{|z^j|}{M} + 1\right)M^{\frac{1}{2}}|\hat{x}_M - x_0|^{1+\beta}$$

The variable $\sum_{j=1}^{M}\frac{|z^j|}{M}$ is bounded in L^2. Let $1 < \delta < 2$. We have

$$E|X_{3M}|^\delta \leq C\left(\left[\left(E\left(\sum_{j=1}^{M}\frac{|z^j|}{M}\right)^2\right)^{\frac{\delta}{2}} + 1\right](E(M^{\frac{1}{2(1+\beta)}}|\hat{x}_M - x_0|)^{\frac{2\delta(1+\beta)}{2-\delta}})^{\frac{2-\delta}{2}}\right)$$

$$\leq C(E(M^{\frac{1}{2(1+\beta)}}|\hat{x}_M - x_0|)^{\frac{2\delta(1+\beta)}{2-\delta}})^{\frac{2-\delta}{2}}$$

But, from Proposition 7, we have

$$E(M^{\frac{1}{2(1+\beta)}}|\hat{x}_M - x_0|)^{\frac{2\delta(1+\beta)}{2-\delta}} \to 0$$

hence

$$X_{3M} \to 0, \text{ in } L^\delta \tag{96}$$

Consider next

$$X_{2M} = \Pi_M^{-\frac{1}{2}}(x_0)\sum_{j=1}^M (z^j - \mu^j(x_0))\varphi"((h^j)^*(x_0))d\lambda h^j(h^j)^*(\hat{x}_M - x_0)$$

We have

$$|X_{2M}| \leq ||\Pi_M^{-\frac{1}{2}}(x_0)\sum_{j=1}^M (z^j - \mu^j(x_0))\varphi"((h^j)^*(x_0))d\lambda h^j(h^j)^*|| \, |\hat{x}_M - x_0|$$

For $1 < \delta < 2$, we obtain

$$E|X_{2M}|^\delta \leq \left(E||\Pi_M^{-\frac{1}{2}}(x_0)\sum_{j=1}^M (z^j - \mu^j(x_0))\varphi"((h^j)^*(x_0))d\lambda h^j(h^j)^*||^2\right)^{\frac{\delta}{2}} (E|\hat{x}_M - x_0|^{\frac{2\delta}{2-\delta}})^{\frac{2-\delta}{2}}$$

We can take as a norm of a matrix A, $||A||^2 = \text{tr}\, A^*A$. Therefore

$$E||\Pi_M^{-\frac{1}{2}}(x_0)\sum_{j=1}^M (z^j - \mu^j(x_0))\varphi"((h^j)^*(x_0))d\lambda h^j(h^j)^*||^2 =$$

$$\sum_{j=1}^M V^j(x_0)(\varphi"((h^j)^*(x_0)))^2 \text{tr} h^j(h^j)^*\Pi_M^{-1}(x_0)h^j(h^j)^* \leq C$$

and thus also

$$X_{2M} \to 0, \text{ in } L^\delta \tag{97}$$

Setting finally

$$X_{1M} = \Pi_M^{-\frac{1}{2}}(x_0)\sum_{j=1}^M W^j(x_0)g'(\mu^j(x_0))(z^j - \mu^j(x_0))h^j$$

We want to prove that

$$X_{1M} \to N(0, I) \tag{98}$$

for the convergence in law.

What we must prove is

$$E \exp i\lambda^* X_{1M} \to \exp -\frac{1}{2}|\lambda|^2 \tag{99}$$

for any $\lambda \in R^n$, with $i = \sqrt{-1}$.

From the independence of variables z^j we have

$$\begin{aligned} E \exp i\lambda^* X_{1M} &= \prod_{j=1}^{M} E \exp i\lambda^* (\Pi^M(x_0))^{-\frac{1}{2}} h^j (z^j - \mu^j(x_0)) W^j(x_0) g'(\mu^j(x_0)) \\ &= \prod_{j=1}^{M} E \exp i\chi_j^M \end{aligned} \tag{100}$$

Write

$$E \exp i\chi_j^M = 1 - a_j^M$$

Since $E\chi_j^M = 0$, we can write

$$a_j^M = E\,[(\chi_j^M)^2 \int_0^1 \int_0^1 u \exp(iuv\chi_j^M) du dv]$$

Therefore

$$|a_j^M| \le \frac{1}{2} E\,(\chi_j^M)^2$$

and from (90) we get

$$|E\,(\chi_j^M)^2| \le \frac{C}{M}$$

hence also

$$\begin{aligned} |(a_j^M)^k| = |a_j^M|^k &\le \frac{1}{2^k} \left(E\,(\chi_j^M)^2\right)^k \\ &= \frac{1}{2^k} (\frac{C}{M})^k \end{aligned}$$

Since M is large, we can assume that $|a_j^M| < 1, \forall j$. We use the definition, for $|\eta| < 1$

$$\log(1 - \eta) = -\sum_{k=1}^{\infty} \frac{\eta^k}{k}$$

therefore

$$\log E \exp i \chi_j^M = -\sum_{k=1}^{\infty} \frac{(a_j^M)^k}{k}$$

Next, we have

$$\log E \exp i \lambda^* X_{1M} = \sum_{j=1}^{M} \log E \exp i \chi_j^M$$
$$= -\sum_{j=1}^{M} \sum_{k=1}^{\infty} \frac{\left(a_j^M\right)^k}{k}$$

We first consider

$$\left|\sum_{j=1}^{M} \sum_{k=2}^{\infty} \frac{(a_j^M)^k}{k}\right| \leq \sum_{j=1}^{M} \sum_{k=2}^{\infty} \frac{|(a_j^M)^k|}{k}$$
$$\leq \sum_{j=1}^{M} \sum_{k=2}^{\infty} \frac{1}{k2^k} \left(\frac{C}{M}\right)^k$$
$$\leq \sum_{k=2}^{\infty} \frac{1}{k2^k} \frac{C^k}{M^{k-1}}$$
$$= \frac{C^2}{4M} \sum_{k=0}^{\infty} \frac{1}{(k+2)2^k} \left(\frac{C}{M}\right)^k$$

This implies

$$\sum_{j=1}^{M} \sum_{k=2}^{\infty} \frac{(a_j^M)^k}{k} \to 0, \text{ as } M \to +\infty \tag{101}$$

Next

$$\sum_{j=1}^{M} a_j^M = \sum_{j=1}^{M} E\,(\chi_j^M)^2 \int_0^1 \int_0^1 u \exp(iuv\chi_j^M)dudv$$
$$= \frac{1}{2} \sum_{j=1}^{M} E\,(\chi_j^M)^2 + \gamma^M$$

with

$$\gamma^M = \sum_{j=1}^{M} E\,(\chi_j^M)^2 \int_0^1 \int_0^1 u\,(\exp(iuv\chi_j^M) - 1)dudv$$
$$= \sum_{j=1}^{M} E\,(\chi_j^M)^3 \int_0^1 \int_0^1 \int_0^1 iu^2v\,\exp(iuvw\chi_j^M)dudvdw$$

Therefore

$$|\gamma^M| \leq \frac{1}{6}\sum_{j=1}^{M} |E\,(\chi_j^M)^3|$$
$$= \frac{1}{6}\sum_{j=1}^{M} |(\lambda^*(\Pi^M(x_0))^{-\frac{1}{2}}h^j)^3(W^j(x_0)g'(\mu^j(x_0)))^3 E(z^j - \mu^j(x_0))^3|$$

and from (6) we have

$$|\gamma^M| \leq \frac{1}{6}\sum_{j=1}^{M} |(\lambda^*(\Pi^M(x_0))^{-\frac{1}{2}}h^j)^3(W^j(x_0)g'(\mu^j(x_0)))^3 b'''(\theta^j(x_0))|$$

and from the assumptions, we obtain

$$|\gamma^M| \leq \frac{C}{\sqrt{M}}$$

Finally

$$\frac{1}{2}\sum_{j=1}^{M} E\,(\chi_j^M)^2 = \frac{1}{2}\sum_{j=1}^{M} (\lambda^*(\Pi^M(x_0))^{-\frac{1}{2}}h^j)^2 V^j(x_0)(W^j(x_0)g'(\mu^j(x_0)))^2$$
$$= \frac{1}{2}\sum_{j=1}^{M} (\lambda^*(\Pi^M(x_0))^{-\frac{1}{2}}h^j)^2 W^j(x_0)$$

from which it follows immediately that

$$\frac{1}{2}\sum_{j=1}^{M} E\,(\chi_j^M)^2 = \frac{1}{2}|\lambda|^2 \tag{102}$$

Therefore

$$\sum_{j=1}^{M} a_j^M \to \frac{1}{2}|\lambda|^2$$

Collecting results we obtain

$$\log E \exp i\lambda^* X_{1M} \to -\frac{1}{2}|\lambda|^2$$

which is equivalent to (98).

Finally, going back to (94), we have

$$\Pi_M^{\frac{1}{2}}(x_0)(\hat{x}_M - x_0) = X_{1M} + X_{2M} + X_{3M}$$

and X_{1M} converges in law to $N(0, I)$ and $X_{2M} + X_{3M} \to 0$ in L^δ. This implies easily (93) and completes the proof.■

□

5 Vector Case

5.1 *Notation and Preliminaries*

In the preceding sections, we have considered that the observation z is a scalar, whereas the unknown parameter x is a vector in $R^{n\cdot}$ At the beginning, see (1) we had recalled the classical linear model, in which the observation is a vector in R^d. So for the sake of completeness, we return here to the vector case. We begin with a density $f(z, \theta)$ and the loglikelihood $L(z, \theta) = \log f(z, \theta)$, with

$$L(z, \theta) = z^* \Sigma^{-1}\theta - b(\theta) + c(z) \tag{103}$$

in which $z \in R^d, \theta \in R^d$ and Σ is a symmetric invertible $d \times d$ matrix. The function $b : R^d \to R$ will satisfy properties given below. The parameter θ is the canonical parameter. We note again z the random variable, whose probability density is $f(z, \theta)$, to save notation.

We have

$$Ez = \mu = \Sigma D_\theta b(\theta) \tag{104}$$

and

$$E(z - \mu)(z - \mu)^* = V = \Sigma D_\theta^2 b(\theta)\Sigma \tag{105}$$

We relate the unknown variable x to the mean, via the link function, so we write

$$g(\mu) = Hx + h \tag{106}$$

in which $H \in \mathcal{L}(R^n; R^d)$, and $g : R^d \to R^d$. So the canonical parameter is linked to the unknown variable x by the relation,

$$g(\Sigma D_\theta b(\theta)) = Hx + h \tag{107}$$

We shall assume the map $\theta \to g(\Sigma D_\theta b(\theta))$, invertible and thus define the function $\varphi(\eta)$ by solving

$$g(\Sigma D_\theta b(\varphi(\eta))) = \eta, \forall \eta \in R^d \tag{108}$$

So the canonical parameter is linked to the unknown variable, by the relation

$$\theta = \theta(x) = \varphi(Hx + h) \tag{109}$$

hence

$$D\theta(x) = D\varphi(Hx + h)H$$

We set, with abuse of notation,

$$L(z, x) = L(z, \theta(x)) \tag{110}$$

We compute easily

$$D_x L(z, x) = H^* D\varphi^*(Hx + h)(\Sigma^{-1} z - D_\theta b(\theta(x))) \tag{111}$$

In order to differentiate a second time in x, it is convenient to write the preceding relation as follows

$$D_x L(z, x) = H^* \sum_{k=1}^{d} ((\Sigma^{-1} z)_k - D_{\theta_k} b(\theta(x))) D\varphi_k(Hx + h) \tag{112}$$

which implies

$$\begin{aligned} D_x^2 L(z, x) = H^*[& - D\varphi^*(Hx + h) D^2 b(\theta(x)) D\varphi(Hx + h) + \\ & + \sum_{k=1}^{d} ((\Sigma^{-1} z)_k - D_{\theta_k} b(\theta(x))) D^2 \varphi_k(Hx + h)]H \end{aligned} \tag{113}$$

From the relation (108) we can write

$$I = Dg(\mu) \Sigma D^2 b(\varphi(\eta)) D\varphi(\eta)$$

hence

$$\Sigma^{-1} D\varphi(\eta) = (\Sigma D^2 b(\varphi(\eta))\Sigma)^{-1} (Dg(\mu))^{-1} \tag{114}$$

with

$$\mu = \Sigma D_\theta b(\varphi(\eta))$$

Using $\eta = Hx + h$, we get

$$\begin{aligned} \theta(x) &= \varphi(Hx + h) \\ \mu(x) &= \Sigma D_\theta b(\theta(x)) \\ V(x) &= \Sigma D^2 b(\theta(x)) \Sigma \end{aligned} \tag{115}$$

hence

$$\Sigma^{-1}D\varphi(Hx+h)=V(x)^{-1}(Dg(\mu(x)))^{-1} \tag{116}$$

therefore, using (111) we can write

$$D_xL(z,x)=H^*((Dg(\mu(x)))^*)^{-1}V(x)^{-1}(z-\mu(x)) \tag{117}$$

We introduce the weight matrix $W(x)$ defined by

$$W(x)=((Dg(\mu(x)))^*)^{-1}V(x)^{-1}(Dg(\mu(x)))^{-1} \tag{118}$$

So we can write

$$D_xL(z,x)=H^*W(x)Dg(\mu(x))(z-\mu(x)) \tag{119}$$

Using

$$ED_x^2L(z,x)=-ED_xL(z,x)(ED_xL(z,x))^*$$

we get

$$ED_x^2L(z,x)=-H^*W(x)Dg(\mu(x))V(x)(Dg(\mu(x)))^*W(x)H$$

and from (118) it follows

$$ED_x^2L(z,x)=-H^*W(x)H \tag{120}$$

We can summarize results as follows

Proposition 9 *We assume that $b(\theta)$ is C^2 on R^d and $D^2b(\theta)$ is strictly positive on compacts sets. We assume that the link function $g(\mu)$ is C^1 from R^d to R^d and invertible with bounded inverse on compact sets. The function $\theta\to g(\Sigma D_\theta b(\theta))$ is then invertible with inverse $\varphi(\eta)$ C^1 on compact sets. The canonical parameter, the mean and variance are expressed as functions of x by formulas (115). The weight matrix function is then defined by formula (118) and is continuous on compact sets. We can then express the gradient of the loglikelihood with respect to x, by formula (119) and we have formula (120).*

5.2 MLE Estimate

We consider now a sequence of independent random variables $z^1,\cdots,z^M$. Each of them follows a GLM distribution. We assume that the unknown variable x is the same for all random variables, however the matrix H and vector h vary from experiment to experiment, so we have a sequence H^j,h^j. Similarly, we have functions $b^j(\theta),\ g^j(\mu)$

and matrices Σ^j depending on the experiment, with identical properties. So we get functions $\theta^j(x), \mu^j(x), V^j(x), W^j(x)$ thus

$$D_x L^j(z,x) = (H^j)W^j(x)Dg^j(\mu^j(x))(z - \mu^j(x))$$

The MLE $\hat{x}_M$ if it exists is the solution of the following system of nonlinear equation*s

$$\sum_{j=1}^{M} (H^j)^* W^j(\hat{x}_M) Dg^j(\mu^j(\hat{x}_M))(z^j - \mu^j(\hat{x}_M)) = 0 \tag{121}$$

We propose to solve (121) by an iterative method, adapted from Newton's method. We shall define a sequence $\hat{x}^k_M$, written $\hat{x}^k$ to save notation as follows

$$\sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k)(H^j \hat{x}^{k+1} + h^j) = \sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k)(H^j \hat{x}^k + h^j) + \\ + \sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k) Dg^j(\mu^j(\hat{x}^k))(z^j - \mu^j(\hat{x}^k))$$

It is clear that, if this iteration converges, the limit point is a solution of (121). Noting that

$$H^j \hat{x}^k + h^j = g^j(\mu^j(\hat{x}^k))$$

We can rewrite the iteration as follows

$$\sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k)(H^j \hat{x}^{k+1} + h^j) = \sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k)[g^j(\mu^j(\hat{x}^k)) + \\ + Dg^j(\mu^j(\hat{x}^k))(z^j - \mu^j(\hat{x}^k))]$$

Assuming $\sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k) H^j$ to be invertible, we get the following iteration

$$\hat{x}^{k+1} = (\sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k) H^j)^{-1} \Bigg(\sum_{j=1}^{M} (H^j)^* W^j(\hat{x}^k)[-h^j + g^j(\mu^j(\hat{x}^k)) \\ + Dg^j(\mu^j(\hat{x}^k))(z^j - \mu^j(\hat{x}^k))] \Bigg) \tag{122}$$

5.3 The Gaussian Case

The Gaussian case corresponds to the model

$$z^j = H^j x + h^j + \epsilon^j \tag{123}$$

where the variables ϵ^j are independent, gaussian with mean 0 and covariance matrix Σ^j. Indeed, if we take in the model (103)

$$b^j(\theta) = \frac{1}{2}\theta^*(\Sigma^j)^{-1}\theta,\ c^j(z) = -\frac{1}{2}z^*(\Sigma^j)^{-1}z - \frac{d}{2}\log(2\pi|\Sigma^j|)|$$
$$g^j(\mu) = \mu$$

then we have

$$g^j(\Sigma^j D_\theta b^j(\theta)) = \theta$$

hence $\varphi(\eta) = \eta$. It follows that

$$\theta^j(x) = H^j x + h^j$$
$$\mu^j(x) = H^j x + h^j$$
$$V^j(x) = \Sigma^j$$
$$W^j(x) = (\Sigma^j)^{-1}$$

We clearly have

$$f^j(z,x) = \frac{\exp -\frac{1}{2}(z - (H^j x + h^j))^*(\Sigma^j)^{-1}(z - (H^j x + h^j))}{(2\pi|\Sigma^j|)^{\frac{d}{2}}}$$

which is equivalent to (123).

The system (121) becomes

$$\sum_{j=1}^{M}(H^j)^*(\Sigma^j)^{-1}(z^j - H^j\hat{x}_M - h^j) = 0 \tag{124}$$

therefore $\hat{x}_M$ is given explicitly by the formula

$$\hat{x}_M = (\sum_{j=1}^{M}(H^j)^*(\Sigma^j)^{-1}H^j)^{-1}\sum_{j=1}^{M}(H^j)^*(\Sigma^j)^{-1}(z^j - h^j) \tag{125}$$

We can check immediately that the iteration (122) leads to $\hat{x}^k = \hat{x}_M, \forall k$.

5.4 Recursivity

It is well known that formula (125), although non recursive in M can be given a recursive form. Similarly, the iterative algorithm (122) is not recursive in M. To get a recursive formula we define a sequence $\hat{x}_j$ as follows. Given $\hat{x}_j$, define $\hat{\mu}_j$ by

$$g(\hat{\mu}_j) = H^j \hat{x}_j + h^j \tag{126}$$

then $\hat{\theta}_j$ by

$$(\Sigma^j)^{-1} \hat{\mu}_j = Db(\hat{\theta}_j) \tag{127}$$

next $\hat{V}_j$ by

$$\hat{V}_j = \Sigma^j D^2 b(\hat{\theta}_j) \Sigma^j \tag{128}$$

Finally $\hat{W}_j$ is given by

$$\hat{W}_j = ((Dg(\hat{\mu}_j))^*)^{-1} \hat{V}_j^{-1} (Dg(\hat{\mu}_j))^{-1} \tag{129}$$

We write

$$\hat{x}_{M+1} = \left(\sum_{j=1}^{M} (H^j)^* \hat{W}_j H^j \right)^{-1} \left(\sum_{j=1}^{M} (H^j)^* \hat{W}_j [- h^j + g^j(\hat{\mu}_j) + Dg^j(\hat{\mu}_j)(z^j - \hat{\mu}_j)] \right) \tag{130}$$

We can then give a recursive form to this formula. We set

$$P_M = \left(\sum_{j=1}^{M} (H^j)^* \hat{W}_j H^j \right)^{-1} \tag{131}$$

then

$$P_M^{-1} = P_{M-1}^{-1} + (H^M)^* \hat{W}_M H^M$$

from which we obtain easily the following recursive relation

$$P_M = P_{M-1} - P_{M-1}(H^M)^* \left(H^M P_{M-1} (H^M)^* + \hat{W}_M^{-1} \right)^{-1} H^M P_{M-1} \tag{132}$$

Let us introduce the *corrected* observation

$$\hat{\zeta}_M = g^M(\hat{\mu}_M) + Dg^M(\hat{\mu}_M)(z^M - \hat{\mu}_M) \tag{133}$$

then, from (130) we get

$$
\begin{aligned}
P_M^{-1}\hat{x}_{M+1} &= \sum_{j=1}^{M}(H^j)^*\hat{W}_j(-h^j+\hat{\zeta}_j)\\
&= P_{M-1}^{-1}\hat{x}_M + (H^M)^*\hat{W}_M(-h^M+\hat{\zeta}_M)
\end{aligned}
$$

hence

$$\hat{x}_{M+1} = P_M P_{M-1}^{-1}\hat{x}_M + P_M(H^M)^*\hat{W}_M(-h^M+\hat{\zeta}_M)$$

Using

$$P_M P_{M-1}^{-1} = I - P_M(H^M)^*\hat{W}_M H^M$$

we obtain

$$\hat{x}_{M+1} = \hat{x}_M + P_M(H^M)^*\hat{W}_M(\hat{\zeta}_M - H^M\hat{x}_M - h^M) \tag{134}$$

or, equivalently

$$\hat{x}_{M+1} = \hat{x}_M + P_M(H^M)^*\hat{W}_M Dg(\hat{\mu}_M)(z^M - \hat{\mu}_M) \tag{135}$$

In such a recursive algorithm, the initial condition $\hat{x}_1$ is arbitrary and corresponds to the best prior estimate of x, without any observation. We then define the values of $\hat{\mu}_1, \hat{\theta}_1, \hat{V}_1, \hat{W}_1$ by formulas (126), (127), (128), (129) and

$$P_1 = ((H^1)^*\hat{W}_1 H^1)^{-1} \tag{136}$$

5.5 *Examples*

5.5.1 Binomial Distribution

The observation takes only finite values $0, 1, \cdots, q$. We take $d = 1$, $\Sigma = 1$,

$$b(\theta) = q\log(1+\exp\theta),\ c(z) = \log C_q^z$$

$$g(\mu) = \log\frac{\mu}{q-\mu},\ 0 < \mu < q$$

so noting

$$\pi = \frac{\exp\theta}{1+\exp\theta}$$

we get easily

$$f(z,\theta) = C_q^z \pi^z (1-\pi)^{q-z}$$

We take

$$Hx + h = h^* x$$

with the abuse of notation as regards h. So

$$\mu(x) = q \frac{\exp h^* x}{1 + \exp h^* x}, \; \theta(x) = h^* x$$

$$V(x) = q \frac{\exp h^* x}{(1 + \exp h^* x)^2}, \; W(x) = V(x)$$

so $W(x)g'(\mu(x)) = 1$. Therefore the maximum likelihood estimator is the solution of the system on nonlinear equation*s, see (121)

$$\sum_{j=1}^{M} h^j \left(z^j - q \frac{\exp h^* \hat{x}_M}{1 + \exp h^* \hat{x}_M} \right) = 0 \tag{137}$$

The recursive algorithm (135) reduces to

$$\hat{x}_{M+1} = \hat{x}_M + P_M h^M (z^M - \hat{\mu}_M) \tag{138}$$

with

$$P_M = P_{M-1} - \frac{P_{M-1} h^M (h^M)^* P_{M-1}}{(h^M)^* P_{M-1} h^M + \hat{W}_M^{-1}} \tag{139}$$

If we want to solve the system (137) by the iterative method (122) we get the sequence

$$\hat{x}^{k+1} = \hat{x}^k + \left(\sum_{j=1}^{M} W^j(\hat{x}^k) h^j (h^j)^* \right)^{-1} \sum_{j=1}^{M} (z^j - \mu^j(\hat{x}^k)) h^j \tag{140}$$

5.5.2 Poisson Distribution

We consider

$$d = 1, \; b(\theta) = \exp\theta, \; c(z) = -\log z!, \; \Sigma = 1, \; g(\mu) = \log\mu \tag{141}$$

$$Hx + h = h^* x$$

and z is integer, so

$$f(z,\theta) = \frac{\theta^z}{z!} \exp -\theta$$

Therefore

$$\mu(x) = \exp\, h^* x,\ \varphi(\eta) = \eta,\ \theta(x) = h^* x$$
$$V(x) = \exp\, h^* x,\ \ W(x) = \exp\, h^* x$$

then the system (121) becomes

$$\sum_{j=1}^{M} (z^j - \exp\,((h^j)^* \hat{x}_M))h^j = 0 \tag{142}$$

The algorithm (122) becomes

$$\hat{x}^{k+1} = \hat{x}^k + \left(\sum_{j=1}^{M} \mu^j(\hat{x}^k) h^j (h^j)^* \right)^{-1} \sum_{j=1}^{M} (z^j - \mu^j(\hat{x}^k))h^j \tag{143}$$

and the recursive algorithm (135) yields

$$\hat{x}_{M+1} = \hat{x}_M + P_M h^M (z^M - \hat{\mu}_M) \tag{144}$$

with

$$P_M = P_{M-1} - \frac{P_{M-1} h^M (h^M)^* P_{M-1}}{(h^M)^* P_{M-1} h^M + \hat{\mu}_M^{-1}} \tag{145}$$

5.5.3 Gamma Distribution

This is a little bit more complex. We take again $d = 1$, and (ν is a positive constant)

$$\Sigma = \frac{1}{\nu},\ b(\theta) = -\nu \log(-\theta),\ \theta \in R^-$$
$$g(\mu) = \frac{1}{\mu},\ \mu > 0,\ c(z) = (\nu - 1)\log z - \log \Gamma(\nu) + \nu \log \nu,\ z > 0$$

We have

$$f(z,\theta) = \frac{(-\theta\nu)^\nu}{\Gamma(\nu)} z^{\nu-1} \exp \nu\theta z$$

recalling that $\theta < 0$. Then $Hx + h = h^* x$. We can easily check that

$$\varphi(\eta) = -\eta,\ \theta(x) = -h^* x$$

We note that x must satisfy the constraint

$$h^* x \geq 0 \tag{146}$$

then

$$\mu(x) = \frac{1}{h^* x}, \; V(x) = \frac{1}{\nu (h^* x)^2}, \; W(x) = \frac{\nu}{(h^* x)^2} \tag{147}$$

So we can write

$$D_x L^j(z, x) = -h^j \nu (z - \frac{1}{(h^j)^* x}) \tag{148}$$

Because of the constraint (146), we cannot write directly (121). The necessary condition of optimality of $\hat{x}_M$ can be written as follows (Kuhn-Tucker condition)

$$\sum_{j=1}^{M} (D_x L^j(z, \hat{x}_M))^* x \leq 0, \; \forall x, \text{such that}, \; (h^j)^* x \geq 0, \; \forall j \tag{149}$$

$$\sum_{j=1}^{M} (D_x L^j(z, \hat{x}_M))^* \hat{x}_M = 0, \; (h^j)^* \hat{x}_M \geq 0, \; \forall j \tag{150}$$

and using formulas (147) we obtain

$$\sum_{j=1}^{M} (h^j)^* x (z^j - \frac{1}{(h^j)^* \hat{x}_M}) \geq 0, \; \forall x, \text{such that}, \; (h^j)^* x \geq 0, \; \forall j \tag{151}$$

$$\sum_{j=1}^{M} (h^j)^* \hat{x}_M (z^j - \frac{1}{(h^j)^* \hat{x}_M}) = 0, \; (h^j)^* \hat{x}_M \geq 0, \; \forall j \tag{152}$$

6 Dynamic Models

6.1 *General Bayesian Approach*

6.1.1 Preliminaries

In all the preceding sections, we have been considering a fixed parameter x.The problem is thus an estimation problem, and the maximum likelihood is an appropriate method to achieve this estimation. When the parameter itself evolves with time, it can be considered as the state of a dynamic system. In general there is an evolution law for this state, but uncertainties affect this evolution. The problem is to estimate the current state, and thus it is a tracking problem. An adequate approach is the

Bayesian approach. We are going to describe it in general, then to apply it for dynamic generalized models.

Instead of a parameter, we shall speak of the state of the system at time j, denoted by $x^j \in R^n$. The observation is still denoted $z^j \in R^d$. The pair x^j, z^j evolves as a Markov chain, with the particularity that the transition probability depends only on x. In other words, we consider a sequence of functions $\varpi^j(\eta, \zeta, x)$ where $\eta, x \in R^n$ and $\zeta \in R^d$. Defining the $\sigma-$algebra generated by $z^1, \cdots, z^j$

$$\mathcal{Z}^j = \sigma(z^1, \cdots, z^j)$$

and, similarly

$$\mathcal{F}^j = \sigma(x^0, x^1, z^1, \cdots, x^j, z^j)$$

we have, for a continuous bounded function on $R^n \times R^d$ $\varphi(x, z)$

$$E(\varphi(x^{j+1}, z^{j+1})|\mathcal{F}^j) = \int\int \varphi(\eta, z)\varpi^j(\eta, z, x^j)d\eta dz \tag{153}$$

To complete the description of the evolution, we need an initial probability density for x^0, denoted by $\mu(\eta)$.

6.1.2 Recursion Formulas

We begin by considering the joint probability density of the variables $z^1, \cdots, z^M, x^M$ given by

$$\pi^M(\zeta^1, \cdots, \zeta^M, \eta^M) = \int \cdots \int \mu(\eta^0)\varpi^0(\eta^1, \zeta^1, \eta^0) \cdots \varpi^{M-1}(\eta^M, \zeta^M, \eta^{M-1})d\eta^0 \cdots d\eta^{M-1} \tag{154}$$

and we see immediately that it satisfies a recursion equation

$$\pi^{M+1}(\zeta^1, \cdots, \zeta^{M+1}, \eta^{M+1}) = \int \pi^M(\zeta^1, \cdots, \zeta^M, \eta)\varpi^M(\eta^{M+1}, \zeta^{M+1}, \eta)d\eta \tag{155}$$

We then derive the conditional probability density of x^M, given the $\sigma-$algebra $\mathcal{Z}^M$. For a given bounded continuous function φon R^n, we consider $E[\varphi(x^M)|\mathcal{Z}^M]$. It a random variable $\mathcal{Z}^M$- measurable. It is standard to check that it obtained through a conditional probability density by the formula

$$E[\varphi(x^M)|\mathcal{Z}^M] = \int p^M(z^1, \cdots, z^M, \eta)\varphi(\eta)d\eta \tag{156}$$

and the function $p^M(\zeta^1, \cdots, \zeta^M, \eta)$ is given by the formula

$$p^M(\zeta^1, \cdots, \zeta^M, \eta) = \frac{\pi^M(\zeta^1, \cdots, \zeta^M, \eta)}{\int \pi^M(\zeta^1, \cdots, \zeta^M, \eta')d\eta'} \tag{157}$$

From the recursion (155) we obtain easily a recursion for the function p^M. We have

$$p^{M+1}(\zeta^1,\cdots,\zeta^{M+1},\eta)=\frac{\int p^M(\zeta^1,\cdots,\zeta^M,\eta')\varpi^M(\eta,\zeta^{M+1},\eta')d\eta'}{\int\int p^M(\zeta^1,\cdots,\zeta^M,\eta')\varpi^M(\eta",\zeta^{M+1},\eta')d\eta' d\eta"} \tag{158}$$

We start this recursion with

$$p^0(\eta)=\mu(\eta)$$

6.2 Dynamic GLM

6.2.1 Conditional Probability

At time j, the observation z^{j+1} has a conditional probability density, when $x^j = x$, given by

$$f^j(z,x)=\exp\left(\theta^j(x)^*(R^j)^{-1}z-b^j(\theta^j(x))+c^j(z)\right) \tag{159}$$

The dynamic system x^j evolves according to the model

$$x^{j+1}=F^jx^j+f^j+G^jw^j \tag{160}$$
$$x^0=N(\bar{\xi},P_0)$$

in which the w^j are independent random variables, which are normal with mean 0 and covariance matrix Q^j. The variables w^j take values in R^m. They are independent of x^0. Also for given x^j, the variables x^{j+1} and z^{j+1} are independent. Therefore the pair x^j, z^j is a Markov chain, as described in Sect. 6.1.1. The function $\varpi^j(\eta,\zeta,x)$, defined in (153) is given by

$$\varpi^j(\eta,\zeta,x)$$
$$=\frac{\exp\left[-\frac{1}{2}(\eta-F^jx-f^j)^*(G^jQ^j(G^j)^*)^{-1}(\eta-F^jx-f^j)+\theta^j(x)^*(R^j)^{-1}\zeta-b^j(\theta^j(x))+c^j(\zeta)\right]}{(2\pi)^{\frac{n}{2}}|G^jQ^j(G^j)^*|^{\frac{1}{2}}} \tag{161}$$

Consider the conditional probability of x^M, given the filtration $\mathcal{Z}^M$, denoted by $p^M(\eta)$, in which we omit the dependence with respect to the arguments $\zeta^1,\cdots\zeta^M$. To simplify notation, we define

$$\gamma^j(\eta,x)=\frac{\exp\left[-\frac{1}{2}(\eta-F^jx-f^j)^*(G^jQ^j(G^j)^*)^{-1}(\eta-F^jx-f^j)\right]}{(2\pi)^{\frac{n}{2}}|G^jQ^j(G^j)^*|^{\frac{1}{2}}}$$

and

$$g^j(\zeta,x)=\exp\left(\theta^j(x)^*(R^j)^{-1}\zeta-b^j(\theta^j(x))\right)$$

then formula (158) leads to

$$p^{M+1}(\eta) = \frac{\int p^M(\eta')\gamma^M(\eta,\eta')g^M(\zeta^{M+1},\eta')d\eta'}{\int p^M(\eta')g^M(\zeta^{M+1},\eta')d\eta'} \tag{162}$$

and obtain the following

Proposition 10 *For the model (159), (160), the conditional probability density of x^M, given the $\sigma-$algebra $\mathcal{Z}^M$, denoted $p^M(\eta) = p^M(\zeta^1, \cdots, \zeta^M, \eta)$ is defined recursively by formula (162) with $p^0(\eta) = \mu(\eta) = N(\bar{\xi}, P_0)$.*

6.2.2 First Two Moments

The best estimate of x^M, denoted $\hat{x}_M$ is defined from the conditional probability density $p^M(\eta)$, simply by the formula

$$\hat{x}_M = \int \eta p^M(\eta)d\eta \tag{163}$$

Unfortunately, there is no recursive formula for $\hat{x}_M$. Noting that

$$\int \eta\gamma^M(\eta,\eta') = F^M\eta' + f^M$$

we can write

$$\hat{x}_{M+1} = F^M\hat{y}_M + f^M \tag{164}$$

with

$$\hat{y}_M = \frac{\int \eta\, p^M(\eta)g^M(\zeta^{M+1},\eta)d\eta}{\int p^M(\eta)g^M(\zeta^{M+1},\eta)d\eta} \tag{165}$$

It is possible to get a recursive formula for $\hat{x}_M$, considering apprimations. The idea is to introduce the covariance operator

$$P_M = \int \eta\eta^* p^M(\eta)d\eta - \hat{x}_M(\hat{x}_M)^* \tag{166}$$

and to approximate $p^M(\eta)$ by a gaussian

$$p^M(\eta) = \frac{\exp\left[-\frac{1}{2}(\eta-\hat{x}_M)^*P_M^{-1}(\eta-\hat{x}_M)\right]}{(2\pi)^{\frac{n}{2}}|P_M|^{\frac{1}{2}}}$$

therefore

$$\hat{y}_M = \frac{\int \eta\, g^M(\zeta^{M+1},\eta)\exp\left[-\frac{1}{2}\eta^*P_M^{-1}\eta + (\hat{x}_M)^*P_M^{-1}\eta\right]d\eta}{\int g^M(\zeta^{M+1},\eta)\exp\left[-\frac{1}{2}\eta^*P_M^{-1}\eta + (\hat{x}_M)^*P_M^{-1}\eta\right]d\eta} \tag{167}$$

Recalling (164) we see that $\hat{x}_{M+1}$ can be obtained from the knowledge of $\hat{x}_M$ and P_M.

We next have to define P_{M+1}. We use

$$P_{M+1} = \int \eta\eta^* p^{M+1}(\eta)d\eta - \hat{x}_{M+1}(\hat{x}_{M+1})^*$$

with $p^{M+1}(\eta)$ given by (162) and $\hat{x}_{M+1}$ given by (164). Introduce

$$\Gamma_M = \frac{\int \eta\eta^* g^M(\zeta^{M+1},\eta)\exp[-\frac{1}{2}\eta^* P_M^{-1}\eta + (\hat{x}_M)^* P_M^{-1}\eta]d\eta}{\int g^M(\zeta^{M+1},\eta)\exp[-\frac{1}{2}\eta^* P_M^{-1}\eta + (\hat{x}_M)^* P_M^{-1}\eta]d\eta} - \hat{y}_M\hat{y}_M^* \tag{168}$$

$$= \frac{\int (\eta-\hat{y}_M)(\eta-\hat{y}_M)^* g^M(\zeta^{M+1},\eta)\exp[-\frac{1}{2}\eta^* P_M^{-1}\eta + (\hat{x}_M)^* P_M^{-1}\eta]d\eta}{\int g^M(\zeta^{M+1},\eta)\exp[-\frac{1}{2}\eta^* P_M^{-1}\eta + (\hat{x}_M)^* P_M^{-1}\eta]d\eta}$$

We check easily the formula

$$P_{M+1} = F^M\Gamma_M(F^M)^* + G^M Q^M (G^M)^* \tag{169}$$

So we propose the recursive algorithm for $\hat{x}_M$ and P_M, defined by formulas (164) and (169), in which the random quantities $\hat{y}_M$ and Γ_M are given by formulas (167) and (168).

6.3 Applications

6.3.1 Kalman Filter

We consider the situation

$$\theta^j(x) = H^j x + h^j,\ b^j(\theta) = \frac{1}{2}\theta^*(R^j)^{-1}\theta$$

so

$$g^j(\zeta,x) = \exp((H^j x + h^j)^*(R^j)^{-1}\zeta - \frac{1}{2}(H^j x + h^j)^*(R^j)^{-1}(Hx + h^j))$$

We will check that $p^M(\eta)$ is indeed a gaussian. Assuming it is the case for $p^M(\eta)$, we prove it for $p^{M+1}(\eta)$ by computing the characteristic function

$$\mathcal{L}^{M+1}(\lambda) = \int \exp i\lambda^*\eta\, p^{M+1}(\eta)d\eta$$

Since

$$\int \exp i\lambda^*\eta\,\gamma^M(\eta,\eta')d\eta = \exp[i\lambda^*(F^M\eta' + f^M) - \frac{1}{2}\lambda^* G^M Q^M (G^M)^*\lambda]$$

we get

$$\mathcal{L}^{M+1}(\lambda) = \exp\left[i\lambda^* f^M - \frac{1}{2}\lambda^* G^M Q^M (G^M)^* \lambda\right] \frac{N(\lambda)}{N(0)}$$

with

$$N(\lambda) = \int \exp\left[-\frac{1}{2}(\eta^*(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)\eta) + (i(F^M)^*\lambda + P_M^{-1}\hat{x}_M + (H^M)^*(R^M)^{-1}(\zeta^{M+1} - h^M))^*\eta\right] d\eta$$

We then check easily that

$$\frac{N(\lambda)}{N(0)} = \exp\left[-\frac{1}{2}\lambda^*(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}\lambda\right] \times \exp i\lambda^* \left(F^M(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(P_M^{-1}\hat{x}_M + (H^M)^*(R^M)^{-1}(\zeta^{M+1} - h^M))\right)$$

Collecting results we see that $\mathcal{L}^{M+1}(\lambda)$ is the exponential of a quadratic form in λ. Therefore $p^{M+1}(\eta)$ is a gaussian with mean

$$\hat{x}_{M+1} = f^M + F^M(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(P_M^{-1}\hat{x}_M + (H^M)^*(R^M)^{-1}(\zeta^{M+1} - h^M))$$

and covariance matrix

$$P_{M+1} = G^M Q^M (G^M)^* + F^M(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(F^M)^*$$

We can rewrite these expressions as follows

$$\hat{x}_{M+1} = F^M\hat{x}_M + f^M + F^M(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(H^M)^*(R^M)^{-1}(\zeta^{M+1} - H^M\hat{x}_M - h^M) \tag{170}$$

Now, we check easily

$$(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1} = P^M - P^M(H^M)^*(R^M + H^M P^M (H^M)^*)^{-1}H^M P^M$$

So

$$P_{M+1} = F^M P^M (F^M)^* - F^M P^M (H^M)^*(R^M + H^M P^M (H^M)^*)^{-1} H^M P^M (F^M)^* + G^M Q^M (G^M)^* \tag{171}$$

We use also

$$(P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(H^M)^*(R^M)^{-1} = P^M(H^M)^*(R^M + H^M P^M (H^M)^*)^{-1}$$

to write

$$\hat{x}_{M+1} = F^M\hat{x}_M + f^M + F^M P^M (H^M)^*(R^M + H^M P^M (H^M)^*)^{-1}(\zeta^{M+1} - H^M\hat{x}_M - h^M) \tag{172}$$

and we obtain the Kalman filter.

Considering formulas (167) and (168) we check easily that

$$\hat{y}_M = (P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(P_M^{-1}\hat{x}_M + (H^M)^*(R^M)^{-1}(\zeta^{M+1} - h^M)) \tag{173}$$

$$\Gamma_M = (P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1} \tag{174}$$

and thus these formulas are no more approximations.

6.3.2 Poisson Distribution

Consider the situation of Sect. 5.5.2, then

$$\theta^j(x) = (h^j)^* x,\ b^j(\theta) = \exp\theta$$
$$g^j(\zeta, x) = \exp(\zeta(h^j)^* x - \exp(h^j)^* x)$$

then formulas (167) and (168) yield

$$\hat{y}_M = \frac{\int \eta \exp[-\frac{1}{2}\eta^* P_M^{-1}\eta - \exp(h^j)^*\eta + ((\hat{x}_M)^* P_M^{-1} + \zeta^{M+1}(h^j)^*)\eta] d\eta}{\int \exp[-\frac{1}{2}\eta^* P_M^{-1}\eta - \exp(h^j)^*\eta + ((\hat{x}_M)^* P_M^{-1} + \zeta^{M+1}(h^j)^*)\eta] d\eta} \tag{175}$$

$$\Gamma_M = \frac{\int (\eta - \hat{y}_M)(\eta - \hat{y}_M)^* \exp[-\frac{1}{2}\eta^* P_M^{-1}\eta - \exp(h^j)^*\eta + ((\hat{x}_M)^* P_M^{-1} + \zeta^{M+1}(h^j)^*)\eta] d\eta}{\int \exp[-\frac{1}{2}\eta^* P_M^{-1}\eta - \exp(h^j)^*\eta + ((\hat{x}_M)^* P_M^{-1} + \zeta^{M+1}(h^j)^*)\eta] d\eta} \tag{176}$$

and $\hat{x}_{M+1}$, P_{M+1} are given by (164) and (169).

6.3.3 Kalman Filter Revisited

We consider formulas (173), (174). Note that the Kalman filter $\hat{x}_{M+1}$ and the covariance error P_{M+1} are given by formulas (164) and (169). We rewrite them as

$$\hat{y}_M = \hat{x}_M + (P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1}(H^M)^*(R^M)^{-1}(z^{M+1} - (H^M\hat{x}_M + h^M)) \tag{177}$$

$$\Gamma_M = (P_M^{-1} + (H^M)^*(R^M)^{-1}H^M)^{-1} \tag{178}$$

We have reinstated z^{M+1} in formula (177) in lieu of ζ^{M+1} to consider $\hat{y}_M$ as a random variable $\mathcal{Z}^{M+1}$measurable. In fact, in this section, we will use the same notation for random variables and arguments, to save notation. We want to prove the following result

Proposition 11
We have

$$\hat{y}_M = E[x_M|\mathcal{Z}^{M+1}] \tag{179}$$
$$\Gamma_M = E[(x_M - \hat{y}_M)(x_M - \hat{y}_M)^*|\mathcal{Z}^{M+1}] \tag{180}$$

Note that this interpretation fits perfectly with $\hat{x}_{M+1} = E[x^{M+1}|\mathcal{Z}^{M+1}]$ and

$$x^{M+1} = F^M x^M + f^M + G^M w^M$$

and noting that w^M is independent of $\mathcal{Z}^{M+1}$.

Proof In proving Proposition 11, we shall use an approach inspired from West et al. 1985. This approach focuses on the canonical parameter, and the GLM form of the observation probability density. So we introduce the sequence of random variables

$$\theta^M = H^M x^M + h^M$$

The probability density of z^{M+1} given θ^M is a Gaussian with mean θ^M and covariance matrix R^M. We write this as

$$\text{Prob}(z^{M+1}|\theta^M) = N(\theta^M|R^M)$$

We recall that we use the same notation for random variables and arguments representing their values. We write this probability in GLM format as follows

$$\text{Prob}(z^{M+1}|\theta^M) = B(z^{M+1}, (R^M)^{-1}) \exp[(z^{M+1})^*(R^M)^{-1}\theta^M - b(\theta^M, (R^M)^{-1})] \tag{181}$$

with, of course,

$$B(z^{M+1}, (R^M)^{-1}) = \frac{\exp -\frac{1}{2}(z^{M+1})^*(R^M)^{-1}z^{M+1}}{(2\pi)^{\frac{d}{2}}|R^M|^{\frac{1}{2}}}$$

and the function $b(\theta, \Sigma)$ defined on $R^d \times \mathcal{L}(R^d, R^d)$ is given by

$$b(\theta, \Sigma) = \frac{1}{2}\theta^* \Sigma \theta$$

We recognize the function b depending on the canonical argument. We have introduced an additional dependence, with respect to a symmetric matrix. This dependence is linear. This will play a key role in the following. Since we operate a recursive argument, we know that

$$\text{Prob}(\theta^M|\mathcal{Z}^M) = N(H^M\hat{x}_M + h^M, H^M P_M (H^M)^*)$$

The key idea is to write it as follows

$$\text{Prob}(\theta^M|\mathcal{Z}^M) = c(\alpha_M, \beta_M) \exp[\alpha_M^*\theta^M - b(\theta^M, \beta_M)] \tag{182}$$

where α_M is a vector in R^d and $\beta_M \in \mathcal{L}(R^d, R^d)$. They are given explicitly by

$$\alpha_M = (H^M P_M (H^M)^*)^{-1}(H^M\hat{x}_M + h^M)$$
$$\beta_M = (H^M P_M (H^M)^*)^{-1}$$

and

$$c(\alpha, \beta) = \frac{|\beta|^{\frac{1}{2}} \exp -\frac{1}{2}\alpha^*\beta^{-1}\alpha}{(2\pi)^{\frac{d}{2}}}$$

We note the analogy between formulas (181) and (182), in term of using the function b, but with different arguments.

From the probabilities (181) and (182) we deduce, using the linearity of b with respect to the second argument

$$\begin{aligned}\text{Prob}(z^{M+1}|\mathcal{Z}^M) &= c(\alpha_M, \beta_M)B(z^{M+1}, (R^M)^{-1}) \\ &\quad \int \exp\,[(\alpha_M + (R^M)^{-1}z^{M+1})^*\theta - b(\theta, (R^M)^{-1} + \beta_M)]d\theta\end{aligned}$$

hence, clearly

$$\text{Prob}(z^{M+1}|\mathcal{Z}^M) = \frac{c(\alpha_M, \beta_M)B(z^{M+1}, (R^M)^{-1})}{c(\alpha_M + (R^M)^{-1}z^{M+1}, (R^M)^{-1} + \beta_M)} \tag{183}$$

We can then compute $\text{Prob}(\theta^M|\mathcal{Z}^{M+1})$. Indeed

$$\begin{aligned}\text{Prob}(\theta^M|\mathcal{Z}^{M+1}) &= \frac{\text{Prob}(\theta^M, z^{M+1}|\mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\ &= \frac{\text{Prob}(z^{M+1}|\theta^M, \mathcal{Z}^M)\text{Prob}(\theta^M|\mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\ &= \frac{\text{Prob}(z^{M+1}|\theta^M)\text{Prob}(\theta^M|\mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)}\end{aligned}$$

and thus

$$\begin{aligned}\text{Prob}(\theta^M|\mathcal{Z}^{M+1}) &= c(\alpha_M + (R^M)^{-1}z^{M+1}, (R^M)^{-1} + \beta_M) \\ &\quad \times \exp\,[(\alpha_M + (R^M)^{-1}z^{M+1})^*\theta^M - b(\theta^M, (R^M)^{-1} + \beta_M)]\end{aligned} \tag{184}$$

Using the value of the α_M, β_M, the value of the functions c and b, we obtain after a lengthy calculation that

$$\text{Prob}(\theta^M|\mathcal{Z}^{M+1}) = N(g_M, \Pi_M) \tag{185}$$

with

$$g_M = H^M\hat{x}_M + h^M + \Pi_M(R^M)^{-1}(z^{M+1} - H^M\hat{x}_M - h^M) \tag{186}$$

$$\Pi_M = \left((R^M)^{-1} + (H^M P_M (H^M)^*)^{-1}\right)^{-1} \tag{187}$$

In order to compute the conditional probability $\text{Prob}(x^M|\mathcal{Z}^{M+1})$, we compute the joint conditional probability $\text{Prob}(x^M, \theta^M|\mathcal{Z}^{M+1})$. We have

$$\text{Prob}(x^M, \theta^M|\mathcal{Z}^{M+1}) = \text{Prob}(x^M, \theta^M|z^{M+1}, \mathcal{Z}^M)$$

$$
\begin{aligned}
&= \frac{\text{Prob}(x^M, \theta^M, z^{M+1}|\mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\
&= \frac{\text{Prob}(z^{M+1}|x^M, \theta^M, \mathcal{Z}^M)\text{Prob}(x^M, \theta^M|\mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\
&= \frac{\text{Prob}(z^{M+1}|\theta^M)\text{Prob}(x^M, \theta^M|\mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\
&= \frac{\text{Prob}(z^{M+1}|\theta^M)\text{Prob}(\theta^M|\mathcal{Z}^M)\text{Prob}(x^M|\theta^M, \mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\
&= \frac{\text{Prob}(z^{M+1}, \theta^M|\mathcal{Z}^M)\text{Prob}(x^M|\theta^M, \mathcal{Z}^M)}{\text{Prob}(z^{M+1}|\mathcal{Z}^M)} \\
&= \text{Prob}(\theta^M|\mathcal{Z}^{M+1})\text{Prob}(x^M|\theta^M, \mathcal{Z}^M)
\end{aligned}
$$

We know $\text{Prob}(\theta^M|\mathcal{Z}^{M+1})$. We can define $\text{Prob}(x^M|\theta^M, \mathcal{Z}^M)$ from the knowledge of the joint conditional probability of the pair x^M, θ^M given the $\sigma-$algebra $\mathcal{Z}^M$. However, this joint probability has no density, since θ^M is linked to x^M by a deterministic relation. However, since the pair is gaussian, it is well known that the conditional probability is also Gaussian with mean $E[x^M|\theta^M, \mathcal{Z}^M]$ and covariance

$$E\left((x^M - E[x^M|\theta^M, \mathcal{Z}^M])(x^M - E[x^M|\theta^M, \mathcal{Z}^M])^*|\theta^M, \mathcal{Z}^M\right) = \varDelta_M$$

Classical linear estimation theory for gaussian variables tells that

$$E[x^M|\theta^M, \mathcal{Z}^M] = \hat{x}_M + \Theta_M(\theta^M - H^M\hat{x}_M - h^M) \tag{188}$$

with

$$\Theta_M = P_M(H^M)^*(H^M P_M(H^M)^*)^{-1}$$

and

$$\Delta_M = P_M - Q_M H^M P_M(H^M)^* Q_M^* \tag{189}$$

Hence we have

$$\text{Prob}(x^M, \theta^M|\mathcal{Z}^{M+1}) = N(\theta^M; g_M, \Pi_M)N(x^M; \hat{x}_M + \Theta_M(\theta^M - H^M\hat{x}_M - h^M), \varDelta_M) \tag{190}$$

From this formula the conditional probability $\text{Prob}(x^M|\mathcal{Z}^{M+1})$ is obtained by integrating in θ^M. It is Gaussian with mean

$$
\begin{aligned}
E(x^M|\mathcal{Z}^{M+1}) &= \hat{x}_M + \Theta_M(g_M - H^M\hat{x}_M - h^M) \\
&= \hat{x}_M + \Theta_M\Pi_M(R^M)^{-1}(z^{M+1} - H^M\hat{x}_M - h^M)
\end{aligned}
$$

It remains to show that

$$\Gamma_M(H^M)^* = \Theta_M\Pi_M$$

which is left to the reader.We have proven the result (179). Formula (188) shows also that

$$
\begin{aligned}
&E\left((x^M - E(x^M|\mathcal{Z}^{M+1}))(x^M - E(x^M|\mathcal{Z}^{M+1}))^*|\mathcal{Z}^{M+1}\right)\\
&\quad = E\left((x^M - E(x^M|\theta^M,\mathcal{Z}^M))(x^M - E(x^M|\theta^M,\mathcal{Z}^M))^*|\mathcal{Z}^{M+1}\right)\\
&\qquad + E\left((E(x^M|\theta^M,\mathcal{Z}^M) - E(x^M|\mathcal{Z}^{M+1}))(E(x^M|\theta^M,\mathcal{Z}^M) - E(x^M|\mathcal{Z}^{M+1}))^*|\mathcal{Z}^{M+1}\right)\\
&\quad = \Delta_M + \Theta_M\Pi_M\Theta_M^*
\end{aligned}
$$

and it remains to show that

$$\Delta_M + \Theta_M\Pi_M\Theta_M^* = \Gamma_M$$

which completes the proof. ■

6.4 *First Two Moments Revisited*

6.4.1 General Ideas

In Sect. 6.2.2 we have formulated an approximate recurrence for $\hat{x}_M$ and P_M, namely formulas (164), (167) and (168), (169). It is obtained in two steps, defining quantities $\hat{y}_M$ and Γ_M in terms of $\hat{x}_M$ and P_M, then $\hat{x}_{M+1}$ and P_{M+1}. The major approximation was in considering that the conditional probability of x^M given $\mathcal{Z}^M$ was a Gaussian with mean $\hat{x}_M$ and covariance matrix P_M. In the case of the Kalman filter, we have interpreted $\hat{y}_M$ as $E[x^M|\mathcal{Z}^{M+1}]$ and Γ_M as $\mathrm{Cov}(x^M|\mathcal{Z}^{M+1})$. Also, in the case of the Kalman filter, the Gaussian property is not an approximation. We have revisited the Kalman filter, focusing on the canonical parameter θ^M, instead of the state x^M. We have considered the canonical parameter as a random variable, linked to x^M by a deterministic relation. Thanks to linearity, we could remain in the gaussian framework, and recover all formulas.

In this section, we will follow the same idea, for the dynamic GLM, and focus on the canonical parameter. It is still linked to the state, but this time through a nonlinear deterministic relation. To get a recurrence, an approximation will be needed, but of a different type. This method has been introduced by West et al. 1985 in the case the canonical parameter is a scalar.

6.4.2 Model and Approximation

We follow the notation of Sect. 6.3.3., see Proposition 11. We have first to make more precise the probability density (154), which defines the dependendence of the observation on the canonical parameter. We write

$$f^j(z,\theta) = \exp\left(\theta^*(R^j)^{-1}z - b(\theta,(R^j)^{-1})\right)B(z,(R^j)^{-1}) \tag{191}$$

So the function $b^j(\theta)$, entering into the definition of (154) is clarified in formula (191). The function b is linear in the second argument. The relation, between the canonical parameter and the state is defined as follows

$$\theta^j(x) = \varphi^j(H^j x + h^j)$$

or by the inverse

$$H^j x + h^j = \gamma^j(\theta^j) \tag{192}$$

We next consider the recurrence from M to $M+1$. We suppose we know

$$\hat{x}_M = E[x^M|\mathcal{Z}^M], \; P_M = \mathrm{Cov}(x^M|\mathcal{Z}^M) \tag{193}$$

Note that P_M is not necessarily deterministic.

We define the random variable θ^M by

$$H^M x^M + h^M = \gamma^M(\theta^M)$$

and recall that

$$x^{M+1} = F^M x^M + f^M + G^M w^M$$

therefore

$$E[\gamma^M(\theta^M)|\mathcal{Z}^M) = H^M \hat{x}^M + h^M, \, \mathrm{Cov}(\gamma^M(\theta^M)|\mathcal{Z}^M) = H^M P_M (H^M)^* \tag{194}$$

and

$$\hat{x}^{M+1} = F^M \, E[x^M|\mathcal{Z}^{M+1}] + f^M \tag{195}$$

$$P_{M+1} = F^M \, \mathrm{Cov}(x^M|\mathcal{Z}^{M+1})(F^M)^* + G^M Q^M (G^M)^* \tag{196}$$

Conversely to the Gaussian case, we do not know the conditional probability of θ^M, given $\mathcal{Z}^M$, except for two relations which must be satisfied, namely (194). We then postulate that it has the form (182)

$$\mathrm{Prob}(\theta^M|\mathcal{Z}^M) = c(\alpha_M, \beta_M) \exp\left[\alpha_M^* \theta^M - b(\theta^M, \beta_M)\right] \tag{197}$$

where α_M, β_M are parameters, which we can define, by writing conditions (194). So we write

$$H^M \hat{x}^M + h^M = c(\alpha_M, \beta_M) \int \gamma^M(\theta) \exp\left[\alpha_M^* \theta - b(\theta, \beta_M)\right] d\theta \tag{198}$$

$$\begin{aligned} &H^M P_M (H^M)^* + (H^M \hat{x}^M + h^M)(H^M \hat{x}^M + h^M)^* \\ &= c(\alpha_M, \beta_M) \int \gamma^M(\theta)(\gamma^M(\theta))^* \exp\left[\alpha_M^* \theta - b(\theta, \beta_M)\right] d\theta \end{aligned} \tag{199}$$

These two relations allow, in principle to compute α_M, β_M.
We can then proceed as in Proposition 11, to show that

$$\text{Prob}(\theta^M|\mathcal{Z}^{M+1}) = c(\alpha_M + (R^M)^{-1}z^{M+1}, (R^M)^{-1} + \beta_M) \times \exp[(\alpha_M + (R^M)^{-1}z^{M+1})^*\theta^M - b(\theta^M, (R^M)^{-1} + \beta_M)] \tag{200}$$

and

$$\text{Prob}(x^M, \theta^M|\mathcal{Z}^{M+1}) = \text{Prob}(\theta^M|\mathcal{Z}^{M+1})\text{Prob}(x^M|\theta^M, \mathcal{Z}^M)$$

Therefore again

$$E[x^M|\mathcal{Z}^{M+1}] = E\left(E[x^M|\theta^M, \mathcal{Z}^M]|\mathcal{Z}^{M+1}\right) \tag{201}$$

$$\text{Cov}\,(x^M|\mathcal{Z}^{M+1}) = E\,\left(\text{Cov}\,(x^M|\theta^M, \mathcal{Z}^M)|\mathcal{Z}^{M+1}\right) + \text{Cov}\left(E[x^M|\theta^M, \mathcal{Z}^M]|\mathcal{Z}^{M+1}\right) \tag{202}$$

However, unlike the Kalman filter case, we do not know the conditional probability $\text{Prob}(x^M|\theta^M, \mathcal{Z}^M)$. To compute the quantities on the left hand side of (201), (202), we do not need the full conditional probability. We only need the quantities

$$E[x^M|\theta^M, \mathcal{Z}^M],\ \text{Cov}\,(x^M|\theta^M, \mathcal{Z}^M)$$

The first term is the best estimate of x^M, given $\theta^M, \mathcal{Z}^M$ and the second one is the covariance of the estimation error. Since knowing θ^M is equivalent to knowing $\gamma^M(\theta^M)$, we need to compute

$$E[x^M|\gamma^M(\theta^M), \mathcal{Z}^M],\ \text{Cov}\,(x^M|\gamma^M(\theta^M), \mathcal{Z}^M)$$

The first quantity is the best estimate of x^M, given $\gamma^M(\theta^M), \mathcal{Z}^M$ and the second is the covariance of the residual error. We cannot compute these quantities, because we do not know the conditional probability of x^M, given $\gamma^M(\theta^M), \mathcal{Z}^M$. However we can compute the best linear estimate, because we know

$$E\begin{pmatrix} x^M \\ \gamma^M(\theta^M) \end{pmatrix} |\mathcal{Z}^M) = \begin{pmatrix} \hat{x}^M \\ H^M\hat{x}^M + h^M \end{pmatrix}$$

and

$$\text{Cov}\begin{pmatrix} x^M \\ \gamma^M(\theta^M) \end{pmatrix} |\mathcal{Z}^M) = \begin{pmatrix} P_M & H^M P_M \\ P_M(H^M)^* & H^M P_M(H^M)^* \end{pmatrix}$$

This best linear estimate has been obtained in the case of the Kalman filter. We thus take the approximation

$$E[x^M|\theta^M, \mathcal{Z}^M] \sim \hat{x}_M + \Theta_M(\gamma^M(\theta^M) - H^M\hat{x}_M - h^M) \tag{203}$$

with

$$\Theta_M = P_M (H^M)^* (H^M P_M (H^M)^*)^{-1} \tag{204}$$

and

$$\text{Cov}\,(x^M | \theta^M, \mathcal{Z}^M) \sim \Delta_M = P_M - \Theta_M H^M P_M (H^M)^* \Theta_M^* \tag{205}$$

Next, using (200) we have

$$\begin{aligned} \hat{y}_M = E[x^M | \mathcal{Z}^{M+1}] \sim \hat{x}_M + c(\alpha_M + (R^M)^{-1} z^{M+1}, (R^M)^{-1} + \beta_M)\Theta_M \\ \times \left(\int \gamma^M(\theta) \exp\left[(\alpha_M + (R^M)^{-1} z^{M+1})^* \theta - b(\theta, (R^M)^{-1} + \beta_M)\right] d\theta \right) \\ - \Theta_M (H^M \hat{x}_M + h^M) \end{aligned} \tag{206}$$

and from (202), (205), (203) we obtain

$$\text{Cov}\,(x^M | \mathcal{Z}^{M+1}) \sim \Delta_M + \Theta_M \Pi_M \Theta_M^* \tag{207}$$

with

$$\begin{aligned} \Pi_M &= c(\alpha_M + (R^M)^{-1} z^{M+1}, (R^M)^{-1} + \beta_M) \\ &\times \left(\int \gamma^M(\theta)(\gamma^M(\theta))^* \exp\left[(\alpha_M + (R^M)^{-1} z^{M+1})^* \theta - b(\theta, (R^M)^{-1} + \beta_M)\right] d\theta \right) - \\ &- (c(\alpha_M + (R^M)^{-1} z^{M+1}, (R^M)^{-1} + \beta_M))^2 \\ &\times \left(\int \gamma^M(\theta) \exp\left[(\alpha_M + (R^M)^{-1} z^{M+1})^* \theta - b(\theta, (R^M)^{-1} + \beta_M)\right] d\theta \right)^2 \end{aligned} \tag{208}$$

So summarizing, $\hat{x}_{M+1}$ and P_{M+1} are obtained by formulas (195),(196) with $E[x^M | \mathcal{Z}^{M+1}]$ given by formula (206), $\text{Cov}\,(x^M | \mathcal{Z}^{M+1})$ given by formulas (207), (208) in which α_M and β_M are computed from relations (198), (199).

6.4.3 Further Approximation

Define

$$\hat{\theta}_M = E[\theta^M | \mathcal{Z}^M], \ \text{Cov}(\theta^M | \mathcal{Z}^M) = E(\theta^M (\theta^M)^* | \mathcal{Z}^M) - \hat{\theta}_M (\hat{\theta}_M)^*$$

then we consider the following approximation

$$\gamma_i^M(\theta^M) = \gamma_i^M(\hat{\theta}_M) + D\gamma_i^M(\hat{\theta}_M)(\theta^M - \hat{\theta}_M) + \frac{1}{2} \text{tr} \left(D^2 \gamma_i^M(\hat{\theta}_M)(\theta^M - \hat{\theta}_M)(\theta^M - \hat{\theta}_M)^* \right) \tag{209}$$

so we can write

$$E[\gamma_i^M(\theta^M)|\mathcal{Z}^M] = \gamma_i^M(\hat{\theta}_M) + \frac{1}{2}\mathrm{tr}\left(D^2\gamma_i^M(\hat{\theta}_M)\mathrm{Cov}(\theta^M|\mathcal{Z}^M)\right) \tag{210}$$

the second term being small with respect to the first one. Similarly, we can check that

$$\begin{aligned}\mathrm{Cov}_{ij}(\gamma^M(\theta^M)|\mathcal{Z}^M) = \frac{1}{2}\mathrm{tr}\big((D\gamma_i^M(\hat{\theta}_M)(D\gamma_j^M(\hat{\theta}_M))^* \\ +D\gamma_i^M(\hat{\theta}_M)(D\gamma_j^M(\hat{\theta}_M))^*)\mathrm{Cov}(\theta^M|\mathcal{Z}^M)\big)\end{aligned}$$

Consider the family of matrices

$$K_{ij}(\hat{\theta}_M) = \frac{1}{2}\left(D\gamma_i^M(\hat{\theta}_M)(D\gamma_j^M(\hat{\theta}_M))^* + D\gamma_i^M(\hat{\theta}_M)(D\gamma_j^M(\hat{\theta}_M))^*\right) \tag{211}$$

then we write

$$\mathrm{Cov}_{ij}(\gamma^M(\theta^M)|\mathcal{Z}^M) = \mathrm{tr}(K_{ij}(\hat{\theta}_M)\mathrm{Cov}(\theta^M|\mathcal{Z}^M)) \tag{212}$$

We can rewrite (198), (199) as

$$H^M\hat{x}^M + h^M = \gamma^M(\hat{\theta}_M) \tag{213}$$

$$(H^M P_M(H^M)^*)_{ij} = \mathrm{tr}(K_{ij}(\hat{\theta}_M)\mathrm{Cov}(\theta^M|\mathcal{Z}^M)) \tag{214}$$

Now $\hat{\theta}_M$ and $\mathrm{Cov}(\theta^M|\mathcal{Z}^M)$ can be expressed as functions of α_M, β_M. We have, from (182),

$$\hat{\theta}_M = -D_{\alpha_M}\log c(\alpha_M,\beta_M) \tag{215}$$

$$\mathrm{Cov}(\theta^M|\mathcal{Z}^M) = -\frac{D^2_{\alpha_M}c}{c}(\alpha_M,\beta_M) + D_{\alpha_M}\log c(\alpha_M,\beta_M)(D_{\alpha_M}\log c(\alpha_M,\beta_M))^* \tag{216}$$

and thus (213), (214) is a nonlinear system of algebraic equation*, but does not involves integrals. Similarly, define

$$\hat{\theta}_M^{M+1} = E[\theta^M|\mathcal{Z}^{M+1}],\ \mathrm{Cov}(\theta^M|\mathcal{Z}^{M+1}) = E(\theta^M(\theta^M)^*|\mathcal{Z}^{M+1}) - \hat{\theta}_M^{M+1}(\hat{\theta}_M^{M+1})^* \tag{217}$$

These quantities can be computed from the probability (200), but this time we cannot avoid computing the integrals

$$\begin{aligned}\hat{\theta}_M^{M+1} &= c(\alpha_M + (R^M)^{-1}z^{M+1}, (R^M)^{-1} + \beta_M) \\ &\quad\times\int \theta\exp\left[(\alpha_M + (R^M)^{-1}z^{M+1})^*\theta - b(\theta^M, (R^M)^{-1} + \beta_M)\right]d\theta\end{aligned} \tag{218}$$

$$E(\theta^M(\theta^M)^*|\mathcal{Z}^{M+1}) = c(\alpha_M + (R^M)^{-1}z^{M+1}, (R^M)^{-1} + \beta_M) \tag{219}$$
$$\times \int \theta\theta^* \exp[(\alpha_M + (R^M)^{-1}z^{M+1})^*\theta - b(\theta^M, (R^M)^{-1} + \beta_M)]d\theta$$

Knowing these quantities we can approximate formulas (206) and (208). We write

$$\hat{y}_M = E[x^M|\mathcal{Z}^{M+1}] = \hat{x}_M + \Theta_M(\gamma^M(\hat{\theta}_M^{M+1}) - (H^M\hat{x}_M + h^M)) \tag{220}$$

and

$$(\Pi_M)_{ij} = \text{tr}(K_{ij}(\hat{\theta}_M^{M+1})\text{Cov}(\theta^M|\mathcal{Z}^{M+1})) \tag{221}$$

6.5 Example of a Beta Model

We report here a simplified example discussed in (da-Silva 2011), with state $x \in R^n$ and observation $z \in (0,1)$. The canonical parameter $\theta \in (0,1)$. The probability density, for a value θ of the canonical parameter is given by

$$f(z,\theta) = \frac{z^{\theta-1}(1-z)^{-\theta}}{\Gamma(\theta)\Gamma(1-\theta)} \tag{222}$$

This is not in the GLM form, but the methodology will be easily adapted. At each experiment (each time) j, the canonical parameter will depend on the state x^j by the relation

$$\theta^j(x^j) = \frac{\exp(h^j)^*x^j}{1+\exp(h^j)^*x^j} \tag{223}$$

and the evolution of the state is given by

$$x^{j+1} = F^jx^j + f^j + G^jw^j \tag{224}$$

We describe the procudure between M and $M+1$.We know

$$\hat{x}_M = E[x^M|\mathcal{Z}^M], \; P_M = \text{Cov}(x^M|\mathcal{Z}^M).$$

Considering the function

$$\gamma(\theta) = \log\frac{\theta}{1-\theta}, \; 0 < \theta < 1$$

we define the variable θ^M by the relation

$$\gamma(\theta^M) = (h^M)^*x^M$$

and we pick as Prob$(\theta^M|\mathcal{Z}^M)$ the probability density, depending on parameters α_M, β_M

$$\text{Prob}(\theta^M|\mathcal{Z}^M) = \frac{\Gamma(\alpha_M+\beta_M)}{\Gamma(\alpha_M)\Gamma(\beta_M)}(\theta^M)^{\alpha_M-1}(1-\theta^M)^{\beta_M-1} \tag{225}$$

Setting

$$p^M(\theta) = \frac{\Gamma(\alpha_M+\beta_M)}{\Gamma(\alpha_M)\Gamma(\beta_M)}(\theta)^{\alpha_M-1}(1-\theta)^{\beta_M-1}$$

$$= \exp\left[(\alpha_M-1)\log\theta + (\beta_M-1)\log(1-\theta) - B(\alpha_M,\beta_M)\right]$$

with

$$B(\alpha_M,\beta_M) = -\log\Gamma(\alpha_M+\beta_M) + \log\Gamma(\alpha_M) + \log\Gamma(\beta_M)$$

We check easily the formulas

$$\frac{\partial B}{\partial\alpha_M} = \int_0^1 \log\theta\; p^M(\theta)d\theta \tag{226}$$

$$\frac{\partial B}{\partial\beta_M} = \int_0^1 \log(1-\theta)\; p^M(\theta)d\theta \tag{227}$$

and thus

$$\frac{\partial B}{\partial\alpha_M} - \frac{\partial B}{\partial\beta_M} = \int_0^1 \gamma(\theta)p^M(\theta)d\theta$$

Therefore,

$$(h^M)^*\hat{x}_M = \Psi(\alpha_M) - \Psi(\beta_M) \tag{228}$$

with $\Psi(x) = \frac{d}{dx}\log\Gamma(x)$.

Equation (228) provides a first relation to compute the pair α_M, β_M. To get a second one, we note that

$$\text{Var}(\gamma(\theta)|\mathcal{Z}^M) = \text{Var}(\log\theta^M|\mathcal{Z}^M) + \text{Var}(\log(1-\theta^M)|\mathcal{Z}^M)$$
$$- 2\text{Cov}(\log\theta^M, \log(1-\theta^M)|\mathcal{Z}^M)$$

We then check

$$\text{Var}(\log\theta^M|\mathcal{Z}^M) = \frac{\partial^2 B}{\partial\alpha_M^2} = \Psi'(\alpha_M) - \Psi'(\alpha_M+\beta_M)$$

$$\text{Var}(\log(1-\theta^M)|\mathcal{Z}^M) = \frac{\partial^2 B}{\partial\beta_M^2} = \Psi'(\beta_M) - \Psi'(\alpha_M+\beta_M)$$

$$\mathrm{Cov}(\log \theta^M, \log(1-\theta^M)|\mathcal{Z}^M) = \frac{\partial^2 B}{\partial \alpha_M \partial \beta_M} = -\Psi'(\alpha_M + \beta_M)$$

Collecting results, we obtain

$$\mathrm{Var}(\gamma(\theta)|\mathcal{Z}^M) = \Psi'(\alpha_M) + \Psi'(\beta_M)$$

which leads to the second relation

$$(h^M)^* P_M h^M = \Psi'(\alpha_M) + \Psi'(\beta_M) \tag{229}$$

and relations (228), (229) allow to obtain α_M, β_M. If we accept the approximation $\Psi(x) \sim \log x$,we obtain

$$\alpha_M = \frac{1 + \exp(h^M)^* \hat{x}_M}{(h^M)^* P_M h^M} \tag{230}$$
$$\beta_M = \frac{1 - \exp(h^M)^* \hat{x}_M}{(h^M)^* P_M h^M}$$

We can next formulate $\mathrm{Prob}(\theta^M|\mathcal{Z}^{M+1})$. From (222) we have

$$\mathrm{Prob}(z^{M+1}|\theta^M, \mathcal{Z}^M) = \frac{(z^{M+1})^{\theta^M - 1}(1 - z^{M+1})^{-\theta^M}}{\Gamma(\theta^M)\Gamma(1-\theta^M)}$$

and using (225) we can write

$$\mathrm{Prob}(\theta^M|\mathcal{Z}^{M+1}) = \frac{(z^{M+1})^{\theta^M - 1}(1 - z^{M+1})^{-\theta^M}(\theta^M)^{\alpha_M - 1}(1-\theta^M)^{\beta_M - 1}}{\Gamma(\theta^M)\Gamma(1-\theta^M)D(z^{M+1})} \tag{231}$$

in which

$$D(z^{M+1}) = \int_0^1 \frac{(z^{M+1})^{\theta - 1}(1 - z^{M+1})^{-\theta}(\theta)^{\alpha_M - 1}(1-\theta)^{\beta_M - 1}}{\Gamma(\theta)\Gamma(1-\theta)} d\theta \tag{232}$$

We can then compute

$$\begin{aligned}\hat{\theta}_M^{M+1} &= E[\theta^M|\mathcal{Z}^{M+1}] \\ &= \int_0^1 \frac{(z^{M+1})^{\theta - 1}(1 - z^{M+1})^{-\theta}\theta^{\alpha_M}(1-\theta)^{\beta_M - 1}}{\Gamma(\theta)\Gamma(1-\theta)D(z^{M+1})} d\theta\end{aligned}$$

$$V_M^{M+1} = \mathrm{Var}(\theta^M|\mathcal{Z}^{M+1})$$

$$= \int_0^1 \frac{(z^{M+1})^{\theta - 1}(1 - z^{M+1})^{-\theta}\theta^{\alpha_M + 1}(1-\theta)^{\beta_M - 1}}{\Gamma(\theta)\Gamma(1-\theta)D(z^{M+1})} d\theta \; - (\hat{\theta}_M^{M+1})^2$$

Set next

$$\Theta_M = \frac{P_M h^M}{(h^M)^* P_M h^M}$$

then we have

$$\hat{y}_M = E[x^M | \mathcal{Z}^{M+1}] = \hat{x}_M + \Theta_M \left(\log \frac{\hat{\theta}_M^{M+1}}{1 - \hat{\theta}_M^{M+1}} - (h^M)^* \hat{x}_M \right) \tag{233}$$

We next set

$$\Delta_M = P_M - \frac{P_M h^M (P_M h^M)^*}{(h^M)^* P_M h^M}$$

and

$$K(\theta) = (\gamma'(\theta))^2 = \frac{1}{\theta^2 (1-\theta)^2}$$

$$\mathrm{Cov}(x^M | \mathcal{Z}^{M+1}) = P_M + \frac{P_M h^M (P_M h^M)^*}{((h^M)^* P_M h^M)^2} (K(\hat{\theta}_M^{M+1}) V_M^{M+1} - (h^M)^* P_M h^M) \tag{234}$$

Finally

$$\hat{x}_{M+1} = F^M \, \hat{y}_M + f^M \tag{235}$$

$$P_{M+1} = F^M \, \mathrm{Cov}(x^M | \mathcal{Z}^{M+1})(F^M)^* + G^M Q^M (G^M)^* \tag{236}$$

References

da-Silva, C.Q., H.S. Migon, and L.T. Correia. 2011. Dynamic Bayesian beta models. *Computational Statistics and Data Analysis* 55:2074–2089

Ibramovic, I.A., and R.Z. Has'minskii. 1981. *Statistical estimation*. New York: Springer

Loève, M. 1978. *Probability theory*. New York: Springer

McCullagh, P., and J.A. Nelder. 1983. *Generalized linear models*. London: Chapman and Hall

Nelder, J.A., and R.W.M. Wedderburn. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* 135 (3): 370–384

Sorenson, W.W. 1980. *Parameter estimation: Principles and problems*. New York: Marcel Dekker

West, M., P.J. Harrison, and H.S. Migon. 1985. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* 80 (389): 75–83.

Distortion Risk Measure or the Transformation of Unimodal Distributions into Multimodal Functions

Dominique Guégan and Bertrand Hassani

1 Introduction

A commonly used risk metrics is the standard deviation. For examples mean-variance portfolio selection maximises the expected utility of an investor if the utility is quadratic or if the returns are jointly normal. Mean-variance portfolio selection using quadratic optimisation was introduced by Markowitz (1959) and became the standard model. This approach was generalized for symmetrical and elliptical portfolio (Ingersoll 1987; Huang and Litzenberger 1988). However, the assumption of elliptically symmetric return distributions became increasingly doubtful (Bookstaber and Clarke 1984; Chamberlain 1983) to characterize the returns distributions making standard deviation an intuitively inadequate risk measure.

Recently the financial industry has extensively used quantile-based downside risk measures based on the Value-at-Risk (VaR_α for confidence level α). While the VaR_α measures the losses that may be expected for a given probability it does not address how large these losses can be expected when tail events occur. To address this issue the mean excess function has been introduced, (Rockafellar and Uryasev 2000; Embrechts et al. 2005; Artzner et al. 1999) and Delbaen (2000) describe the properties that risk measures should satisfy including their coherence in particular the VaR is not a coherent risk measure, failing to be sub-additive.

When we use a sub-additive measure the diversification of the portfolio always leads to risk reduction while if we use measures violating this axiom the diversification benefit may be lost even if partial risks are triggered by mutually exclusive events.

D. Guégan (✉)
University Paris 1 Panthéon-Sorbonne et New York University Polytechnic School of Engineering, Brooklyn, New York, USA

CES UMR 8174, 106 boulevard de l'Hopital, 75647 Paris Cedex 13, France
tel +33144078298
e-mail: dguegan@univ-paris1.fr

B. Hassani
University Paris 1 Panthéon-Sorbonne, CES UMR 8174, 106 boulevard de l'Hopital, 75647 Paris Cedex 13, France tel +44 (0)2070860973
e-mail: bertrand.hassani@gmail.com

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,*
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_2,

The sub-additive property is required for capital adequacy purposes in banking supervision: for instance if we consider a financial institution made of several subsidiaries or business units, if the capital requirement of each of them is dimensioned to its own risk profile authorities. Consequently it has appeared relevant to construct a more flexible risk measure which is sub-additive.

Nevertheless, the VaR remains preeminent even though it suffers from the theoretical deficiency of not being sub-additive. The problem of sub-additivity violations is not as important for assets verifying the regularity conditions[1] than for those which do not and for most assets these violations are not expected. Indeed, in most practical applications the VaR_α can have the property of sub-additivity. For instance, when the return of an asset is heavy tailed, the VaR_α is sub-additive in the tail region for high level of confidence if it is computed with the heavy tail distribution (Ingersoll 1987; Danielson et al. 2005; Embrechts et al. 2005). Non sub-additivity of the VaR_α is highlighted when assets have very skewed return distributions. When the distributions are smooth and symmetric, when assets dependency is highly asymmetric, and when underlying risk factors are dependent but heavy-tailed, it is necessary to consider other risks measures.

Unfortunately, non sub-additivity is not the only problem characterizing the VaR. First VaR only measures distribution percentiles and thus disregards any loss beyond its confidence level. Due to combined effects of this limitation and the occurrence of extreme losses there is a growing interest for risk managers to focus on the tail behavior and its Expected Shortfall[2] (ES_α) since it shares properties that are considered desirable and applicable in a variety of situations. Indeed, expected shortfall considers the loss beyond the VaR_α confidence level and is sub-additive and therefore it ensures the coherence of the risk measure (Rockafellar and Uryasev 2000).

Since using expected utility, the axiomatic approach to risk theory has expanded dramatically as illustrated by (Yaari 1987; Panjer et al. 1997; Artzner et al. 1999; De Giorgi 2005; Embrechts et al. 2005; Denuit et al. 2006) among others. Thus other classes of risk measures were proposed each with their own properties including convexity (Follmer and Shied 2004), spectral properties (Acerbi and Tasche 2002), notion of deviation (Rockafellar et al. 2006) or distortion (Wang et al. 1997). Acerbi and Tasche (2002) studied spectral risk measures which involve a weighted average of expected shortfalls at different levels. Then, the dual theory of choice under risk leads to the class of distortion risk measures developed by Yaari (1987) and Wang (2000), which transforms the probability distribution shifting it in order to better quantify the risk in the tails instead of modifying returns as in the expected utility framework.

Whatever the risk measures considered, the value associated to each measure is based and depends on the distribution fitted on the underlying data set by risk managers strategy. Mostly of the part the distributions belong to the elliptical domain,

---

[1] Regularly varying (heavy tailed distributions, fat tailed) non-degenerate tails with tail index $\eta > 1$ for more detail see Danielson et al. (2005).

[2] The terminology “Expected shortfall” was proposed by Acerbi and Tasche (2002) . A common alternative denotation is “Conditional Value at Risk” or CVaR that was suggested by Rockafellar and Uryasev (2002).

recently risk managers and researchers have focused on a class of distributions exhibiting asymmetry and producing heaving tails, All these distributions belong to the Generalized Hyperbolic class of distributions (Barndorff-Nielsen 1977), to the α-stable distributions (Samorodnitsky and Taqqu 1994) or the g- and -h distributions among others.

Nevertheless nearly all these distributions are unimodal. However, since the 2000s bubbles and financial crises and extreme events became more and more important, restricting unimodal distributions models for risk measures. Recently debates have been opened to convince economists to consider bimodal distributions instead of unimodal distributions to explain the evolution of the economy since the 2000s (Bhansali 2012). The debate about the choice of distributions characterized by several modes is timely. We propose an approach to build and fit these distributions on real data sets. An objective of this paper is to discuss this new approach and propose a theoretical framework to build multi-modal distributions to create new coherent risk measures.

The paper is organized as follows. In Section two we recall some principles and history of the risk measures: the VaR, the ES and the spectral measure. In Section three we discuss the notion of distortion to create new distributions. Section four proposes an application which illustrates the impact of the choice of unimodal or bimodal distribution associated to different risk measures to provide a value for the corresponding risk. Section five concludes.

2 Quantile-Based and Spectral Risk Measures

Traditional deviation risks measures such as the variance, the mean-variance analysis and the standard deviation, are not sufficient within the context of capital requirements. In this section we recall the definitions of several quantile-based risk measures:[3] the Value-at-Risk introduced in the 1980s, the Expected Shortfall proposed by Acerbi and Tasche (2002), the Tail Conditional Expectation suggested by Rockafellar and Uryasev (2002), and the spectral measure introduced by Acerbi and Tasche (2002).

Value at Risk initially used to measure financial institutions market risk, was mainly popularised by J.P. Morgan's RiskMetrics (1995). This measure indicates the maximum probable loss, given a confidence level and a time horizon. The VaR is sometimes referred as the "unexpected" loss.

Definition 1 Given a confidence level $\alpha \in (0, 1)$, the VaR is the relevant quantile[4] of the loss distribution: $VaR_\alpha(X) = \inf\{x \mid P[X > x] \leqslant 1-\alpha\} = \inf\{x \mid F_X(x) \geqslant \alpha\}$ where X is a risk factor admitting a loss distribution F_X .

---

[3] Artzner (2002) proposes a natural way to define a measure of risk as a mapping $\rho : L^\infty \to \mathbb{R} \cup \infty$.

[4] $VaR_\alpha(X) = q_{1-\alpha} = F_X^{-1}(\alpha)$

Table 1 Expression of ES_α as function of VaR_α for some usual distributions in finance

Distribution	$ES_\alpha = f(VaR_\alpha)$
Normal	$\mu + \frac{\sigma}{\sqrt{2\pi}} \frac{exp\left[-\frac{1}{2}\left(\frac{VaR_\alpha-\mu}{\sigma}\right)^2\right]}{1-\alpha}$
Student-t	$\mu + \frac{\sigma}{1-\alpha}\frac{\sqrt{\eta}}{(\eta-1)\sqrt{\pi}}\frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)}\left(1+\frac{(VaR_\alpha(X)-\mu)^2}{\sigma^2\eta}\right)^{-\frac{\eta+1}{2}}$
Logistic	$\mu + \frac{\sigma}{1-\alpha}\left(\ln\left(1+e^{VaR_\alpha(X)}\right) - VaR_\alpha(X)\left[1+e^{-(VaR_\alpha(X))}\right]^{-1}\right)$
Exponential power	$\mu + \frac{\sigma\beta^{\left(\frac{1}{\beta}-1\right)}}{2(1-\alpha)\Gamma(1+1/\beta)}\Gamma\left(\frac{2}{\beta},\frac{1}{\beta}\left(\frac{VaR_\alpha(X)-\mu}{\sigma}\right)^\beta\right)$
Generalized hyperbolic	$\mu + \beta\mathbb{E}(W) + \frac{\sigma}{\sqrt{2\pi}}\frac{exp\left[-\frac{1}{2}\left(\frac{VaR_\alpha-\mu}{\sigma}\right)^2\right]}{1-\alpha}\mathbb{E}(\sqrt{W})$
Generalized pareto	$\frac{VaR_\alpha(X)}{1-\xi} + \frac{\sigma-\xi u}{1-\xi}$
g and h	$\mu + \frac{\sigma}{g(1-\alpha)\sqrt{1-h}}\left[e^{(g^2/2(1-h))}\bar{\phi}\left(\sqrt{1-h}z_\alpha - \frac{g}{1-h}\right) - \bar{\phi}\left(\sqrt{1-h}z_\alpha\right)\right]$

As discussed in the Introduction, the VaR does not always appear sufficient. When a tail event occurs in a unimodal distribution, the loss in excess of the VaR is not captured. To avoid this problem we consider the expected shortfall (ES_α) proposed by Artzner et al. (1999). This measure is more conservative than the VaR_α as it captures the information contained in the tail. The expected shortfall is defined as follows:

Definition 2 The Expected Shortfall (ES_α) is defined as the average of all losses which are greater or equal than VaR_α:

$$ES_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\alpha dp$$

The Expected Shortfall has a number of advantages over the VaR_α. Accordingly the ES takes accounts for the tail risk and fulfills the sub-additive property[5] (Acerbi and Tasche 2002)[6] . Table 1 summarizes the link between ES_α and VaR_α for some distributions given α.

Expected Shortfall is the smallest coherent risk measure that dominates the VaR. Acerbi and Tasche (2002) derived from this concept a more general class of coherent risk measures called spectral risk measures[7]. Spectral risk measures are a subset of coherent risk measures. Instead of averaging losses beyond the VaR, a weighted

[5] An extension can be found in Inui and Kijima (2005).

[6] In this last paper, the difference between ES and TCE is conceptual and is only related to the distributions. If the distribution is continuous then the expected shortfall is equivalent to the tail conditional expectation.

[7] If ρ_i is coherent risk measures for $i = 1 \ldots n$, then, any convex combination $\rho = \sum_1^n \beta_i \rho_i$ is a coherent risk measure (Acerbi and Tasche 2002).

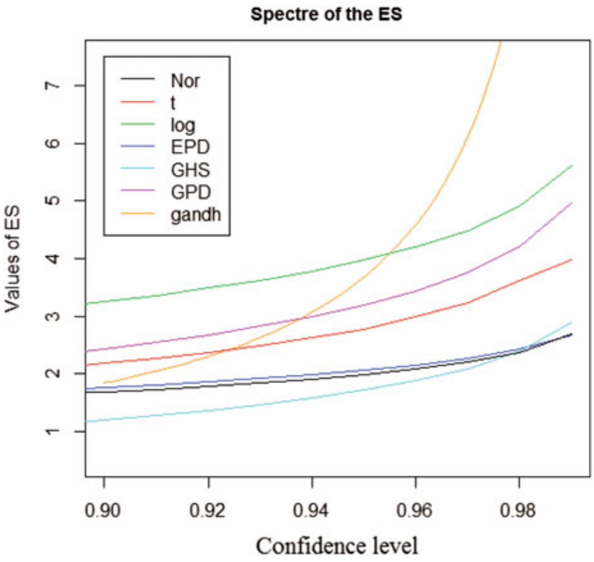


Fig. 1 Spectrum of the ES for some well known distributions for several $\alpha \in [0.9, 0.99]$. Each *line* corresponds to the graph of the ES as a function of α for each distribution introduced in Table 1

average of different levels of ES_α is used. These weights characterize risk aversion: different weights are assigned to different α levels of ES_α in the left tail. The associated spectral measure could be $\sum_\alpha w_\alpha ES_\alpha$, where $\sum_\alpha w_\alpha = 1$. In Fig. 1 we exhibit a spectrum corresponding to the sequence of ES_α for different α.

Figure 1 points out that the spectrum of the ES is an increasing function of the confidence level α. It expresses the risk aversion as a weighted average for different level of ES_α to generate the spectral risk measure. This is one advantage when using a spectral risk measure. Moreover a spectral risk measure being a convex combination of ES_α for $\alpha \in [0.9, 0.99]$, it accounts for more information than only considering one value of α.

However the choice of weights is sensitive and need to be studied more carefully (Dowd et al. 2008). Finally, in practice the relation between spectral risk measure and risk aversion is not obvious depending on the choice of the weights.

3 Distortion Risk Measures

3.1 Notion of Distortion Risk Measures

Distortion risk measures have their origin in Yaari's (1987) dual theory of choice under risk that consists in measuring the risks by applying a distortion function g on the cumulative distribution function F_X. In order to transform a distribution into a new distribution we need to specify the property of the distorsion function g.

Definition 3 A function $g : [0, 1] \rightarrow [0; 1]$ is a distortion function if:

1. $g(0) = 0$ and $g(1) = 1$,
2. g is a continuous increasing function.

In order to quantify the risk instead of modifying the loss distribution (as with the expected utility framework), the distortion approach modifies the probability distribution. The risk measures (VaR and ES) derived from this transformation were originally applied to a wide variety of financial problems such as the determination of insurance premiums (Wang 2000), economic capital (Hürlimann 2004), and capital allocation (Tsanakas 2004). Acerbi (2002) suggests that they can be used to set capital requirements or obtain optimal risk-expected return trade-offs and could also be used by clearing-houses to set margin requirements that reflect their corporate risk aversion (Cotter and Dowd 2006).

One possibility is to shift the distribution function towards the left or the right sides to account for extreme values. Wang et al. (1997) developed the concept of distortion[8] risk measure by computing the expected loss from a non-linear transformation of the cumulative probability distribution of the risk factor. A formal definition of this risk measure computed from a distortion of the original distribution has been derived (Wang et al. 1997).

Definition 4 The distorted risk measure $\rho_g(X)$ for a risk factor X admitting a cumulative distribution $S_X(x) = \mathbb{P}(X > x)$, with a distortion function g, is defined[9] as:

$$\rho_g(x) = \int_{-\infty}^{0} [g(S_X(x)) - 1]dx + \int_{0}^{+\infty} g(S_X(x))dx. \tag{1}$$

Such a distortion risk measure corresponds to the expectation of a new variable whose probabilities have been re-weighted.

Finding appropriate distorted risk measures reduces to the choice of an appropriate distortion function g. Properties for the choice of a distortion function include continuity, concavity, and differentiability. Assuming g is differentiable on $[0, 1]$ and $S_X(x)$ is continuous, then a distortion risk measure can be re-written as:

$$\rho_g(X) = \mathbb{E}[Xg'(S_X(X))] = \int_{0}^{1} F_X^{-1}(1-p)dg(p) = \mathbb{E}_g\left[F_X^{-1}\right]. \tag{2}$$

Distortion functions arose from empirical[10] observations that people do not evaluate risk as a linear function of the actual probabilities for different outcomes but

[8] The distortion risk measure is a special class of the so-called Choquet expected utility, i.e. the expected utility calculated under a modified probability measure.

[9] Both integrals in (1) are well defined and take a value in $[0, +\infty]$. Provided that at least one of the two integrals is finite, the distorted expectation $\rho_g(X)$ is well defined and takes a value in $[-\infty, +\infty]$.

[10] This approach towards risk can be related to investor's psychology as in Kahneman and Tversky (1979).

rather as a non-linear distortion function. It is used to transform the probabilities of the loss distribution to another probability distribution by re-weighting the original distribution. This transformation increases the weight given to desirable events and deflates others. Different distortions g have been proposed in the literature. A wide range of parametric families of distortion functions is mentioned in Wang (2000), and Hardy and Wirch (2001). For well known utility functions we provide the function g in Table 2, where the parameters k and γ represent the confidence level corresponding and the level of risk aversion.

Table 2 Examples of utility functions with their associated convex spectrum

	Utility function	Parameters	Spectrum function
Exponential	$U_1(x) = -e^{-kx}$	$k > 0$	$g(p, k) = \frac{ke^{-k(1-p)}}{1-e^{-k}}$
Power	$U_2(x) = x^{1-\gamma}$	$\gamma \in (0, 1)$	$g(p, \gamma) = \gamma(1-p)^{\gamma-1}$
Power	$U_3(x) = x^{1-\gamma}$	$\gamma > 1$	$g(p, \gamma) = \gamma(p)^{\gamma-1}$

When g is a concave function its first derivative g' is an increasing function, $g'(S_X(x))$ is a decreasing function[11] in x and $g'(S_X(x))$ represents a weighted coefficient which discounts the probability of desirable events while loading the probability of adverse events. Moreover, Hardy and Wirch (2001) have shown that distorted risk measure $\rho_g(X)$ introduced in (2) is sub-additive and coherent if and only if the distortion function is concave.

In his article, Wang (2000) specifies that the distortion operator g can be applied to any distribution. Nevertheless in applications due to technical practical reasons he restricts the illustration of his methodology to a function g defined as follows:

$$g_\alpha(u) = \Phi\left[\Phi^{-1}(u) + \alpha\right], \tag{3}$$

where Φ is the Gaussian cumulative distribution. In other words he applies the same perspective of preference to quantify the risk associated to gain and risk. Thus, a risk manager evaluates the risk associated to the upside and downside risks with the same function g implying a symmetric consideration for the two effects due to the distortion. Moreover it induces the same confidence level for the losses and the gain which implies the same level of risk aversion associated to the losses and the gains.

In Fig. 2 we illustrate the impact of the Wang (2000) distortion function introduced in Eq. (3) on the logistic distribution provided in Table 1. We can remark that the distorted distribution is always symmetrical under this kind of distortion function, and we observe a shift of the mode of the initial distribution towards the left.

To avoid the problem of symmetry in the previous distorsion, Sereda et al. (2010) propose to use two different functions issued from the same polynomial with different coefficients, say:

$$\rho_{g_i}(X) = \int_{-\infty}^{0} [g_1(S_X(x)) - 1]\, dx + \int_{0}^{+\infty} g_2(S_X(x))dx. \tag{4}$$

[11] This property involves that $g'(S_X(x))$ becomes smaller for large values of the random variable X.

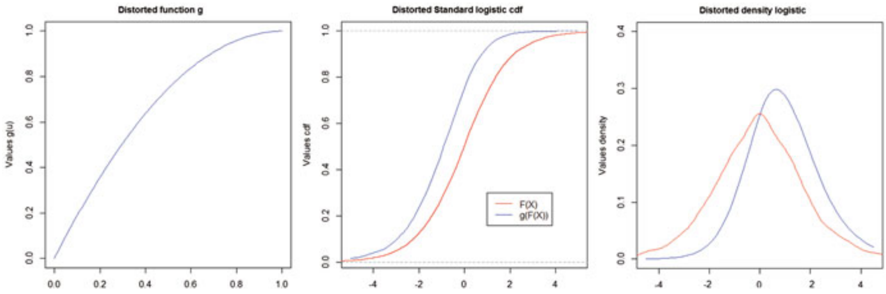


Fig. 2 Distortion of logistic distribution with mean 0 using a Wang distortion function with confidence level 0.65. It illustrates the effect of distortion

with $g_i(u) = u + k_i\left(u - u^2\right)$ for $k_i \in]0, 1]$ et $\forall i \in \{1, 2\}$. With this approach one models loss and gains differently relatively to the values of the parameters $k_i, i = 1, 2$. Thus upside and downside risks are modeled in different ways. Nevertheless the calibration of the parameters $k_i, i = 1, 2$ remains an open problem.

To create bimodal or multi-modal distributions we have to impose other properties to the distortion function g. Indeed, transforming an unimodal distribution into a bimodal one provides different approaches to the risk aversion of losses and gains. This will allow us to introduce a new coherent risk measure in that latter case.

3.2 A New Coherent Risk Measure

We begin to discuss the choice of the function g to obtain a bimodal distribution. To do so we need to use a function g which creates saddle points. The saddle point generates a second hump in the new distribution which allows us to take into account different patterns located in the tails. The distortion function g fulfilling this objective is an inverse S-shaped polynomial function of degree 3 given by the following equation and characterized by two parameters δ and β:

$$g_\delta(x) = a\left[\frac{x^3}{6} - \frac{\delta}{2}x^2 + \left(\frac{\delta^2}{2} + \beta\right)x\right]. \tag{5}$$

We remark that $g_\delta(0) = 0$, and to get $g_\delta(1) = 1$ this implies that the coefficient of normalization is equal $a = \left(\frac{1}{6} - \frac{\delta}{2} + \frac{\delta^2}{2} + \beta\right)^{-1}$. The function g_δ will increase if $g'_\delta > 0$ requiring $0 < \delta < 1$. The parameter $\delta \in [0, 1]$ allow us to locate the saddle point. The curve exhibits a concave part and a convex part. The parameter $\beta \in \mathbb{R}$ controls the information under each hump in the distorted distribution. To illustrate the role of δ on the location of the saddle points, we provide several simulations.

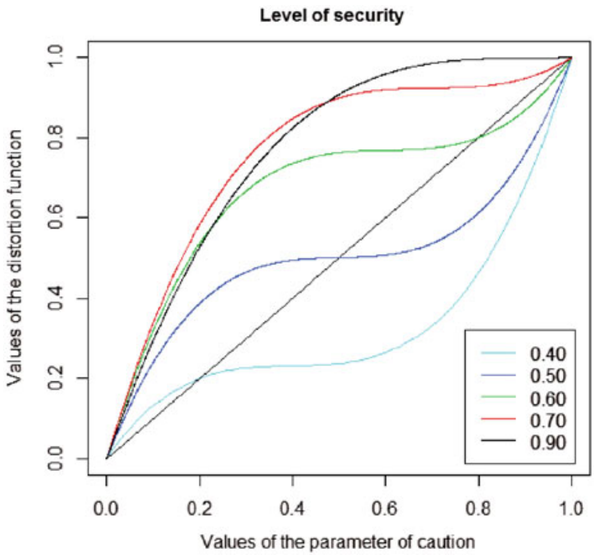


Fig. 3 Curves of the distortion function g_δ introduced in Eq. (5) for several value of δ and fixed values of $\beta = 0.001$

In Fig. 3, the value of the level of the discrimination of an event is given by $\beta = 0.001$ then we plot the function g_δ for different values of δ. This parameter β illustrates the fact that some events are discriminating more than others. Figure 3 shows the location of the saddle point creating convex and concave parts inside the domain $[0, 1]$. The convex part can be associated to the negative values of the returns associated to the losses and the concave part is associated to positive returns. We observe in this picture that for high values of δ the concave part diminishes and then the effect of saddle point decreases.

Variations in β in Fig. 4 exhibit different patterns for a fixed value of δ.

To understand the influence of the parameter β on the shape of the distortion function we use three graphs in Fig. 4. The two left graphs correspond to the same value of the parameters. The middle figure zooms on the x-axis from $[0, 1]$ to $[-4, 4]$. We show that the function g may not have a saddle point on $]0, 1[$ depending on the values of β. The right graph provides different representations of the distorsion function for several values of β. We observe that if β tends to 1 then the distortion function g tends to the identity mapping and when β tends to 0 the curve is more important and the effect of g on the distribution will be more important.

Figure 5 illustrates the effect of distortion of the Gaussian distribution for several values of β and fixed $\delta = 0.50$. We observe the same effects as in Fig. 4. For small values of the parameter β (0.00005 or 0.005) the distortion function has two distinct parts, one convex part for $x \in]0, 0.5[$ and one concave part for $x \in]0.5, 1[$. Moreover when β is close to 1 then the distorted cumulative distribution tends to the initial Gaussian variable.

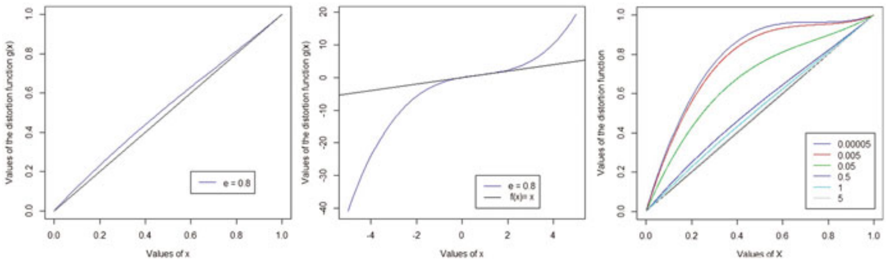


Fig. 4 The effect of β on the distortion function for a level of security $\delta = 0.75$ showing that if β tends to 1, the distortion function tends to the identity function

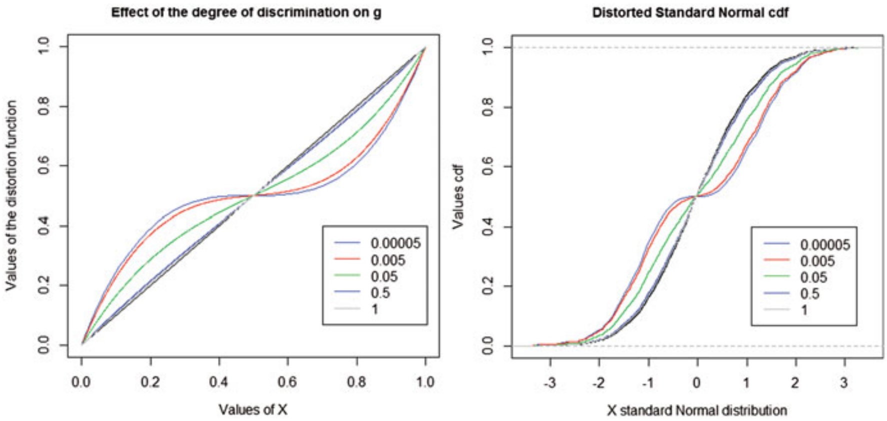


Fig. 5 The effect of β on the cumulative Gaussian distribution for $\delta = 0.50$

Figure 6 points out the effect of distortion on the density of the Gaussian distribution using the same values of the parameters than those used in Fig. 5. Again we generate a new distribution with two humps. Making both parameters varying permits to solve one of our objective: to create a asymmetrical distribution with more than one hump.

It is important to notice that the function g_δ creates a distorted density function which associates a small probability in the centre of the distribution and put greater weight in the tails. This phenomenon is illustrated in Fig. 7 where the derivative of g (density) indicates how weights on the tails can be increased.

Such discrimination is also illustrated in Fig. 8 which exhibits the particular effect of parameter β when δ is fixed to 0.75 for the creation of humps. From a Gaussian distribution, applying g_δ defined in (5), with $\delta = 0.75$ and $\beta = 0.48$ we create a distribution for which the probability of occurrences of the extremes in the right part is bigger than the probability of occurrence of the extremes in the left part which can be counter-intuitive for risk management but interesting from a theoretical point of view.

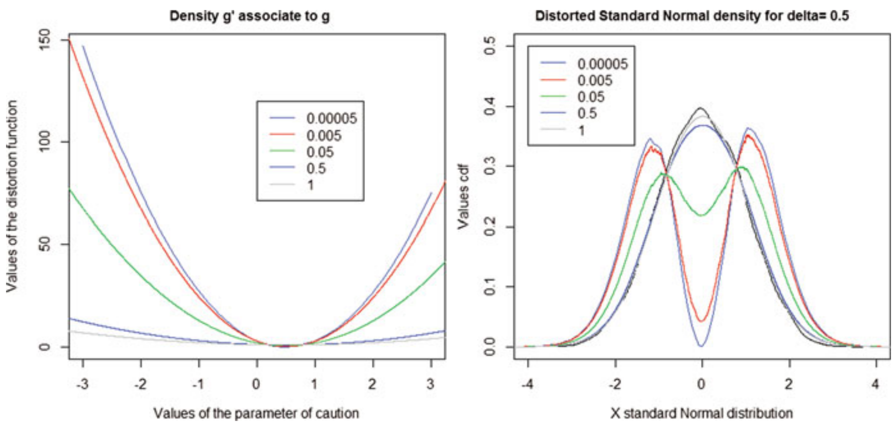


Fig. 6 The effect of β on the Gaussian density function for $\delta = 0.50$

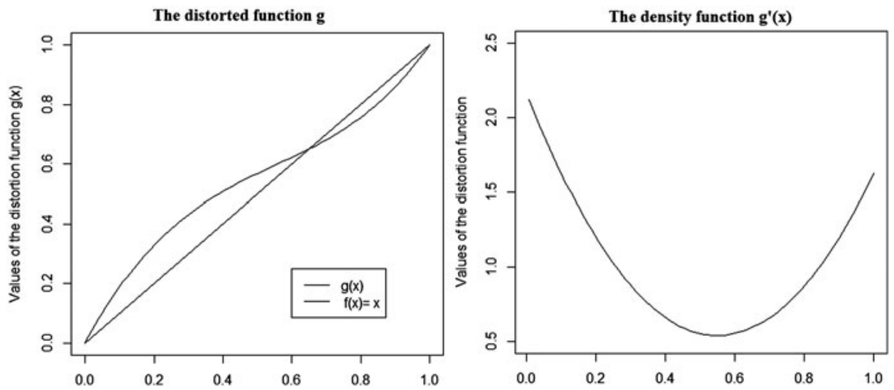


Fig. 7 The density g' associate to the distortion function g_δ with $\delta = 0.75$ which illustrate the fact that the effect of the saddle point discriminate the middle part of the quantile and put all the weight in the tail part

In order to associate a risk measure for such distorted function, we can remark that in all our examples we have $g(x) \geqslant t$ for all $x \in [0, 1]$ and then $\rho_g(X) \geqslant \mathbb{E}[X]$. This property characterizes the risk adverse behavior of managers. Nevertheless this last property does not guarantee the coherence of the risk measure ρ_g introduced in (2). Indeed, the function g_δ used to obtain these results can be convex and concave. In order to have a sub-additive risk measure and then to get coherence we propose to define a new risk measure in the following way:

$$\rho(X) = \mathbb{E}_g \left[F_X^{-1}(X) | F_X^{-1}(X) > F_X^{-1}(\delta) \right]. \tag{6}$$

It is a well defined measure, similar to the expected shortfall but computed under the distribution $g \otimes F_X$. Moreover it verifies the coherence axiom. With this new

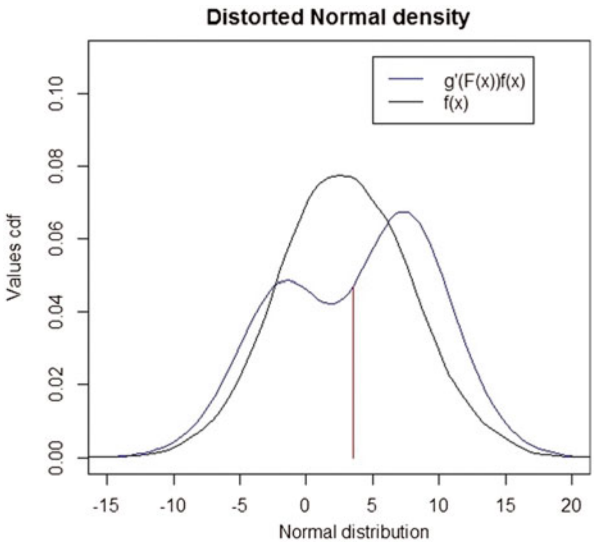


Fig. 8 Distortion of the Gaussian distribution using the function g introduced in Eq. (5) with $\delta = 0.75$ and $\beta = 0.48$. This picture exhibits a bimodal distribution due to the effect of the saddle point

measure we resolve our concern to define a risk measure that takes into account the information in the tails.

To create a multi-modal distribution with more than one hump, we can use a polynomial g of higher degree to have more saddle points in the interval $[0, 1]$. This is important if we seek to model distributions with multiple humps to represent multiple behaviors. For example we can consider a polynomial of degree 5 and 2 saddle points in the interval $[0, 1]$:

$$\begin{aligned} g(x) = & a_0(a_1^2a_3^2\frac{x^5}{5} + a_1^2a_4\frac{x^3}{3} + a_2^2a_3\frac{x^3}{3} + a_2^2a_4^2x - 2a_1^2a_3a_4\frac{x^4}{4} - 2a_1a_2a_3^2\frac{x^4}{4} \\ & + 4a_1a_2a_3a_4\frac{x^3}{3} - 2a_1a_2a_4^2\frac{x^2}{2} - 2a_2^2a_3a_4\frac{x^2}{2}) \end{aligned}$$

with first and second derivatives:

$$\begin{aligned} g'(x) = & a_0(a_1x - a_2)^2(a_3x - a_4)^2 = a_0(a_1a_3x^2 - a_1a_4x - a_2a_3x + a_2a_4)^2 \\ = & a_0(a_1^2a_3^2x^4 + a_1^2a_4x^2 + a_2^2a_3x^2 + a_2^2a_4^2 - 2a_1^2a_3a_4x^3 - 2a_1a_2a_3^2x^3 \\ & + 4a_1a_2a_3a_4x^2 - 2a_1a_2a_4^2x - 2a_2^2a_3a_4x), \\ g''(x) = & 2a_0a_1(a_1x - a_2)(a_3x - a_4)^2 + 2a_0a_3(a_1x - a_2)^2(a_3x - a_4). \end{aligned}$$

This function satisfies all the properties of a distortion function and can be used to generate a trimodal distribution under the condition that:

1. $a_i > 0$ for all $i \in \{1, 2, 3, 4\}$,
2. $\delta_1 = \dfrac{a_2}{a_1}$ and $\delta_2 = \dfrac{a_4}{a_3}$.

As we can see, the number of parameters increases as the number of saddle points increases.

Table 3 Summary statistics of the daily returns of $S\&P500$

Statistics	Mean	Variance	Std. Dev.	Skewness	Kurtosis
Return $(r_t)_t$	0.000016	0.299164	0.546959	0.030772	97.958431

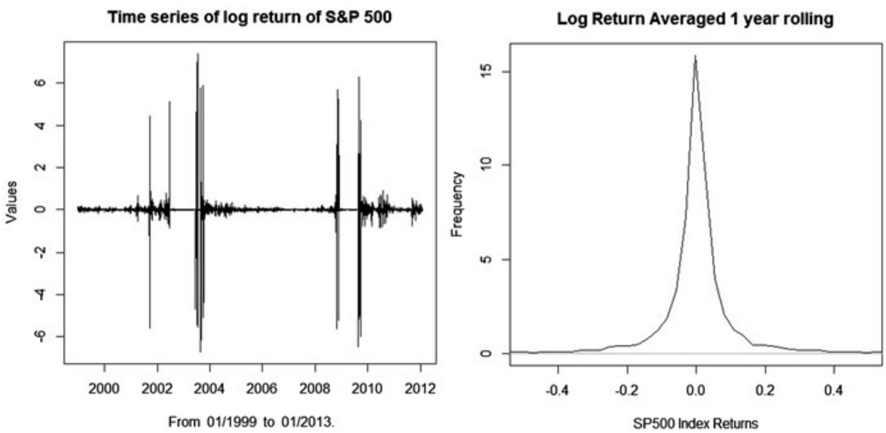


Fig. 9 The $S\&P500$ return index over time and the density

4 The Risk Measurement Using Distortion Measures

In this section distortion risk measures are applied to daily log-returns computed on the $S\&P500$ index collected from 01/01/1999 to 31/12/2011. This sample contains 3270 data points. Table 3 provides the empirical statistics of the data sets. This distribution is right skewed, most values are concentrated on the left of the mean, and some extreme values have been identified in the right tail. The distribution is leptokurtic ($Kurtosis > 3$) and sharper than a Gaussian distribution. Figure 9 exhibits the related time series and the empirical cumulative distribution.

Prior to applying any distortion, the underlying distribution has to be selected. For this exercise, consider the shape of returns as a Gaussian distribution. Then, considering Eq. (1), the empirical distribution is distorted successively using the Gini, the exponential and the Wang distortion functions while the polynomial distortion is calibrated to a Gaussian distribution.

To adjust the distortion, the following approach is implemented. First, the confidence level δ is set, then the parameter β is estimated using market information. In this paper, extreme value theory is used to estimate the kurtosis of the tail part associated to the losses. Then, its truncated kurtosis is divided by the kurtosis of the entire data set to evaluate the discrimination level β. Finally, the distortion using the function S inverse polynomial with the parameters δ and β can be applied.

Finally we focus on the properties of the resulting risk measures. We first compute the VaR, the ES and the spectral risk measure using both exponential and power

Table 4 Values of VaR, ES, exponential spectral risk measure and power spectral risk measure for different α. This table shows that the power spectral risk measure is not consistent with the concept of risk aversion

α	VaR	ES	Exp. spectral	Power spectral
0.90	0.095406	0.489994	2.210340	0.031422
0.91	0.107440	0.532119	2.226816	0.027705
0.92	0.117975	0.584883	2.243202	0.024120
0.93	0.136280	0.650673	2.259491	0.020690
0.94	0.151314	0.733222	2.275701	0.017383
0.95	0.188124	0.846702	2.291823	0.013892
0.96	0.220421	1.008289	2.307860	0.010814
0.97	0.273118	1.253847	2.323810	0.007881
0.98	0.371772	1.719202	2.339676	0.005319
0.99	0.605753	2.963937	2.355458	0.003081

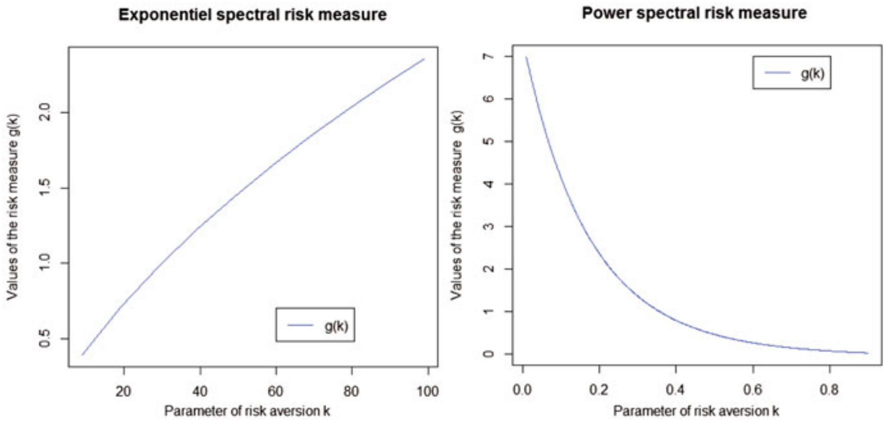


Fig. 10 This figure presents the level of risk with respect to the risk aversion parameters

spectrum functions using the original data set. These results are provided in Table 4 for different confidence level α. For both the VaR and the ES, the values of the risk measures increase with α. in this particular case, both the VaR and the ES are consistent risk measures. The spectral measures are provided first with exponential weights and second considering power weights. Looking at Table 4 fourth column, we note that the spectral power risk measures are not consistent[12] with the concept of risk aversion because they decrease with the level of confidence. Although, the value of the spectral exponential risk measures are consistent. In practice, this means that it makes no sense to use the power spectral measure. These two behaviors are presented in Fig. 10.

[12] $0.90 < 0.99$ but $0.031422 > 0.003081$.

Table 5 Values of distorted risk measures of the log returns of the $S\&P500$ using different distortion functions: Polynomial, Gini, exponential and Wang

Level	Saddle	Exp. distortion	Gini	Wang
0.90	1.695396	2.210329	0.096792	1.392441
0.95	2.156548	2.300160	1.105478	2.400308
0.99	2.978867	2.354956	1.108967	5.338054
0.995	3.275869	2.560075	2.106531	6.667918

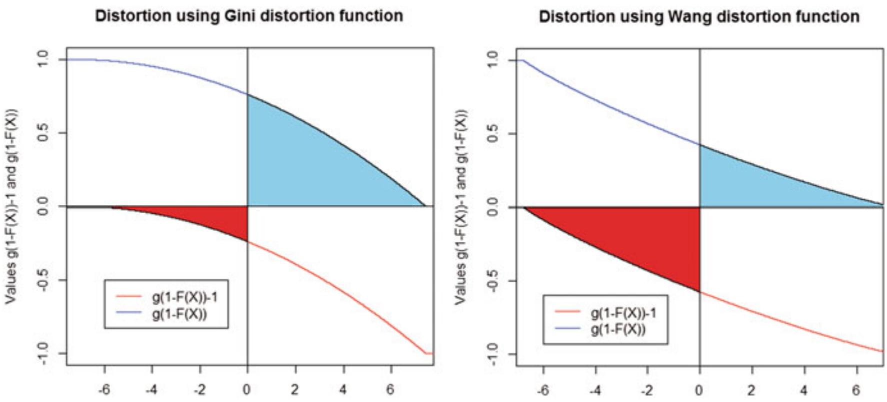


Fig. 11 On the *left graph*, the distortion using the Gini function is exhibited, and on the *right graph* the distortion implied by Wang. The *red line* corresponds to the negative part of the returns and the *blue line* of the positive part. Using Gini distortion the weights on the negative part are smaller than the weights obtained using Wang distortion. On the contrary, the weights on the positive part using Gini are larger than those implied by Wang on the same portion. The same function g is used to build the positive and the negative part of the distribution

In a second step, various distortion approaches presented previously (Polynomial, Gini, Exponential and Wang) are applied to the data, and the associated risk measures are computed using Eq. (1). The risk measures obtained for each of the four methodologies are given in Table 5. These measures are all more conservative than the empirical VaR which may be used as a benchmark. The impacts of the distortions using these functions are represented in Figs. 11 and 12.

In our methodologies, most of the distortion functions are symmetric while the underlying information is usually asymmetric. To address we propose to use two different functions for the losses and the gains. Figure 13 exhibits the Sereda et al. (2010) distortion function overcoming the symmetry issue.

Unfortunately, it is not sufficient to consider the same function with two different parameters. As observed in Fig. 13, distortions that are applied on both sides are convex which is not consistent with a risk aversion property. It is important to consider a different behaviour to analyse separately the losses and the gains. Indeed, if a convex distortion function is considered for the losses then a concave distortion function should be considered for the gains (Bhansali 2012).

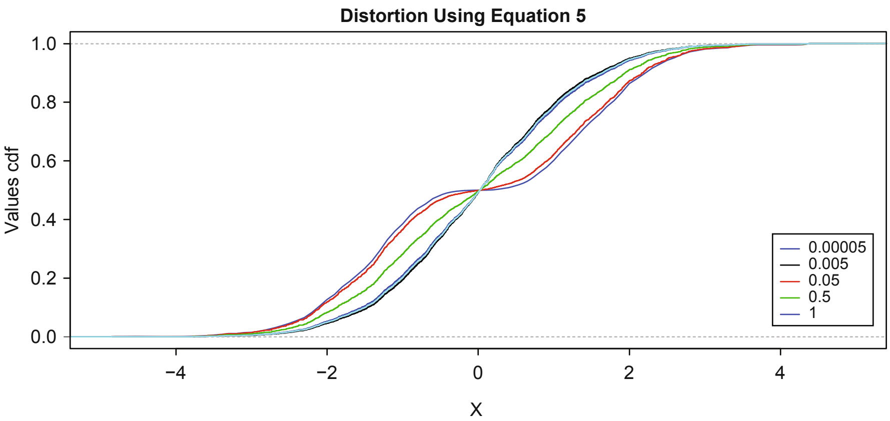


Fig. 12 On this figure we present the distorted empirical distribution of the returns using the function g introduced in (5) with $0.00005 \leq \beta \geq 1$ and $\delta = 0.5$

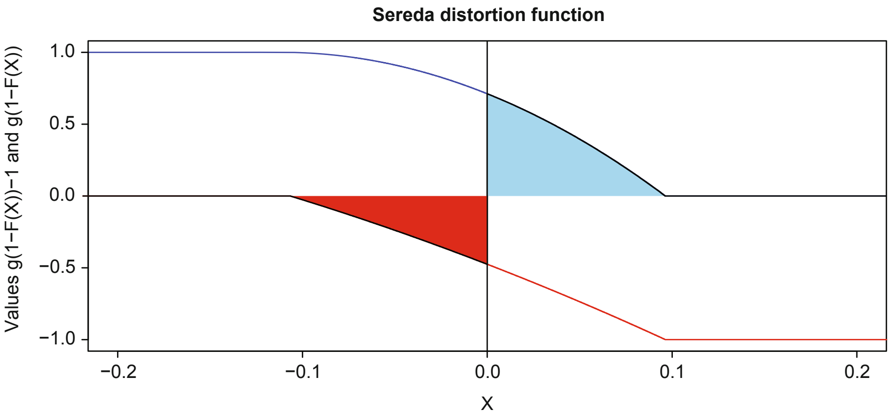


Fig. 13 In this picture we present two convex distortion functions in order to create asymmetry: The *red line* illustrates the function $g_2(S(X)) - 1$ and the *blue line* represents the function $g_1(S(X))$ where g is a convex distortion function. Two different functions g_1 and g_2 with two levels of confidence $k_1 = 0.95$ and $k_2 = 0.2$ are considered to get the positive part and the negative part of the distribution

5 Conclusion

This paper has summarized different notions of risk measures developed in the literature: the quantile based risk measure (VaR and ES), the spectral risk measure and distortion risk measure. This review has amplified the difficulty encountered in terms of financial regulation (as demanded by Bale III and Solvency II) using these risk measures. We recall that the VaR is not coherent while the ES cannot account for risk aversion (because it is risk neutral), and spectral risk measure depends on the choice of the weights limiting the quantification of risk. One alternative to this limitation is provided by distortion theory developed under convex function. This approach represents an appropriate way to consider and analyze risk because it is always possible to define a distortion function such as Wang's distortion function generated from a Gaussian distribution. The distortion risk measure provides an equivalent approach to measure the risk under the convex distortion function.

Nevertheless using the same convex function for upside and downside risks like in Wang (2000) imposes a similar approach for both parts. This represents a limitation for the use of convex distortion function to analyze upside and downside risks at the same time. Alternatively we can consider a distortion risk measure with a S-inverse function with a concave part and a convex part generating decreasing risk measure with respect to the confidence level.

In summary we have provided a general framework that combines expected shortfall of the quantile of the risk measure with S-inverse shaped distortion function. This new risk measure is coherent, satisfies all the axiom of risk measures and is consistent with risk aversion concept. A statistical methodology for the estimation of such measures remains yet to be developed.

References

Acerbi, C., and D. Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26 (7): 1487–1503.

Artzner, P., F. Delbaen, J.-M. Eber, and D Heath. 1999. Coherent risk measures. *Mathematical Finance* 9 (3): 203–228.

Bhansali, Vineer. 2012. Asset allocation and risk management in a bimodal world. Financial analysts seminar improving the investment decision-making process held in Chicago on 23–27 July 2012 in partnership with the CFA Society of Chicago.

Bookstaber, R., and R. Clarke. 1984. Option portfolio strategies: Measurement and evaluation. *Journal of Business* 57 (4): 469–492.

Barndorff-Nielson, O. E. 1977. Exponentially decreasing distribution for the logarithm of the particle size. *Proceedings of the Royal society, Series A: Mathematical and physical sciences* 350:401–419.

Chamberlain, G. 1983. A characterization of the distribution that imply mean-variance utility functions. *Journal of Economic Theory* 29:185–201.

Cotter, J., and K. Dowd. 2006. Extreme spectral risk measures: An application to futures clearinghouse margin requirements. *Journal of Banking and finance* 30:3469–3485.

Danielson, J., B. N. Jorgensen, M. Sarma, G. Samorodnitsky, and C. G. de Vries. 2005. Subadditivity re-examined: The case for value at risk. www.RiskResearch.org. (Unfinished working paper).

De Giorgi, E. 2005. Reward risk portfolio selection and stochastic dominance. *Journal of Banking and Finance* 29 (4): 895–926.

Delbaen, F. 2000. Coherent risk measures. Lecture notes of the Cattedra Galileiana at the Scuola Normale di Pisa, March 2000.

Denuit, M., J. Dhaene, and M. J. Goovaerts. 2006. *Actuarial theory for dependent risks*. New York: Wiley.

Dowd, K., G. Sorwar, and J. Cotter. 2008. Estimating power spectral risk measures. Mimeo. Nottingham University Business School.

Embrechts, P., R. Frey, and A. McNeil. 2005. *Quantitative risk management: Concepts, techniques and tools*. Princeton: Princeton University Press.

Föllmer, H., and A. Schied. 2004. *Stochastic finance: An introduction in discrete time* (2nd revised and extended edition). Berlin: Walter de Gruyter & Co. (de Gruyter Studies in Mathematics 27).

Hardy, M. R., and J. Wirch. 2001. Distortion risk measures: coherence and stochastic dominance. Working Paper, University of Navarra.

Huang, C.F., and R. H. Litzenberger. 1988. *Foundations for financial economics*. Englewood Cliffs: Prentice Hall.

Härlimann, W. 2004. Distortion risk measures and economic capital. *North American Actuarial Journal* 8 (1): 86–95.

Ingersoll, J. E. Jr. 1987. *Theory of financial decision making*. Totowa: Rowman and Littlefield.

Inui, K., and M. Kijima. 2005. On the significance of expected shortfall as a coherent risk measures. *Journal of Banking and Finance* 29:853–864.

Morgan, J. P. 1995. *RiskMetrics technical document* (4th ed.). NewYork: Morgan Guaranty Trust Company. http://www.riskmetrics.com/rm.html

Kahneman, D., & A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:313–327.

Markowitz, H. M. 1959. *Portfolio selection: Efficient diversification of Investments*. New York: Wiley (Oxford: Basil Blackwell, Yale University Press 1970, 1991).

Rockafellar, R. T., and S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2 (3): 21–41.

Rockafellar, R. T., and S. Uryasev. 2002. Conditional value at risk for general loss distributions. *Journal of Banking and Finance* 26 (7): 1443–1471.

Rockafellar, R. T., S. Uryasev, and M. Zabarankin. 2006. Optimality conditions in portfolio analysis with generalized deviation measures. *Mathematical Programming* 108 (2–3): 515–540.

Samorodnitsky, G., and M. S. Taqqu. 1994. *Stable non-Gaussian random processes: Stochastic models with infinite variance*. New York: Chapman and Hall.

Sereda, E. N., E. M. Bronshtein, S. T. Rachev, F. J. Fabozzi, W. Sun, and S. V. Stoyanov. 2010. Handbook of portfolio construction, Volume 3 of Business and economics, Chapter Distortion risk measures in portfolio optimisation, 649–673. Springer US.

Tasche, D. 2002. Expected shortfall and beyond. *Journal of Banking and Finance* 26 (7): 1519–1533.

Tsanakas, A. 2004. Dynamic capital allocation with distortion risk measures. *Insurance: Mathematics and Economics* 35:223–243.

Wang, S. S., V. R. Young, and H. H. Panjer. 1997. Axiomatic characterisation of insurance prices. *Insurance: Mathematics and Economics* 21:173–183.

Wang, S. S. 2000. A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* 67 (1): 15–36.

Yaari, M.-E. 1987. The dual theory of choice under risk. *Econometrica* 55 (1): 95–115.

Stress Testing Engineering: The Real Risk Measurement?

Dominique Guégan and Bertrand K. Hassani

1 Introduction

Stress testing (Berkowitz 1999; Quagliariello 2009; Siddique and Hasan 2013) is used to determine the stability or the resilience of a given financial institution by deliberately submitting the subject to intense and particularly adverse conditions which has not been considered a priori. This involves testing beyond the traditional capabilities-usually to determine limits-to confirm that intended specifications are both accurate and being met in order to understand the process underlying failures. This exercise does not mean that the entity's failure is imminent, though its purpose is to address and prepare this potential failure. Consequently the stress testing is the quintessence of risk management.

Since the 1990's most financial institutions have conducted stress testing exercises on their balance sheet, but it is only in 2007 following the current crisis, that regulatory institutions became interested in analyzing and measuring the resilience of financial institutions[1] in case of dramatic movements of economic fundamentals such as the GDP. Then, stress tests have been regularly performed by regulators to insure that banks are properly adopting practices and strategies which decrease the chance of a bank fails and jeopardises the entire economy (Berkowitz 1999).

[1] In October 2012, U.S. regulators unveiled new rules expanding this practice by requiring the largest American banks to undergo stress tests twice per year, once internally and once conducted by the regulators.

D. Guégan (✉)
University Paris 1 Panthéon-Sorbonne et New York University Polytechnic School of Engineering, Brooklyn, New York, USA
e-mail: dguegan@univ-paris1.fr

CES UMR 8174, 106 boulevard de l'Hopital, 75647 Paris Cedex 13, France Tel.:+33144078298

B. K. Hassani
Université Paris 1 Panthéon-Sorbonne CES UMR 8174 and Grupo Santander,
106 boulevard de l'Hopital, 75647 Paris Cedex 13, France Tel.:+44(0)2070860973
e-mail: bertrand.hassani@santander.co.uk.

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,*
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_3,

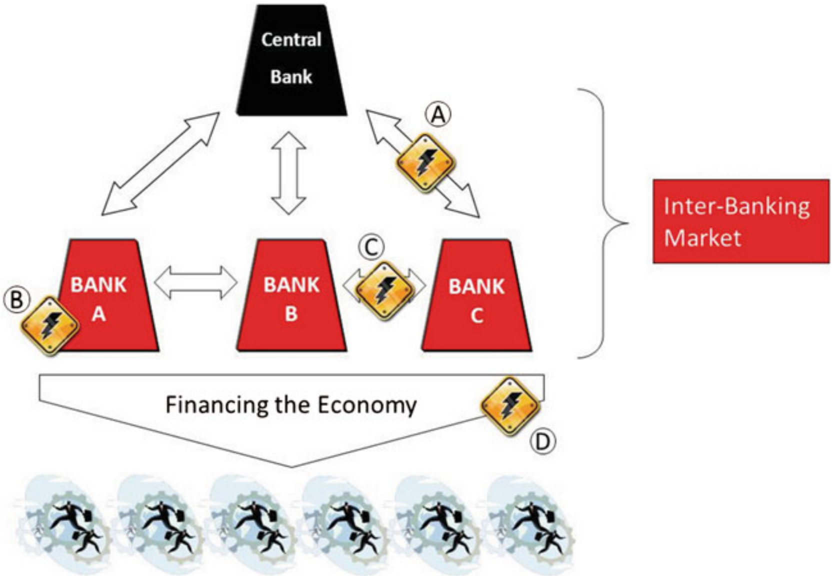


Fig. 1 Financial system

Originally governments and regulators sought to measure financial institutions' resilience-and by extension the entire financial system-in order to avoid future failures, ultimately assumed by the tax payer. Stress testing framework raises the following questions: if a risk is identified even with a small probability why is it not directly integrated in bank's risk models? Why should we have two processes addressing the same risks, first under "normal" conditions and the other under stressed condition knowing that the risk universe includes them both? Therefore, are stress tests discussing in fact the impact of an exogenous event on a balance sheet or of an endogenous failing process. For example, when a model is shown unable to account for certain risks. In other words are risks of failure and threats to the entire system only exogenous?

Assuming that our objective is the protection of the system, a clear definition of the latter is necessary. Figure 1 exhibits a simplified version of the banking system used in most developed countries. It consists of three layers, the "real" economy, the interbanking intermediate market and the central bank. Based on this structure if we want to measure the system's resilience we need to identify what threatens the system. We may reasonably assume that stress-testing and systemic-risk measurement are necessarily connected. As will be discussed below a bank may either fail because of a lack of capital or a lack of liquidity. The risk of illiquidity may be particularly challenging while a bank balance sheet is fairly simple, in fact assets are composed of intangible assets, investments and loans while liabilities are mostly shareholder's equity and subordinated debt. Subsequently comes the wholesale funding and finally clients deposits. Thus by changing the money on short duration, such as savings with money on a longer one, bank's lending exercises a maturity transformation. This leads banks having an unfavorable liquidity position since they have no access to the money they lent while the money they owe to clients can be withdrawn at any time on demand.

Regarding our simplistic representation of the financial system, Fig. 1 exhibits four potential failure spots: (i) The central bank cannot provide any liquidity anymore; (ii) Banks are not funding each other anymore; (iii) A bank of sensible size is failing; (iv) The banks stop financing the economy.

Regarding the first point (i), considering that the purpose of a central bank is refinancing loans provided by commercial banks controlling inflation, unemployment rate, etc., as soon as this central bank ceases functioning properly the entire system collapses (uncontrolled inflation, etc.). This point discusses a model failure and falls beyond the stress-testing exercise. Some examples may be found in the Argentinian or the Russian central banks failure in the 1990's. Considering the second point (ii), if banks are not funding each other then some will face a critical lack of liquidity inter bank trust and may fail (for instance this was the case when banks refused to help Lehman Brothers the 15th of September 2007). This example highlights how financial institutions face liquidity risk. Third point (iii), a bank may fail due to a lack of capital. Consequently the problem arises to evaluate ex-ante its capital requirements when a major risk occurs. In 2008 the insurance company AIG failed due to a lack of capital when most of the CDS they sold had been triggered simultaneously. This is the care of risk management intra-bank. Finally (iv), and assuming that the purpose of banks is to generate a profit while financing the real economy it is highly unlikely that they would do it in case of high stress. A lack of liquidity may lead a bank to stop lending and may engender corporate and retail defaults. Contagion defaults in chain may themselves drive the financial institution to bankruptcy. In that latter case we are clearly discussing the way banks are addressing the credit risk.

Whatever the origin of the stress inside the banks a failure causes important damages: bankruptcy alters the markets confidence. Thus global system disrupted. Therefore stress tests are used to imagine the scenarii illustrating such situations and introduce dynamical and adaptive solutions that can avoid dramatic failures.

In this paper we focus on what may lead banks to fail and how financial resilience can be measured. A system may trigger a chain reaction by measuring its resilience and by extension the systemic resilience. The proper question is what may trigger contagion. Following (Lorenz 2010) we have to capture the butterfly which may engender a twister? Two families of triggers are analyzed: first is based on the impact of external (and/or extreme) events, the second is based on the impacts of inadequate models for predictions or risks measurement: explicitly models defaulting over time due to their non adaptivity to dynamical environmental changes. The first trigger needs to take into account fundamental macro-economic data or massive operational risks while the second trigger deals with the limitations of the quantitative models for forecasting, pricing, evaluating capital or managing the risks. A model's example of system failure was the use of the Gaussian copula to price CDOs (and CDS), mis-capturing the intrinsic upper tail dependencies characterizing CDOs tranches correlations. These ledding to mis-pricing and mis-hedging positions and in fact producing the experienced financial breakdowns. It may be argued that if inside the banks-limitations, pitfalls and other drawbacks of models used were correctly identified, understood and handled, and if the associated products were correctly known, priced and insured, then the effects of the crisis may not have been as important.

This paper is structured as follows. In Section two the stress testing framework is presented. In the third Section the mathematical tools required to develop the stress testing procedures are introduced. Finally, in a fourth section we discuss and illustrate integration of the stress-testing strategy directly into the risk models.

2 The Stress-Testing Framework

A stress-testing exercise means choosing scenarii that are costly and rare which can lead to the financial institution failure, To do so, we integrate these scenarii in a model to measure their impact. The integration process may be a simple linear increase of parameters to augment the outcomes' confidence interval, or switch to more advanced models predicting the potential loss due to an extreme event or a succession of extreme events. To do so we implement various methodologies allowing the capture of multiple behaviours, or adding exogenous variables.

The objective of this exercise is to strengthen management's rate assessment to better understanding extreme exposures, i.e. exposure that may fall beyond the "business as usual" domain of a model. We define the capture domain of a model by its capability to be resilient to the occurrence of an extreme event, i.e. the relevant risk measure would not fluctuate or would do so only in a narrow range of values, or would not breach the selected confidence intervals too often.

When a financial institution proceeds to stress tests, its accept that its exposure only up to a certain extent. This is due to incomplete data sets used to calibrate the models. The information set is partial and is not adapted to the real economy. In fact the data sets contain only past incidents, and if crisis accounted they do not integrate futur extreme events which are by definition unknown. In other words, even if models are conservative enough to consider eventual Black Swans (Taleb 2010), stress-testing provides a greater awareness of Black Swans "with blue eyes and white teeth".

Selecting the appropriate scenario is equivalent to selecting the factors that may have an impact on the models, (e.g. covariates) and to define the stress' level. These scenarii are planned to characterize shocks likely to occur more than what historical observations would indicate: shocks that have never occurred (stress expected loss), shocks reflecting circumstantial break downs, shocks reflecting future structural breaks. Mathematically all new shocks' categories entail drawing some new factor distribution f^* which is not equal to the original distribution f that characterizes the original data set. Every type of shock has to include correlations, co-movements and specific events, such as crash, bankruptcy, systemic failure, etc.

When scenarii are assessed practitioners have to assess their various outcomes. Are they relevant to the goal of stress testing? Are they internally consistent? Are they archetypal? Do they represent relatively stable outcome situations? Risk managers would identify extremes outcomes and their causes and assess their consistency and plausibility. Three key points may be addressed:

1. Time frame: are "new" trends compatible within the time frame in question?
2. Internal consistency: do sources of uncertainty account in probable scenarii?
3. Stakeholder influence: Are scenarii reliable, when considering the potential negative effect to shareholders?

In fact, there are many risk sources to be taken into account inside stress tests. We enumerate specific risks for which particular attention ought to be given, and constantly new information used to continuously update and the likelihood of potential severe outcomes reflected in banks' internal assessments. These are mainly market, credit, operational and liquidity risks.

Some procedures have traditionally been applied to banks' market risks defining trading portfolios by considering multiple states of nature scenarii (some unlikely) impacting various risk factors. Traditionally three kinds of approaches are used, standard, historical and worst-case scenarii. This approach to stress testing is probably too simple and therefore may be incomplete due to over simplicity.

The credit risk stress testing has been integrated into the capital calculations formulas through the Loss Given Default distribution. The credit risk stress testing concerns other domains that the capital allocation. For example Majnoni et al. (2001) linked the ratio non-performing loan over total assets to several fundamentals macroeconomic variables such as nominal interest rates, inflation, GDP, etc.; Bunn et al. (2005) measured the impact of aggregate default rates and LGD evaluation of aggregated write-offs for corporate mortgages and unsecured lendings using standard macroeconomic factors like GDP, unemployment, interest rates, income gearing and loan to value ratios. This last component may be particularly judicious in the UK considering the level of interest rate for mortgages sold in the past few years. Practitioners (Pesola 2007) argue that unexpected shocks should drive loans related losses and the state of system, i.e. a more fragile system would worsen the losses. Therefore factors weakening a financial system should interact in a multiplicative way. Counterparty credit exposure may either be represented by the "current" exposure, the "expected" exposure or the "expected positive" exposures. Stressing the exposure distributions would naturally impact the measures based on them, for instance the Credit Value Adjustement (CVA) (Gregory 2012) via the expected exposure or the expected loss via the expected positive exposure.

Operational risks are stressed through extreme scenarii. Following the demand of the regulator, the implemented methodologies are often conservative providing risk measures larger than what empirical data or traditional approach would give to practitioners. However these are not sufficient studies to provide an accurate representation of these risks over time. Alternative strategies need to be developed such as those presented in the next section.

Liquidity risk arises from situations in which an entity interested in trading an asset cannot do it because nobody wants to buy or sell it with respect to the market conditions. Liquidity risk becomes particularly important to entities which currently hold an asset (or want to held it) since it affects their capability to trade. Manifestations of liquidity risk are very different if it comes from price dropping to zero. In case an asset's price falling to zero the market is saying that the asset is worthless. However if one bank cannot find a counterparty interested in trading the asset this may only be a problem of market equilibrium, i.e. the participants have trouble finding each other. This is why liquidity risk is usually found to be higher in emerging or low-volume markets. Accordingly liquidity risk has to be managed in addition to market, credit and operational risks. Because of its tendency to compound other exposures it is difficult or impossible to isolate liquidity risk. Some ALM techniques can be applied to assessing liquidity risk. A simple test for liquidity risk is to look at future net cash flows on a day-by-day basis where any day that has a sizeable negative net cash flow is of concern. Such an analysis can be supplemented with stress testing.

In this paper we only partially deal with the "micro" liquidity risk, i.e. the liquidity of an asset, by opposition to the "macro" liquidity exposure, i.e. of a financial institution which is an aggregated measure assuming that it is included in market prices dropping up to a certain extent which is captured in the market risk measurement. The liquidity position of a financial institution is measured by the quantity of assets to be sold immediately to face the liquidity requirements, even considering a haircut, while the price of an asset on the market is illiquid if there is no demand and its price is actually equal to 0, and consequently the measure should be forward. As a result considering the previous statement-in the next section-we focus on methodologies to measure and stress the solvency of financial institutions in relation to market, credit and operational risks.

3 Tools

In this section some of the tools required to develop stress test strategies are introduced. The main ingredient which is determinant in our point of view is the information set on which our work relies. This set is definitively determinant and whatever the methodologies we will use latter the conclusions cannot be done without referring to this information set. After discussing the role of the data set we briefly recall the measures of risks which can be used and how they can be computed. This leads us in a first step to introduce the distributions appearing relevant to obtained realistic risks measures (from an univariate point of view) and second the notion of dependence permitting to capture interdependences between the risks in order to properly evaluate their risk measures in a multivariate framework. Finally the question of dynamics that should be captured in all strategies of risk management is also discussed.

An *a priori* which is important to note is that financial data sets are always formed with discrete time data and they cannot be directly associated to continuous data thus in this chapter the techniques we present are adapted to this kind of data set.

3.1 Data Mining

Feeding the scenario analysis and evaluating the potential outcomes lie on the quality of the information set used. Therefore a data mining process should be undertaken. Data mining is the computational process of discovering patterns in large data sets. The objective process is to extract information from a data set, make it understandable and prepare it for further use. In our case the data mining step is equivalent to a pre-processing exercise. Data mining involves almost six common classes of tasks (Fayyad et al. 1996):

(a) Anomaly detection–This is the search of events which do not conform to expected patterns. These anomalies are often translated into actionable information.
(b) Dependency analysis–The objective is to detect interesting relationships between variables in databases.
(c) Clustering–Cluster analysis consists in the task of creating groupings of similar objects.
(d) Classification–The classification consists in generalizing known structure to be applicable to new data sets.
(e) Regression & fittings–This statistical process purpose is to find the appropriate model, distribution or function which represent the "best" (in a certain sense) fit for the data set.
(f) Summarisation–This step purpose is to provide a synthetic representation of the data set.

These classical techniques are recalled as they could be interesting in certain cases to determine the sets or subsets on which we should work. They could also be useful to exhibit the main features of the data before beginning a probabilistic analysis (for the risk measure) or doing a time series analysis to examine the dynamics which seems the more appropriate for stress testing purposes.

3.2 Risk Measures

Even if at the beginning the risk in the banks was evaluated using the standard deviation applied to different portfolios the financial industry uses now the quantile-based downside risk measures including the Value-at-Risk (VaR_α for confidence level α) and Expected Shortfall. The VaR_α measures the losses that may be expected for a given probability, and corresponds to the quantile of the distribution which characterizes the asset or the type of events for which the risk has to be measured. Thus, the fit of an adequate distribution to the risk factor is definitively an important task to obtain a reliable value of the risk. Then, in order to measure the importance of the losses beyond the VaR percentile and to capture the diversification benefits the expected shortfall measure has been introduced.

The definitions of these two risks measures are recalled below:

Definition 3.1 *Given a confidence level* $\alpha \in (0, 1)$*, the* VaR_α *is the relevant quantile*[2] *of the loss distribution,* $VaR_\alpha(X) = \inf\{x \mid P[X > x] \leqslant 1 - \alpha\} = \inf\{x \mid F_X(x) \geqslant \alpha\}$ *where* X *is a risk factor admitting a loss distribution* F_X .

Definition 3.2 *The Expected Shortfall (*ES_α*) is defined as the average of all losses which are equal or greater than* VaR_α*:*

$$ES_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\alpha dp$$

The Value at Risk initially used to measure financial institutions market risk was popularised by Riskmetrics (1993). This measure indicates the maximum probable loss given a confidence level and a time horizon. The VaR_α is sometimes referred as the "unexpected" loss. The expected shortfall has a number of advantages over the VaR_α because it takes into account the tail risk and fulfills the sub-additive property. It has been widely dealt with in the literature, for instance in Artzner et al. (1999) and Rockafellar and Uryasev (2000, 2002). Relationships between VaR_α and ES_α, for some distributions can be found in this book inside the chapter of Guégan et al. (2014).

Nevertheless even if the regulators asked to the banks to use the VaR_α and recently the ES_α to measure their risks and ultimately provide the capital requirements to avoid bankruptcy these risk measures are not entirely satisfactory:

- They provide a risk measure for an α which is too restrictive considering the risk associated to the various financial products;
- The fit of the distribution functions can be complex or inadequate in particular for the practitioners who want to follow the guidelines proposed by the regulators (Basel II/III guidelines). Indeed, in case of the operational risks the suggestions is to fit a GPD which does not correspond very often to a good fit and whose carrying out can be difficult.
- It may be quite challenging to capture extreme events. Taking into account these events in modelling the tails of the distributions is determinant.
- Finally all the risks are computed considering unimodal distributions which can be non realistic in practice.

Recently several extensions have been analysed to overcome these limitations and to propose new routes for the risk measures. These new techniques are briefly recalled and we suggest the reader to look at the chapter of Guégan et al. (2014) in this book for more details, developments and applications:

- Following our proposal we suggest the practitioners to use several α to obtain a spectrum of their expected shortfall and to visualize the evolution of the ES with respect to these different values. Then, a unique measure can be provided

[2] $VaR_\alpha(X) = q_{1-\alpha} = F_X^{-1}(\alpha)$.

making a convex combination of these different ES with appropriate weights. This measure is called spectral measure (Acerbi and Tasche 2002).

- In the univariate approach if we want to take into account information contained in the tails we cannot restrict to the GPD as suggested in the guidelines provided by the regulators. There exist other classes of distributions which are very interesting, for instance the generalized hyperbolic distribution (Barndorff-Nielsen and Halgreen 1977), the extreme value distributions including the Gumbel, the Frechet and the Weibull distributions (Leadbetter 1983), the α-stable distributions (Taqqu and Samorodnisky 1994) or the g-and-h distributions (Huggenberger and Klett 2009) among others.
- Nevertheless the previous distributions are not always sufficient to properly fit the information in the tails and another approach could be to build new distributions shifting the original distribution on the right or left parts in order to take a different information in the tails. Wang (2000) proposes such a transformation of the initial distribution which provides a new symmetrical distribution. Sereda et al. (2010) extend this approach to distinguish the right and left part of the distribution taking into account more extreme events. The function applied to the initial distribution for shifting is called a distortion function. This idea is ingenious as the information in the tails is captured in a different way that using the previous classes of distributions.
- Nevertheless when the distribution is shifted with a function close to the Gaussian one as in Wang (2000) and Sereda et al. (2010) the shift distribution remains unimodal. Thus we propose to distort the initial distribution with polynomials of odd degree in order to create several humps in the distributions. This permits to catch all the information in the extremes of the distributions, and to introduce a new coherent risk measure $\rho(X)$ computed under the $g \otimes F_X$ distribution where g is the distortion operator and F_X the initial distribution, thus we get:

$$\rho(X) = \mathbb{E}_g \left[F_X^{-1}(x) | F_X^{-1}(x) > F_X^{-1}(\delta) \right]. \tag{1}$$

All these previous risk measures can be included within a stress testing strategy.

3.3 Univariate Distributions

This section proposes several alternatives for the fitting of a proper distribution to the information set related to a risk (losses, scenarios, etc.). The knowledge of the distributions which characterises each risk factor is determinant for the computation of the associated measures and will be also determinant in the case of a stress test. The elliptical domain needs to be left aside to consider distributions which are asymmetric and leptokurtic like the Generalized Hyperbolic distributions, Generalized Pareto distributions, or Extreme Value Distributions among others. Their expressions are recalled in the following.

The Generalized Hyperbolic Distribution (GHD) is a continuous probability distribution defined as a mixture of an inverse Gaussian distribution and a normal distribution. The density function associated to the GHD is:

$$f(x,\theta) = \frac{(\gamma/\delta)^{\lambda}}{\sqrt{2\pi} K_{\lambda}(\delta\gamma)} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu^2)}/\alpha\right)^{1/2-\lambda}}, \tag{2}$$

with $0 \leq |\beta| < \alpha$. This class of distributions is very interesting as it relies on five parameters. If the shape parameter λ is fixed then several well known distributions can be distinguished:

(a) $\lambda = 1$: hyperbolic distribution
(b) $\lambda = -1/2$: NIG distribution
(c) $\lambda = 1$ and $\xi \to 0$: Normal distribution
(d) $\lambda = 1$ and $\xi \to 1$: Symmetric and asymmetric Laplace distribution
(e) $\lambda = 1$ and $\chi \to \pm\xi$: Inverse Gaussian distribution
(f) $\lambda = 1$ and $|\chi| \to 1$: Exponential distribution
(g) $-\infty < \lambda < -2$: Asymmetric Student
(h) $-\infty < \lambda < -2$ and $\beta = 0$: Symmetric Student
(i) $\gamma = 0$ and $0 < \lambda < \infty$: Asymmetric Normal Gamma distribution

The four other parameters can be then associated to the first four moments permitting a very good fit of the distributions to the corresponding losses.

Another class of distributions is the Extreme Value Distributions built on sequences of maxima obtained from the initial data sets. To introduce this class, the famous Fisher-Tippett theorem needs to be recalled:

Theorem 3.1 *Let be (X_n)a sequence of i.i.d.r.v. If it exists constants $c_n > 0$, $d_n \in \mathbb{R}$ and a non degenerated distribution function G_α such that*

$$c_n^{-1}(M_n - d_n) \xrightarrow{\mathcal{L}} G_\alpha \tag{3}$$

then G_α is equal to:

$$G_\alpha(x) = \begin{cases} \exp\left(-(1+\alpha x)^{-\frac{1}{\alpha}}\right) & \alpha \neq 0,\ 1+\alpha x > 0 \\ \exp\left(-e^{-x}\right) & \alpha = 0,\ x \in \mathbb{R} \end{cases}$$

This function $G_\alpha(x)$ contains several classes of extreme values distributions:

$$.\text{Fréchet (type III)}: \Phi_\alpha(x) = G_{1/\alpha}\left(\frac{x-1}{1/\alpha}\right) = \begin{cases} 0 & x \leq 0 \\ \exp\left(-x^{-\alpha}\right), & x > 0,\ \alpha > 0 \end{cases}$$

$$.\text{Weibull (Type II)}: \Psi_\alpha(x) = G_{-1/\alpha}\left(\frac{x+1}{1/\alpha}\right) = \begin{cases} \exp\{-(-x^{\alpha})\}, & x \leq 0,\ \alpha > 0 \\ 1 \end{cases}$$

$$.\text{Gumbel (Type I)}: \Lambda(x) = G_0(x) = \exp\left(-e^{-x}\right),\ x \in \mathbb{R}.$$

Considering only the maxima of a data set is an alternative to model in a robust way the impact of the extremes of a series within a stress testing strategy.

Another class of distributions permitting to model the extremes is the distribution built on a data set defined above or under a threshold. Let X a r.v. with distribution function F and right end point x_F and a fixed $u < x_F$. Then,

$$F_u(x) = P[X - u \leq x | X > u], \quad x \geq 0,$$

is the excess distribution function of the r.v. X (with the df F) over the threshold u, and the function

$$e(u) = E[X - u | X > u]$$

is called the mean excess function of X which can play a fundamental role in risk management. The limit of the excess distribution has the distribution G_ξ defined by:

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}} & \xi \neq 0, \\ 1 - e^{-x} & \xi = 0, \end{cases} .$$

where,

$$\begin{array}{ll} x \geq 0 & \xi \geq 0, \\ 0 \leq x \leq -\frac{1}{\xi} & \xi < 0, \end{array} .$$

The function $G_\xi(x)$ is the standard Generalized Pareto Distribution. One can introduce the related location-scale family $G_{\xi,\nu,\beta}(x)$ by replacing the argument x by $(x - \nu)/\beta$ for $\nu \in \mathbb{R}$, $\beta > 0$. The support has to be adjusted accordingly. We refer to $G_{\xi,\nu,\beta}(x)$ as GPD.

Another class of distributions is the class of α-stable distributions defined through their charateristic function also relying on several parameters. For $0 < \alpha \leq 2$, $\sigma > 0$, $\beta \in [-1, 1]$ and $\mu \in R^+$, $S_\alpha(\sigma, \beta, \mu)$ denotes the stable distribution with the characteristic exponent (index of stability) α, the scale parameter σ, the symmetric index (skewness parameter) β and the location parameter μ. $S_\alpha(\sigma, \beta, \mu)$ is the distribution of a r.v. X with characteristic function,

$$E[e^{ixX}] = \begin{cases} exp(i\mu x - \sigma^\alpha |x|^\alpha (1 - i\beta sign(x) tan(\pi\alpha/2))) & \alpha \neq 1, \\ exp(i\mu x - \sigma |x| (1 + (2/\pi) i\beta sign(x) ln|x|)) & \alpha = 1 \end{cases} ,$$

where $x \in R$, $i^2 = -1$, $sign(x)$ is the sign of x defined by $sign(x) = 1$ if $x > 0$, $sign(0) = 0$ and $sign(x) = -1$ otherwise. A closed form expression for the density $f(x)$ of the distribution $S_\alpha(\sigma, \beta, \mu)$ is available in the following cases: $\alpha = 2$ (Gaussian distribution), $\alpha = 1$ and $\beta = 0$ (Cauchy distribution) and $\alpha = 1/2$ and $\beta = +/-1$ (Levy distributions). The index of stability α characterises the heaviness of the stable distribution $S_\alpha(\sigma, \beta, \mu)$.

Finally we introduce the g-and-h random variable $X_{g,h}$ obtained transforming the standard normal random variable with the transformation function $T_{g,h}$:

$$T_{g,h}(y) = \begin{cases} \frac{exp(gy) - 1}{g} exp\left(\frac{hy^2}{2}\right) & g \neq 0, \\ y exp\left(\frac{hy^2}{2}\right) & g = 0, \end{cases} .$$

Thus

$$X_{g,h} = T_{g,h}(Y), \text{when} \;\; Y \sim N(0,1).$$

This transformation allows for asymmetry and heavy tails. The parameter g determines the direction and the amount of asymmetry. A positive value of g corresponds to a positive skewness. The special symmetric case which is obtained for $g = 0$ is known as h distribution. For $h > 0$ the distribution is leptokurtic with the mass is the tails increasing in h.

Thus to model the margins of all items forming a portfolio we have several choices in order to capture asymmetry, leptokurtosis and extreme events:

- The Generalized Hyperbolic Distribution
- The Extreme Value Distribution G_α, $\alpha \in \mathbb{R}$ which describes the limit distributions of normalised maxima
- The Generalized Pareto Distribution $G_{\xi,\beta}(x)$, $\xi \in \mathbb{R}$, $\beta > 0$ which appears as the limit distribution of scaled excesses over high thresholds.
- The α-stable distributions
- The g-and-h distributions.

Now with respect to the risks we need to measure the estimation and the fitting of the univariate distributions will be adapted to the data sets. The models will be different depending on the kind of risks we would like to investigate.

3.4 Interdependence Between Risks

A necessity of the stress testing is to take into account the interactions or interdependences between the entities, business units, items or risks. In most of the case, a bank will be associated to a unique risk portfolio and this one is often modelled as a weighted sum of all its parties. This approach is very restrictive as even if it captures in a certain sense the correlation between the lines it is not sufficient to model all the dependences between the risks characterising the bank. The same observation can be done when we consider the interactions between the different banks trading by the way the same products. We need to bypass this univariate approach and work with a multivariate approach. This multivariate approach permits to explain and measure the contagion effects between all the parties to model the systemic risks and their possible propagation between the different parties.

A robust way to measure the dependence between large data sets is to compute their joint distribution function. As soon as independence between the assets or risks characterizing the banks or between the banks cannot be assumed measuring interdependence can be done through the notion of copula. Recall that a copula is a multivariate distribution function linking a large set of data through their standard uniform marginal distributions (Bedford and Cooke 2001; Berg and Aas 2009). In the literature, it has often been mentioned that the use of copulas is difficult when we have more than two risks apart from using elliptical copulas such as the Gaussian one or the Student one (Gourier et al. 2009). It is now well known that these restrictions can be released considering recent developments on copulas either using nested copulas (Sklar 1959; Joe 2005) or vine copulas (Mendes et al. 2007; Rodriguez 2007; Weiss 2010; Brechmann et al. 2010; Guégan and Maugis 2010) and (Dissmann et al. 2013). These n-dimensional copulas need to be fed by some marginal distributions. For instance they can correspond to distributions characterizing the various risks faced by a financial institution. The calibration of the exposure distribution plays an important role in the assessment of the risks, whatever the method used for the dependence structure as discussed in the next section.

Until now most practitioners use Gaussian or Student t-copulas however they fail to capture asymmetric (and extreme) shocks (for example operational risks severity distributions are asymmetric-sic!). Using a Student t-copula with three degrees of freedom[3] to capture a dependence between the largest losses would mechanically imply a higher correlation between the very small losses. An alternative is to use Archimedean or Extreme Value copulas (Joe 1997a) which have attracted particular interest due to their capability to capture the dependence embedded in different parts of the marginal distributions (right tail, left tail and body). The mechanism is recalled in the following.

Let $X = [X_1, X_2, \ldots, X_n]$ be a vector of random variables, with joint distribution F and marginal distributions $F_1, F_2, \ldots, F_n$. Sklar (1959) theorem insures the existence of a function $C(.)$ mapping the individual distribution to the joint distribution:

$$F(x) = C\left(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)\right).$$

From any multivariate distribution F, the marginal distributions F_i can be extracted, and also the copula C. Given any set of marginal distributions $(F_1, F_2, \ldots, F_n)$ and any copula C the above formula can be used to compute the joint distribution with the specified marginals and copula. The function C can be an Elliptical copula (Gaussian and Student Copulas) or an Archimedean copulas (defined through a generator) as

$$C(F_X, F_Y) = \phi^{-1}\left[\phi(F_X) + \phi(F_Y)\right],$$

including the Gumbel, the Clayton the Franck copulas, among others. The Archimedean copula are easy to use because of the link existing between the Kendall's tau and the generator ϕ:

$$\tau(C_\alpha) = 1 + 4\int_0^1 \frac{\phi_\alpha(t)}{\phi'_\alpha(t)} dt. \tag{4}$$

If we want to measure the interdependence between more than two risks the Archimedean nested copula is the most intuitive way to build n-variate copulas with

[3] A low number of degrees of freedom imply a higher dependence in the tail of the marginal distributions.

bivariate copulas and consists in composing copulas together yielding formulas of the following type. For instance when $n = 3$:

$$\begin{aligned} F(x_1, x_2, x_3) &= C_{\theta_1,\theta_2}(F(x_1), F(x_2), F(x_3)) \\ &= C_{\theta_1}\left(C_{\theta_2}(F(x_1), F(x_2)), F(x_3)\right) \end{aligned}$$

where $\theta_i, i = 1, 2$ is the parameter of the copula. This decomposition can be done several times, allowing to build copulas of any dimension under specific constraints to insure that it is always a copula. Therefore a large number of multivariate Archimedean structures have been developed for instance, the fully nested structures, the partially nested copulas and the hierarchical ones. Nevertheless all these architectures imply restrictions on the parameters and impose using an Archimedean copula at each node, making their use limited in practice. To bypass the restrictions imposed by the previous nested strategy an intuitive approach proposed by Genest et al. (1995) can be used based on a pair-copula decomposition such as the D-vine (Joe 1997b) or the R-vine (Mendes et al. 2007). These approaches rewrite the n-density function associated with the n-copula as a product of conditional marginal and copula densities. All the conditioning pair densities are built iteratively to obtain the final one representing the entire dependence structure. The approach is simple and has no restriction for the choice of functions and their parameters. Its only limitation is the number of decompositions to consider as the number of vines grows exponentially with the dimension of the data sample and thus requires the user to select a vine from $\frac{n!}{2}$ possible vines, (Capéraà 2000; Galambos 1978; Brechmann et al. 2010; Guégan and Maugis 2010, 2011). These are briefly introduced now.

If f denotes the density function associated with the distribution F of a set of n r.v.X, then the joint n-variate density can be obtained as a product of conditional densities. For instance when $n = 3$ the following decomposition is obtained:

$$f(x_1, x_2, x_3) = f(x_1).f(x_2|x_1).f(x_3|x_1, x_2),$$

where,

$$f(x_2|x_1) = c_{1,2}(F(x_1), F(x_2)).f(x_2),$$

and $c_{1,2}(F(x_1), F(x_2))$ is the density copula associated with the copula C which links the two margins $F(x_1)$ and $F(x_2)$. With the same notations we have:

$$\begin{aligned} f(x_3|x_1, x_2) &= c_{2,3|1}(F(x_2|x_1), F(x_3|x_1)).f(x_3|x_1) \\ &= c_{2,3|1}\,(F(x_2|x_1)\,, F(x_3|x_1)).c_{1,3}(F(x_1), F(x_3)).f(x_3). \end{aligned}$$

Then,

$$\begin{aligned} f(x_1, x_2, x_3) =& f(x_1).f(x_2).f(x_3) \\ & .c_{1,2}(F(x_1), F(x_2)).c_{1,3}(F(x_1), F(x_3)) \\ & .c_{2,3|1}(F(x_2|x_1), F(x_3|x_1)). \end{aligned} \quad (5)$$

Other decompositions are possible using different permutations. These decompositions can be used whatever the links between the r.v. as there is no constraint. To use it in practice and eventually to obtain an initial measure of dependence between the risks inside the banks the first conditioning has to be selected.

3.5 Dynamic Approach

In order to take into account events at the origin of stress some banks consider them as single events by simply summing them thus the complete dependence scheme including their arrival time is not taken into account; it appears unrealistic as it may lead to both inaccurate capital charge evaluation and wrong management decisions. In order to overcome the problems created by these choices we suggest to use the following methodology based on the existence of dependencies between the losses through a time series process.

The dependence between assets, risks, etc. within a bank and with other banks exists and it is crucial to measure it. We propopse in the previous subsection a way to take them into account. If it does correctly it can avoid the creation of systemic risks. Nevertheless sometimes this dependence does not appear and it seems that the independence assumption cannot be rejected *a priori*, thus a profound analysis should always be performed at each step of the cognitive process. Therefore, in order to build a model close to the reality, static models have to be avoided and intrinsic dynamics should be introduced. The knowledge of this dynamical component should allow to build robust stress tests. To avoid bankruptcies and failures, generally we focus on the incidents performing any probabilistic study as proposed in the previous subsections nevertheless analysing the dynamics embedded within the incidents through time series will permit to be more reactive and close to the reality. If $(X_t)_t$ $\forall t$ denotes the losses, then our objective is to propose some time series models permitting to link the losses in time. These dynamics can be expressed in the formal following way:

$$X_t = f\left(X_{t-1,\ldots}\right) + \varepsilon_t, \tag{6}$$

where the function $f(.)$ can take various expressions to model the serial correlations between the losses, and $(\varepsilon_t)_t$ is a strong white noise following any distribution. Various classes of models may be adopted for instance short memory models e.g. AutoRegressive (AR) processes, GARCH models or long term models, e.g. Gegenbauer processes.

Besides the famous ARMA model which corresponds to a linear regression of the observations (losses here) on their known passed values (Brockwell and Davis 1988) characterising the fluctuations of the level of the losses it is possible to measure the amplitude of their volatility using ARCH/GARCH models whose a simple representation is (Engle 1982; Bollerslev 1986),

$$X_t | F_{t-1} \sim D(0, h_t), \tag{7}$$

where h_t can be expressed as:

$$h_t = h\left(X_{t-1}, X_{t-2}, \cdots, X_{t-p}, a\right), \tag{8}$$

where $h(.)$ is a non linear function in the r.v. $X_i, i = 1, \ldots, p$, p is the order of the ARCH process, a is a vector of unknown parameters, $D(.)$ is any distribution previously introduced, and F_{t-1} the σ-algebra generated by the past of the process X_s, for $s < t$, *i.e.*: $F_{t-1} = \sigma(X_s, s < t)$.

If we observe both persistence and seasonality inside the losses those can be modelled in the following way:

$$\prod_{i=1}^{k}\left(I - 2\cos(\lambda_i)B + B^2\right)^{d_i} X_t = u_t, \tag{9}$$

where $k \in N$, λ_i are the k frequencies. This representation is called a Gegenbauer process (Guégan 2005) and corresponds to the k cycles whose periods are equal to $2\pi/\lambda_i$ and d_i are fractional numbers which measure the persistence inside the cycles. This representation includes the FARMA processes (Beran 1994).

Through these time series processes the risks associated to the loss intensity which may increase during crises or turmoil are captured and the existence of correlations, dynamics inside the events, and large events will be reflected in the choice of the residual distributions. With this dynamical approach we reinforce the information done by the marginal distribution: for instance for operational risks using a time series permit to capture the dynamics between the losses and the adequate choice of the distribution of the filtering data set permits to capture the information provided in the fat tails.

4 Stress-Testing: A Combined Application

Most stress testing processes in financial institutions begin with negative economic scenarii and evaluate how the models would react to these shocks. Unfortunately, considering that the stress testing is supposed to evaluate the resilience of the bank in case of an extreme shock, if the model used to evaluate how the capital requirements would react to the integration of extreme information, by definition the model does not fully capture the risks. Obviously, it is not always simple to fully capture a bank exposure through a model, for two reasons: on the first hand, financial institutions may follow an adverse selection process, because the more extreme information you integrate, the larger the capital charge, and on the other hand the risks may not have all been identified, and in this case we are going way beyond the Black Swans.

In this section using an alternative economic reality we present approaches that allow to take into account risk behaviors that may not be captured with traditional strategies. Our objective, is not to say "the larger the capital charge the better", but to integrate all the information left aside by traditional methodologies to understand what would be the appropriate capital requirement necessary to face a extreme shock up to a certain threshold; as even if we are going further than traditional strategies to capture multiple patterns usually left aside, our models-as any model-will have its limitations, but in our case the pros and cons scale inclines toward the pros. The methodologies presented in the previous sections are applied to Market, Credit and Operational Risk data. The data used in this analysis are all genuine and the results reliable. However, in order not to confuse our message, a fictive financial institution is considered where its credit risks arise from loans contracted with foreign countries, the market risks from investments in the French CAC 40 and its operational risks are only characterized by Execution, Delivery and Process Management failures.

The impact of the methodologies are analysed applying four steps approach:

(a) In a first step, the risk measures are computed using a traditional but not conservative approach. The marginal distributions are combined using a Gaussian copula.
(b) In a second step, the marginal distributions are built challenging the traditional aspect either methodologically or with respect to the parameters increasing the conservativeness of the risk metrics.
(c) In a third step, we introduce a dynamical approach.
(d) In a fourth step, the dependence architecture is modified to capture the extreme dependencies through a Gumbel copula.

Then, for all steps listed above the impact on the capital requirement and mechanically on both the risk weight asset (RWA) and the bank balance sheet is illustrated.

4.1 Univariate Approach

For each of the three risks (credit, market and operational risks) the objective is to build profit and loss distributions. Using all data sets three marginal distributions are constructed considering various type of stress testing to evaluate the buffer a financial institution should hold to survive to a shock.

4.1.1 Credit Risks

For stressing the credit risk, various options may be considered from the simplest which would be to shock the parameters of the regulatory approach to the fit of a fat tailed distribution on the P&L function. The way the parameters are estimated may also be revised. In this section, three approaches are presented. The first provides a benchmark following the regulatory way to compute the credit risk capital charge. In the second step, the inputs are changed to reflect an economical downturn scenario. In a third step a Stable distribution (Taqqu and Samorodnisky 1994) is fitted on the P&L distribution to capture extreme shocks.

Table 1 Correlation matrix used to evaluate the credit risk regulatory capital

1	0.4	0.6	0.4	0.5	0.3	0.3	0.2
0.4	1	0.5	0.6	0.5	0.4	0.3	0.2
0.6	0.5	1	0.2	0.3	0.2	0.2	0.2
0.4	0.6	0.2	1	0.2	0.2	0.3	0.4
0.5	0.5	0.3	0.2	1	0.4	0.3	0.2
0.3	0.4	0.2	0.2	0.4	1	0.4	0.4
0.3	0.3	0.2	0.3	0.3	0.4	1	0.2
0.2	0.2	0.2	0.4	0.2	0.4	0.2	1

Traditional Scheme The credit risk rating provided by S&P which characterizes the probability of default of the sovereigns[4]. Table 2 provides the probabilities of moving from a rating to another over a year. For credit risk the regulation (BCBS 2006) provides the following formula to calculate the capital required:

$$K = \left(LGD * \Phi\left(\sqrt{\frac{1}{(1-\rho)}} * \Phi^{-1}(PD) + \sqrt{\frac{\rho}{(1-\rho)}} * \Phi^{-1}(Q)\right) - PD * LGD \right) * \frac{1}{(1-1.5*b)} * (1+(M-2.5)*b), \tag{10}$$

$$RWA = K * 12.5\,\% * EAD \tag{11}$$

where, the *LGD* is the Loss Given Default, the *EAD* is the Exposure At Default, the *PD* is the Probability of Default, $b = ((0.11852 - 0.05478) * \ln(PD))^2$ (maturity adjustment)[5], ρ corresponds to the default correlation, M the number of assets, Φ is the cdf of a Gaussian distribution and Q represents the 99^{th} percentile.

However, in our case the objective is to build a P&L distribution to measure the risk associated to the loans provided through the VaR or the Expected Shortfall, and to use it as a marginal distribution in a multivariate approach. Our approach is not limited to the regulatory capital. Therefore, a methodology identical to the one proposed by Riskmetrics (1993) has been implemented. This one is based on Merton's model (Merton 1972) which draws a parallel between option pricing and credit risk evaluation to evaluate the Probabilities of Default.

The portfolio used in this section contains loans to eight countries, for which the bank exposure is respectively \$ 40, \$ 10, \$ 50, \$ 47, \$ 25, \$ 70, \$ 40, \$ 23 million. The risk free rate is equal to 3 %. The Loss Given Default has been estimated historicaly and is set at 45 % and the correlation matrix is presented in Table 1 and the rating migration matrix in Table 2 (Guégan et al. 2013).

The VaR and the ES obtained are respectively, \$ 35 380 683 and \$ 40 888 259 (Table 5).

[4] http://www.standardandpoors.com/ratings/articles/en/us/?articleType=HTML&assetID=124535 0156739.

[5] The maturity adjustment is not always present as it is contingent to the type of credit.

Table 2 Probability of default: Credit migration matrix (S&P)

	AAA	AA	A	BBB	BB	B	C	D
AAA	92.3	6.9	0.1	0.2	0.5	0.0	0.0	0.0
AA	9.9	82.7	5.0	2.0	0.4	0.0	0.0	0.0
A	0.0	10.8	77.1	9.6	1.7	0.2	0.3	0.3
BBB	0.0	0.0	20.4	69.1	7.2	1.2	0.5	1.6
BB	0.0	0.0	0.0	18.2	67.6	10.4	0.8	3.0
B	0.0	0.0	0.0	0.6	17.1	73.1	2.4	6.8
C	0.0	0.0	0.0	0.0	0.0	32.5	16.6	50.9
D	0	0	0	0	0	0	0	100

Stressing the Input Stressing the input means that the parameters should reflect a crisis business cycle characterised for example by a decreasing GDP, an increasing unemployment rate, etc. resulting in higher probabilities of default, larger exposures at defaults, higher correlation etc. Considering that during an economical downturn, input data are already stressed, is extremely risky, as they do not contain the next extreme events. The objective of the stress testing is to evaluate the resilience of the bank to extreme shocks, where the term extreme characterises events that are worse than what the financial institution already experienced.

Table 3 Probability of default: Stress credit migration matrix

	AAA	AA	A	BBB	BB	B	C	D
AAA	88.81	8.53	0.88	0.36	0.38	0.32	0.31	0.41
AA	0.30	80.65	13.79	2.44	1.56	0.53	0.32	0.41
A	0.00	0.00	74.25	16.98	5.24	2.46	0.51	0.56
BBB	0.00	0.00	3.95	68.93	13.15	7.57	4.92	1.48
BB	0.00	0.00	0.00	5.73	65.25	25.26	2.2	1.56
B	0.00	0.00	0.00	0.00	5.39	73.06	12.54	9.01
C	0.00	0.00	0.00	0.00	0.00	13.01	34.1	52.89
D	0	0	0	0	0	0	0	100

The LGD has been stressed from 0.45 to 0.55, and Table 3 provides the stressed ratings migration matrix. In credit risk management, the simple application of a LGD downturn as prescribed by the regulation is by itself the integration of stress-testing into the traditional credit risk management scheme. The method to evaluate the risk is identical to the one presented in the previous subsection, only the components have been stressed. Compared to the metric obtained using Eq. 10, the value increased by 31.5 % from \$35 380 683 to \$46 556 127 (Table 5).

Stressing the P&L Distribution Though stressing the input may already provide a viable alternative, the creation of the loss distribution is questionable as it may not capture extreme shocks beyond the input parameters. An interesting approach is to fit a Stable distribution on the P&L distribution created using the regulatory scheme. The underlying assumption is that the regulatory distribution is not conservative enough, therefore a more conservative distribution should be fitted to the regulatory P&L function. The stable distribution is leptokurtic and heavy tailed, and its four parameters allow a flexible calibration allowing the capture of embedded

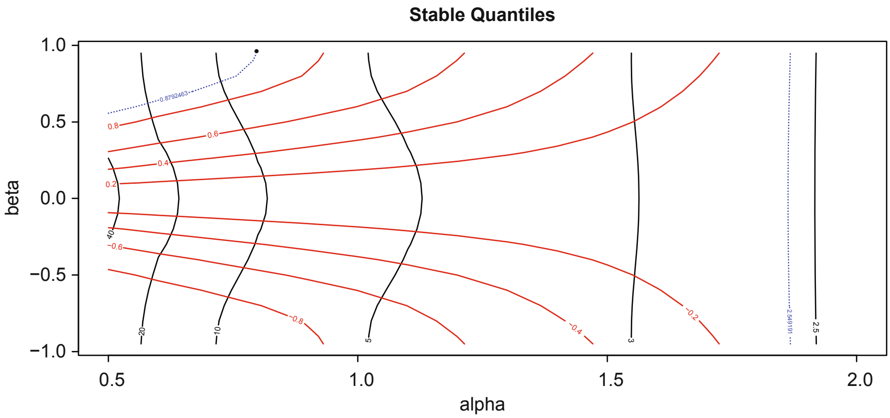


Fig. 2 Estimation of the Stable distribution using the McCulloch method

Table 4 Parameters of the Stable distribution fitted on the stressed P&L distribution obtained by stressing the input of the credit capital model

	α	β	γ	δ
Parameters	0.801	0.950	2915380.878	−1583525.686
s.d.	0.013	0	277961.5	176248.9

or assumed tail behaviours. The parameters of the Stable distribution are estimated using McCulloch approach (Fig. 2) (McCulloch 1996), (Table 4).

Comparing to the two previous other methods the values obtained from the Stable distribution we observe that we get higher values of risk measures potentially unrealistic. For instance the VaR is superior to \$572 million which is almost twice the value of the portfolio, while or loss is limited to amount lent. It is more interesting to analyze the results obtained using the α distribution in terms of the probability of occurences of the events. Indeed the probability of loss using the VaR value obtained from the stressed input approach (i.e. \$46 556 127) is supposed to a 1 % probability of loss but with the stable distribution it is in a 7.7 % probability of loss (Table 5). Therefore, the risk of such a loss is much higher with respect of our assumptions and should be mitigated consequently.

Thus, it appears that stressing the input is not sufficient and the key point in terms of stress testing lies on the choice of the P&L distribution during the crisis.

4.1.2 Market Risks

To illustrate our analysis, the methodologies presented in the previous section have been applied to data extracted from the CAC 40 index. These closing values of the index have been collected from 01 March 1990 to 06 December 2013, on a daily basis. The time series is presented in Fig. 3.

Table 5 This table presents the risk measures computed considering the three approaches presented to model the credit risk, for instance the regulatory approach, the stressed input approach and the fit of a Stable distribution. The more conservative the approach the larger the risk measures. Comparing the values obtained from the Stable distribution to the others exhibits much larger risk measures, potentially unrealistic. Here, we suggest changing the way the results are read. The line labeled "Percentile Equivalent" provides the probability of losing the VaR value obtained from the stressed input approach (i.e. \$ 46 556 127) considering a Stable distribution. What was supposed to be a 1 % probability of loss is in fact a 7.7 % probability of loss considering the Stable distribution

	VaR	ES
Regulatory	\$ 35 380 683	\$ 40 888 259
Stressed Input	\$ 46 556 127	\$ 60 191 821
Stable Distribution	\$ 572 798 381	\$ 13 459 805 700
Percentile Equivalent	92.3 %	NA

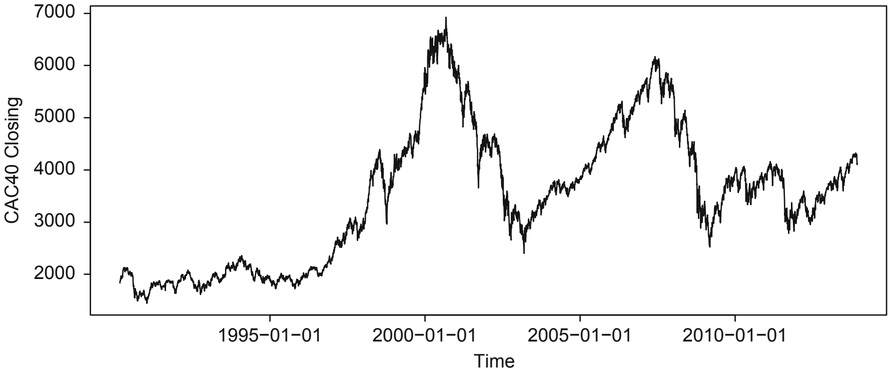


Fig. 3 CAC 40 index values from 01 March 1990 to 06 December 2013

In this subsection, we assume that our fictive financial institution only invested in the assets constituting the CAC 40 index, in the exact proportion that they replicated the index in such a way that daily returns of their portfolio are identical to those of the CAC 40 index. The daily return are computed as follows, $\log\left(\frac{Index_t}{Index_{t-1}}\right)$. The histogram of the daily log returns are represented in Fig. 4.

In this application, an initial investment of 100 million is considered.

Traditional Scheme Two approaches are considered to build the Profit and Loss distributions, the Gaussian approximation and the historical log return on investment. In a first step, a Gaussian distribution is used. The Gaussian VaR is obtained using the following equation,

$$VaR_{Market} = I_0 * \sigma * \phi_{\alpha}^{-1}(0,1) * \sqrt{(}10), \tag{12}$$

where I_0 represents the initial investment, σ is the standard deviation of the log return of the index, ϕ^{-1} is the quantile function of the standard normal distribution, $\sqrt{(}10)$ is the square root of the 10 days and α is the appropriate percentile. Following the current paradigm, in a first step, $\alpha = 0.95$ and $\sigma = 1.42\,\%$ are used.

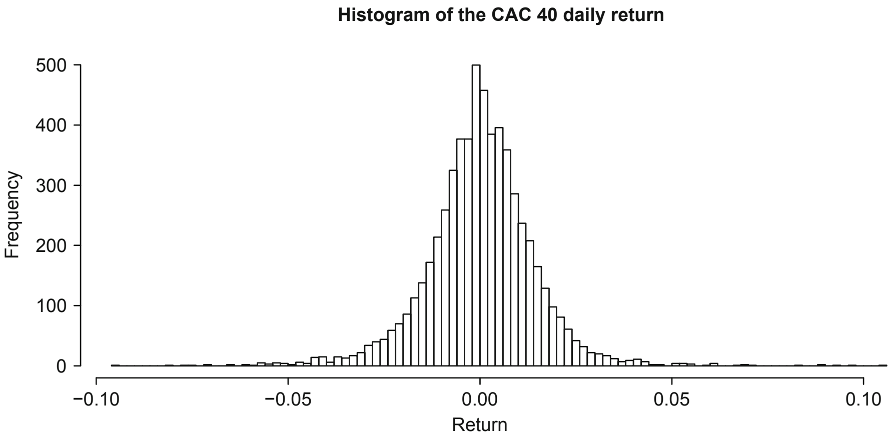


Fig. 4 Histogram of CAC 40 daily return

A common alternative is to calculate the historical VaR applying the 10-day log returns of the index time series to the portfolio value continuously compounded assuming no reduction, increase or alteration of the investment. 95 % VaR and ES have been computed and the results are presented in Table 8.

Stressing the Distribution The market risk measure is stressed switching from the traditional Gaussian distribution to a normal-inverse Gaussian distribution (NIG). As presented above, the NIG is a continuous probability distribution that is defined as the normal variance-mean mixture where the mixing density is the inverse Gaussian distribution. The NIG is a particular case of the generalised hyperbolic (GH) distributions family. This distribution is much more flexible and capture asymmetric shocks and extreme behaviours by integrating the skewness and the kurtosis of the data in the parameterization.

The parameters fitted on the 10-day log returns of the index time series applied to the portfolio value continuously compounded assuming no reduction, increase or alteration of the investment are provided in Table 6 and also their variances. The results for the VaR and the ES are provided in Table 8.

Table 6 Parameters of the NIG fitted on the 10-day log returns of the index time series applied to the portfolio value continuously compounded, and their variance covariance matrix

	$\bar{\alpha}$	μ	σ	γ
Parameters	131.8729	6373148.7957	5401353.9667	168.4646
Var-Cov	$\bar{\alpha}$	μ	σ	γ
$\bar{\alpha}$	1.363399312	1.564915e+00	−0.0023311711	−8.865840e-02
μ	1.564914681	2.280498e+06	0.1230302070	−1.628887e+05
σ	−0.002331171	1.230302e-01	0.0005837707	9.642882e-03
γ	−0.088658405	−1.628887e+05	0.0096428820	3.257814e+05

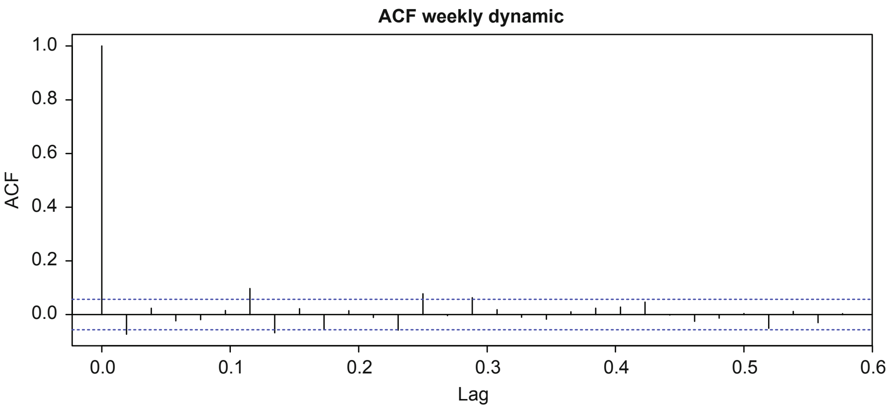


Fig. 5 ACF of the CAC 40 weekly return

Capturing an Intrinsic Dynamic Considering the market data, an ARMA model is substituted to the Gaussian approach and the appropriate distribution (potentially fat tailed) is fitted on the residuals. This approach allows the capture of intrinsic dynamics, i.e. time dependencies, between the various data points representing the returns. This approach enables capturing the patterns embedded during the crisis periods covered by the data sets, patterns which would be diluted in a more traditional approach such as a simple Gaussian or Historical approach. During a crisis, the VaR obtained would be larger as the weight of the latest events would be larger than for the oldest ones.

In a first step, the data are tested to ensure the series be stationary. The initial augmented Dickey-Fuller test (Said and Dickey (1984)) rejects the stationarity assumption, as the plot of the time series does not show any trends, the data are initially filtered to remove the seasonality components using a LOWESS process (Cleveland 1979). The results of the augmented Dickey-Fuller test post filtering is exhibited below.

Dickey-Fuller = -10.6959, Lag order = 10, p-value = 0.01

The p-value lower than 5 % allows not to reject the stationarity assumption. Considering the ACF and the PACF of the time series, respectivelly exhibited in Figs. 5 and 6, an ARMA(1,1) has been adjusted on the data. Figure 6 exhibits some autocorrelations up to 22 weeks before the latest. This could be consistent with the presence of long memory in the process. Unfortunately, the estimation procedure failed estimating the parameters properly for both the ARFIMA and the Gegenbaueur alternatives.

A NIG is fitted on the residuals. Parameters for the ARMA are presented in Table 7 and for the NIG residuals distribution in Table 7.

ARMA	$\phi_1 = -0.3314, \theta_1 = 0.2593$

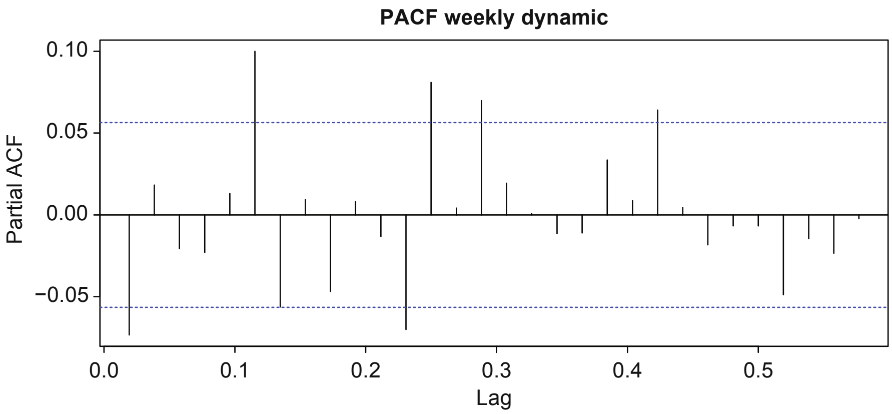


Fig. 6 PACF of the CAC 40 weekly return

Table 7 Parameters of the NIG fitted on the residuals engendered by the ARMA adjusted on the weekly CAC 40 log return. The variance covariance matrix is also provided

	$\bar{\alpha}$	μ	σ	γ
Parameters	2.127453461	0.007864290	0.028526513	− 0.007269106
Var-Cov	$\bar{\alpha}$	μ	σ	γ
$\bar{\alpha}$	0.0577158575	2.513919e-04	− 2.740407e-03	− 2.486429e-04
μ	0.0002513919	7.033579e-06	− 2.430003e-05	− 7.037312e-06
σ	− 0.0027404071	− 2.430003e-05	7.635435e-04	2.291703e-05
γ	− 0.0002486429	− 7.037312e-06	2.291703e-05	7.737263e-06

Table 8 presents risk measures computed for each of the four approach implemented. In our case the Gaussian approach provides values for the risk measures which are lower than the values obtained using historical data, therefore the Gaussian distribution does not capture the tails properly and appears irrelevant. The NIG and the ARMA process are both providing larger risk measures at the 99 % confidence level, which would be irrelevant for a traditional capital requirement calculation, but may be interesting for stress testing as in that kind of exercises, the question is to understand what could lead to the failure of the institution, and more specificaly from a market risk perpective, what could lead to the loss of our asset portfolio[6]. This reverse stress testing process is captured by the model. It is interesting to note that a conservative but static approach (the NIG) provides larger risk measures than a dynamic approach fitting the same distribution on the residuals. This means that the simple capture of the extreme events by calibrating a fat tailed distribution may be misleading regarding our interpretation of the exposure and that the research of the dynamic component is crucial. The threat is represented by an over estimation of the exposure and its implied falacious management decisions.

[6] Note that the ES obtained from the NIG is far superior to the initial investment, but is still consistent regarding a continously coumpounded portfolio.

Table 8 This table presents the risk measures computed considering the three approaches presented to model the market risk, for instance, the traditional approach either calibrating a Gaussian distribution or using an historical approach, fitting a NIG distribution to value of the portfolio and adjusting an ARMA process combined with a NIG on the residuals (Time Series). The risk measure have been computed at the 99 % level

	VaR	ES
Gaussian	7 380 300	9 261 446
Historic	8 970 352	12 871 355
NIG	93 730 084	157 336 657
Time Series	63 781 366	64 678 036

4.1.3 Operational Risks

This section describes how risks are measured considering three different approaches: the first one corresponds to the traditional Loss Distribution Approach (Guégan and Hassani 2009; Hassani and Renaudin 2013; Guégan and Hassani 2012b) the second assumes that the losses are strong white noises (they evolve in time but independently)[7], and the third one filters the data sets using the time series processes developed in the previous sections. In the next paragraphs, the methodologies are detailed in order to associate to each of them the corresponding capital requirement through a specific risk measure. According to the regulation, the capital charge should be a Value-at-Risk (VaR) (Riskmetrics 1993), i.e. the 99.9$^{\text{th}}$ percentile of the distributions obtained from the previous approaches. In order to be more conservative, and to anticipate the necessity of taking into account the diversification benefit (Guégan and Hassani 2013a) to evaluate the global capital charge the expected shortfall (ES) (Artzner et al. 1999) has also been evaluated. The ES represents the mean of the losses above the VaR therefore this risk measure is informed by the tails of the distributions.

Traditional Scheme To build the traditional loss distribution function we proceed as follows. Let $p(k, \lambda)$ be the frequency distribution associated to each data set, $F(x;\theta)$, the severity distribution, then the loss distribution function is given by $G(x) = \sum_{k=1}^{\infty} p(k;\lambda) F^{\otimes k}(x;\theta), x > 0$, with $G(x) = 0, x = 0$. The notation $\otimes$ denotes the convolution (**?**) operator between distribution functions and therefore $F^{\otimes n}$ the n-fold convolution of F with itself. Our objective is to obtain annually aggregated losses by randomly generating the losses. A distribution selected among the Gaussian, the lognormal, the logistic, the GEV (Guégan and Hassani 2012a) and the Weibull is fitted on the severities. A Poisson distribution is used to model the frequencies. As losses are assumed i.i.d., the parameters are estimated by MLE[8].

[7] This section presents the methodologies applied to weekly time series, as presented in the result section. They have also been applied to monthly time series.
[8] Maximum Likelihood Estimation.

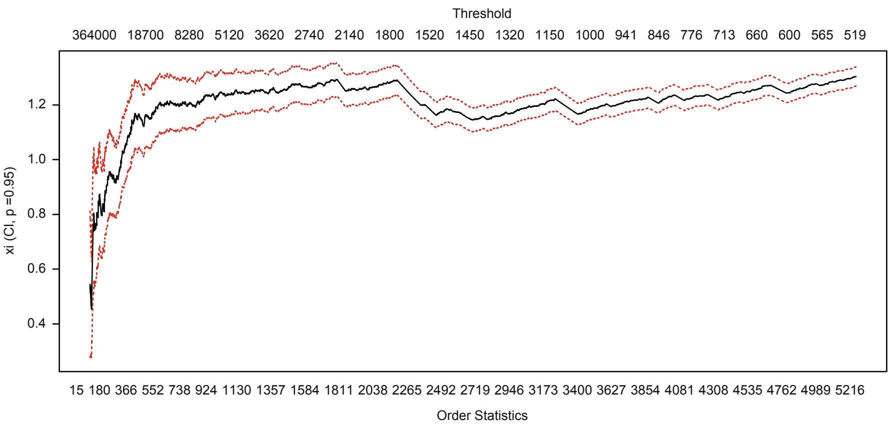


Fig. 7 Hill plot obtained from the data characterising CPBP/Retail Banking collected since 2004

Table 9 Parameters of the GPD fitted on the CPBP/Retail Banking collected from 2004 to 2011, considering an upper threshold of $13 500. The variance covariance matrix is also provided

	ξ	β
Parameters	8.244216e-01	2.172977e+04
Variances/Covariance	$\bar{\alpha}$	μ
ξ	0.008012705	$-$ 72.59912
β	$-$ 72.599117197	2856807.55692

Capturing the Fat Tails The operational risk approach is similar to the one presented in the previous paragraph. A lognormal distribution is used to model the body of the distribution while a GPD is used to characterise the right tail (Guégan et al. 2011). A conditional Maximum likelihood is used to estimate the parameters of the body while a traditional MLE is used for the GPD on the tail.

Using the Hill plot (Fig. 7), the threshold has been set at $ 13 500. This means that 99.3 % of the data are located below. However, 407 data points remains above this threshold. The parameters estimated for the GPD are given in Table 9 along their variance-covariance matrix. The parameters obtained fitting the lognormal distribution on the body of the distribution, i.e. on the data below the threshold, are given in Table 10 along their hessian. The VaR obtained with this approach equals $ 31 438 810 and the Expected Shortfall equals $ 97 112 315 (Fig. 8, 9).

Capturing the Dynamics For the second approach (Guégan and Hassani 2013b), in a first step, the aggregation of the observed losses provides the time series $(X_t)_t$. These weekly losses are assumed to be i.i.d. and the following distributions have been fitted on the severities: the Gaussian, the lognormal, the logistic, the GEV and the Weibull distributions. Their parameters have been estimated by MLE. Then 52 data points have been generated accordingly by Monte Carlo simulations and aggregated to create an annual loss. This procedure is repeated a million times to create a new loss distribution function. Contrary to the next approach, the losses are aggregated

Table 10 Parameters of the Lognormal distribution fitted on the CPBP/Retail Banking collected from 2004 to 2011, considering an upper threshold of \$13 500. The hessian is also provided

	μ	σ
Parameters	4.068128	1.917474
hessian	μ	σ
μ	10318.2793	− 678.7544
σ	− 678.7544	19134.8783

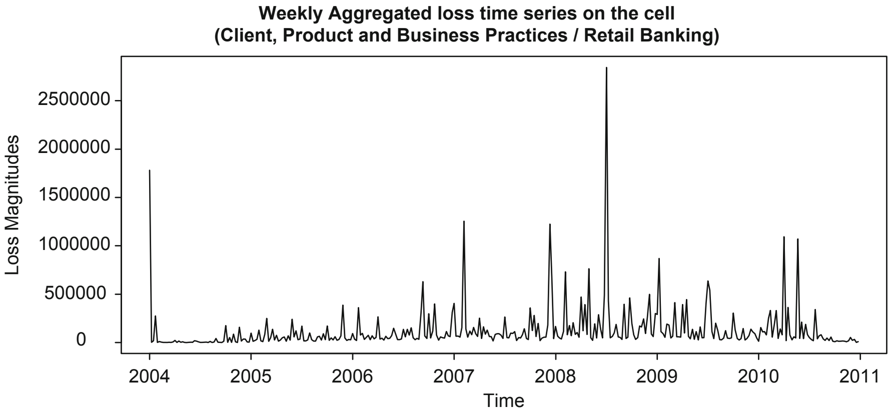


Fig. 8 The figure represents the weekly aggregated loss time series on the cell CPBP/Retail Banking collected since 2004

over a period of time (for instance, a week or a month), but no time series process is adjusted on them, and therefore no autocorrelation phenomenon is being captured.

With the third approach the weekly data sets are modelled using an AR, an ARFI and a Gegenbauer process when it is possible. Table 11 provides the estimates of the parameters for the time series processes. For The residuals a distribution is selected among the Gaussian, the lognormal, the logistic, the GEV and the Weibull distributions. Their parameters are provided in Table 12. To obtain annual losses, 52 points are randomly generated from the residuals' distributions $(\varepsilon_t)_t$ from which the sample mean have been subtracted, proceeding as follows: if $\varepsilon_0 = X_0$ corresponds to the initialisation of the process, X_1 is obtained applying one of the appropriate adjusted stochastic processes to X_0 and ε_1, and so on, and so forth until X_{52}. The 52 weeks of losses are aggregated to provide the annual loss. Repeating this procedure a million times enables creating another loss distribution function.

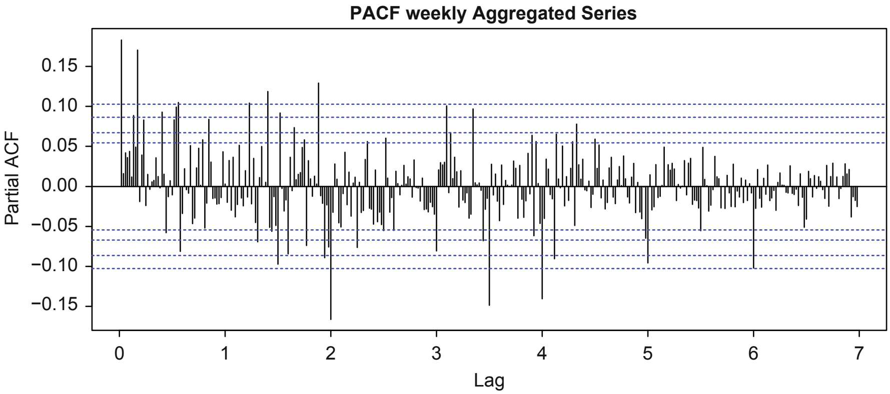


Fig. 9 The PACF of the weekly aggregated losses of the cell CPBP/Retail Banking suggests either an AR at the 5 % level or an long memory process. The order may be higher at a lower confidence level as presented in the figure. The dotted lines represnets respectivally the 95 % (*top line*) confidence intervals, the 90 %, the 80 % and the 70 %

Table 11 The table presents the estimated values of the parameters for the different models adjusted on the data sets, with their standard deviation in brackets, and also the results of the AIC criteria, the Portmanteau test and the Jarque-Bera test. The Portemanteau test has been applied considering various lags, and no serial correlation has been found after the different filterings. However, the "whiteness" of the results may be discussed using the *p*-values. Regarding the *p*-values of the Jarque-Bera test it appears that the residual distributions do not follow a Gaussian distribution

Model		CPBP/RB (W)
AR	Parameterisation	$\phi_1 = 0.1821$ (0.0552)
		$\phi_9 = 0.1892$ (0.0549)
	AIC	9964.2
	Portemanteau	lag/df = 30
		Statistic = 25.4906064
		p-value = 0.7008565
	Jarque-Bera (df = 2)	$\chi^2 = 26517.27$
		p-value < 2.2e-16
ARFI	Parameterisation	$d = 0.184673$ (0.086078), p-value = 0.03192
		$\phi_2 = -0.089740$ (0.052857), p-value = 0.08955
	AIC	−144.7204
	Portemanteau	lag/df = 30
		Statistic = 31.320582
		p-value = 0.3997717
	Jarque-Bera (df = 2)	$\chi^2 = 23875.25$
		p-value < 2.2e–16
Gegenbauer	Parameterisation	$d = 0.822$ (0.067)
		$u = -0.723$ (0.045)
	AIC	−6 466.381
	Portemanteau	lag/df = 30
		Statistic = 12.011896
		p-value = 0.9985863
	Jarque-Bera (df = 2)	$\chi^2 = 14639.36$
		p-value < 2.2e–16

Table 12 The table presents the parameters of the distributions fitted on either the i.i.d. losses or the residuals characterising the CPBP/Retail Banking weekly aggregated. Traditional LDF denotes Frequency ⊗ Severity, Time Series LDF characterises the second approach considering i.i.d. weekly losses, AR denotes the autoregressive process and both ARFI and Gegenbauer denote their related processes. The standard deviations are provided in brackets. The Goodness-of-Fit (GoF) is considered to be satisfactory if the value is $> 5\,\%$. If it is not, then the distribution associated to the largest p-value is retained. The best fit per column are presented in bold characters. Note: NA denotes a model "Not Applicable", and OD denotes results which may be provided "On Demand".

Distribution	Traditional LDF		Time Series LDF		AR(9)		ARFI(d,2)		Gegenbauer	
	Severity	GoF	Severity	GoF	Residuals	GoF	Residuals	GoF	Residuals	GoF
Gaussian	$\mu = 775.60$, $\sigma = 13449.39$ (55.77), (1557.07)	$<2.2e-16$	$\mu = 125676.6$, $\sigma = 240299.7$ (12396.71), (49772.55)	$<2.2e-16$	$\mu = -3.73e-12$, $\sigma = 193951.9$ (10321.18), (31866.62)	$2.716e-13$	$\mu = 213.66$, $\sigma = 210162.3$ (10549.58), (33898.01)	$5.551e-16$	$\mu = -33271.55$, $\sigma = 187160.6$ (13697.62), (46220.53)	$2.677e-09$
Lognormal	$\mu = 4.12$, $\sigma = 2.01$ (0.0147), (0.00998)	$<2.2e-16$	$\mu = 10.85$, $\sigma = 1.55$ (0.081), (0.092)	0.0008751	$\mu = 10.86$, $\sigma = 1.39$ (0.115), (0.0856)	$<2.2e-16$	$\mu = 10.90$, $\sigma = 1.45$ (0.118), (0.099)	$<2.2e-16$	$\mu = 10.28$, $\sigma = 1.43$ (0.171), (0.138)	$<2.2e-16$
Logistic	$\mu = 146.16$, $\beta = 702.67$ (4.199), (2.813)	$<2.2e-16$	$\mu = 117515.05$, $\beta = 72633.31$ (2097.15), (2097.15)	$<2.2e-16$	$\mu = -14762.13$, $\beta = 52516.15$ (2097.15), (1482.91)	0.2351	$\mu = -22832.232$, $\beta = 51028.33$ (2965.82), (2097.15)	0.1166	$\mu = -11419.42$, $\sigma = 38697.341$ (2965.82), (1712.32)	0.1077
GEV	NA NA	NA	$\xi = 7.37e-01$, $\beta = 41154$, $\mu = 34506$ $(6.88e-02)$, (1537), (1387)	0.03361	NA NA	NA	$\xi = -1.74e-02$, $\beta = 173140$, $\mu = -87169$ $(1.57e-02)$, (1483), (2097)	$<2.2e-16$	$\xi = -2.04e-01$, $\beta = 111700$, $\mu = -89320$ $(2.003e-06)$, (595.2), (2760)	$2.932e-10$

Table 12 (continued)

Distribution	Traditional LDF		Time Series LDF		AR(9)		ARFI(d,2)		Gegenbauer	
	Severity	GoF	Severity	GoF	Residuals	GoF	Residuals	GoF	Residuals	GoF
Weibull	$\xi = 0.441$, $\beta = 172.3$ $(1.459e-03)$, (2.096)	$<2.2e-16$	$\xi = 0.62$, 51430 $(2.262e-02)$, (2326)	$<2.2e-16$	$\xi = 0.752$, $\beta = 199529$ $(5.05e-02)$, (4201)	$<2.2e-16$	$\xi = 0.725$, $\beta = 210262$ $(4.94e-02)$, (4199)	$<2.2e-16$	$\xi = 0.662$, $\beta = 186161$ $(6.94e-02)$, (5974)	$<2.2e-16$
Hyperbolic	$\alpha = 569.74$, $\mu -$ $9.37e-04$ $\delta =$ $3.40e-06$, $\beta =$ 569.74	$<2.2e-16$	NA	NA	$\alpha =$ $9.8845e-06$, $\delta = 1031.03$, $\beta =$ $-3.088e-11$, $\mu = 6.003$	NA	NA	NA	$\alpha = 1.33394e-05$, $\delta = 875.97$, $\beta =$ $-2.746983e-06$, $\mu = -1451.36$	NA
	OD		NA		OD	NA	NA			

Table 13 The table presents the Capital charge (VaR) and Expected Shortfall of the cell CPBP/Retail Banking weekly aggregated using different distribution either to model i.i.d. losses or the residuals. Traditional LDF denotes Frequency ⊗ Severity, Time Series LDF characterises the second approach considering i.i.d. monthly losses, AR denotes the autoregressive process and both the ARFI and the Gegenbauer labels are self explanatory. Note: NA denotes a model not applicable

Distribution	Traditional LDF		Time series LDF		AR(9)		ARFI(d,2)		Gegenbauer	
	VaR	ES	VaR	ES	VaR	ES	VaR	ES	VaR	ES
Gaussian	10 417 929	10 677 551	11 798 726	12 219 110	641 575	694 887	1 421 140	1 553 661	1 997 072	2 199 513
Lognormal	6 250 108	7 234 355	39 712 897	52 972 896	1 241 192	1 679 228	4 840 470	7 609 198	6 262 572	9 779 669
Logistic	1 606 725	1 637 926	9 057 659	9 307 824	317 715	350 396	628 637	692 862	750 756	830 100
GEV	NA	NA	182 588 314	519 616 925	NA	NA	1 425 307	1 541 832	1 258 450	1 393 567
Weibull	4 168 547	4 198 009	7 497 567	8 046 971	882 146	955 434	2 587 094	2 967 185	5 892 381	6 992 815

The first remark is that, focusing on the distributions selected before, the adequacy tests may be misleading as the values are not conservative at all. The distributions have been adjusted on the residuals arising from the adjustment of the AR, the ARFI and the Gegenbauer processes. However, to conserve the white noise properties, the mean of the samples has been subtracted from the generated value, therefore, the distribution which should be the best according to the Kolmogorov-Smirnov test may not be in reality the most appropriate. As highlighted in Table 13, the use of two sided distributions lead to lower risk measures while one sided distributions lead to more conservative risk measures. Besides, these are closer to those obtained from the traditional LDA meanwhile the autocorrelation embedded within the data has been captured.

It is also interesting to note that there is not an approach always more or less conservative than the others. The capital charge depends on the strategy adopted and the couple selected: time series process and residuals distribution. For instance a Gegenbauer process associated to a lognormal distribution on CPBP/RB will be slightly more conservative than the traditional approach and enables the capture time dependency, long memory, embedded seasonality and larger tail. As a result, this may be a viable alternative approach to model the risks. The distribution generating the white noise has a tremendous impact on the risk measures. From Table 13, we observe that even if the residuals have an infinite two-sided support, they have some larger tails and an emphasised skewness. Therefore, even if the residuals have been generated using one sided distribution, as the mean of the sample has been subtracted from the values to ensure they remain white noises, the pertaining distributions have only been shifted from a $[0, +\infty[$ support to a $]-\infty, +\infty[$ support. As a result the large positive skewness and kurtosis characteristics of the data have been kept.

4.2 Multivariate Approach

A n-dimensional copulas need to be fed by some marginal distributions. In our case they correspond to the distributions created previously, each of them representing a particular risk. As exhibited in Guégan and Hassani (2013a), the choice of the model to characterise a certain risk plays an important role in the measurement of the exposure whatever the method used for the dependence structure. In the previous section, various methodologies are introduced to fit the appropriate dependence structure, for instance the nested strategies (Partially, Fully and Hierarchical) or the pair-copula decomposition such as the D-Vine and the R-Vine (Joe 1997b). While for the nested structure the dependence intensity has to decrease as the level of nesting increases the limitation of the vines is found in the number of decompositions we have to consider as the number of vines grows exponentially with the dimension of the data set and thus requires the user to select a vine from $\frac{n!}{2}$ possible vines. For optimal selection strategies, we refer the interested reader to (Capéraà et al. 2000; Galambos 1978; Brechmann et al. 2010; Guégan and Maugis 2011).

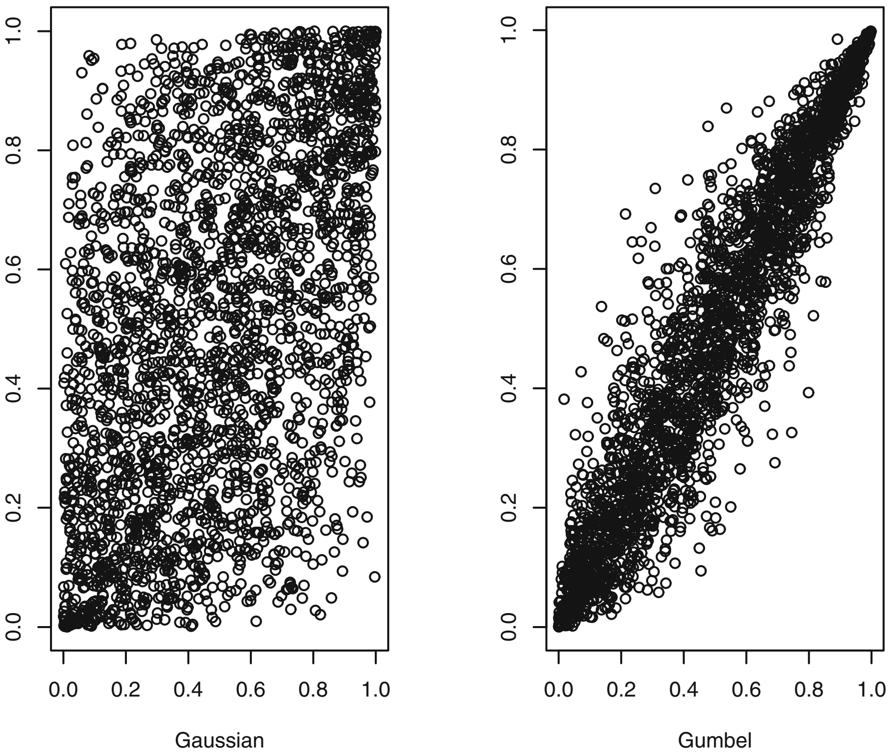


Fig. 10 The figure compares the Gaussian copula with a parameter equals 0.5 and a Gumbel copula with a parameter equals to 5. The Gumbel copula shape exhibit some upper tail dependency

In our case, the number of marginal distributions are limited as we only considered the three main risks therefore the calibration strategies introduced earlier may not be necessary and a simple maximum likelihood estimation associated to the appropriate optimization algorithm may be sufficient. However, as soon as another marginal distribution is considered, these have to be implemented to ensure a proper parametrization of the dependence structure.

A crisis is characterized by asymmetric shocks translated into upper tail dependencies. Traditional approaches use either linear correlation in the sense of Pearson (Pearson 1900), or Gaussian copulas. In the best case scenarios, Student copulas are also used. Unfortunately this latter copula with 3 degrees of freedom is symmetrical therefore the partially captured upper tail dependence is naturally translated in a modeled lower tail dependence even if small losses are independent. Besides, these structure are far from being conservative enough. Our stress-testing objective imposes other copulas like for instance the Archimedean or Extreme value copulas characterized by upper tail dependences such as the Gumbel copula, the Galambos copula (Koehler and Symanowski 1995), the Husler-Reiss (Caputo 1998) or the Tawn copula (Silverman 1986). Figure 10 compare the dependence structure obtained from

Table 14 This table presents the risk measures (at the 99 % confidence level) computed considering both the Gaussian and the Gumbel copula for which the marginal distribution have been constructed implementing traditional approaches

	Credit	
	VaR	ES
Gaussian	37 146 877	44 772 524
Gumbel	45 939 029	48 235 754

a Gaussian copula ($\rho = 0.5$) to the one obtained from a Gumbel ($\phi = 5$). The concentration of dots in the top right hand corner of the figure is characteristic of an upper tail dependence.

Applying these two last copulas to the various traditional marginal distributions constructed in the previous subsections led us to the risk measures presented in the Table 14. We note that the Gumbel copula provides a more conservative capital requirement illustrating the lower impact of the diversification benefit. These results have been obtained for assets whose correlation is not particularly high: $\rho_{Credit,Market} = 0.5$, $\rho_{Credit,Operational} = 0.4$, $\rho_{Market,Operational} = 0.4$ for the Gaussian copula, and $\phi_1 = 5$ for the Gumbel copula.

It is also interesting to note that the capture of an upper tail dependence behaviour implies the capture of a contagion effect, which engenders larger losses during a turmoil than during a calm business cycle.

References

Acerbi, C. and D. Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26 (7): 1487–1503.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9 (3): 203–228.

Barndorff-Nielsen, O., and C. Halgreen. 1977. Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 38 (4): 309–311.

BCBS. 2006. *Basel committee: International convergence of capital measurement and capital standards*. Basel: Bank for International Settlements.

Bedford, T., and R. M. Cooke. 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence* 32:245–268.

Beran, J. 1994. *Statistics for long-memory processes*. New York: CRC Press.

Berg, D., and K. Aas. 2009. Models for construction of multivariate dependence-a comparison study. *The European Journal of Finance* 15:639–659.

Berkowitz, J. 1999. *A coherent framework for stress testing*. Washington DC: Federal Reserve Board.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (3): 307–327.

Brechmann, E. C., C. Czado, and K. Aas. 2010. Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics* 40 (1): 68–85.

Brockwell, P., and R. Davis. 1988. Simple consistent estimation of the coefficients of a linear filter. *Stochastic Processes and Applications* 28 (1): 47–59.

Bunn, P., A. Cunningham, and M. Drehmann. 2005. Stress testing as a tool for assessing systemic risks. *Financial Stability Review* 18:116–126. (Bank of England Stability Review).

Capéraà, P., A. Fougères, and C. Genest. 2000. Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis* 72:30–49.

Caputo, A. 1998. Some properties of the family of koehler-symanowski distributions. Working paper, The collaborative research center (SBF), LMU-München.

Cleveland, W. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74 (368): 829–836.

Dissmann, J., E. Brechmann, C. Czado, and D. Kurowicka. 2013. Selecting and estimating regular vine copulae and application to financial returns. Statistics & Data Analysis, Elsevier.

Engle, R. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50 (4): 987–1007.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth. 1996. From data mining to knowledge discovery in databases. *AI Magazine* 17 (3): 37–54.

Galambos, J. 1978. *The asymptotic theory of extreme order statistics*. Wiley Series in Probability and Mathematical Statistics. Chichester: Wiley.

Genest, C., K. Ghoudi, and L.-P. Rivest. 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543–552.

Gourier, E., W. Farkas, and D. Abbate. 2009. Operational risk quantification using extreme value theory and copulas: From theory to practice. *Journal of Operational Risk* 3 (2009): 1–24.

Gregory, J. 2012. *Counterparty credit risk and credit value adjustment: A continuing challenge for global financial markets*, 2nd ed. London: Wiley.

Guégan, D., B. Hassani, and C. Naud. 2011. An efficient threshold choice for the computation of operational risk capital. *The Journal of Operational Risk* 6 (4): 3–19.

Guégan, D., B. Hassani, and X. Zhao. 2013. *Emerging countries sovereign rating adjustment using market information: Impact on financial institutions investment decisions*. Oxford: Academic Press.

Guégan, D. 2005. How can we define the concept of long memory? An econometric survey. *Econometric Review* 24 (2): 113–149.

Guégan, D., and B. K. Hassani. 2009. A modified panjer algorithm for operational risk capital computation. *The Journal of Operational Risk* 4:53–72.

Guégan, D., and B. Hassani. 2012a. A mathematical resurgence of risk management : An extreme modeling of expert opinions. To appear in Frontier in Economics and Finance, Documents de travail du Centre d'Economie de la Sorbonne 2011.57 - ISSN : 1955-611X.

Guégan, D., and B. Hassani. 2012b. Operational risk: A basel ii++ step before basel iii. *Journal of Risk Management in Financial Institutions* 6 (13): 37–53.

Guégan, D., and B. Hassani. 2013a. Multivariate vars for operational risk capital computation : A vine structure approach. *International Journal of Risk Assessment and Management* 17 (2): 148–170.

Guégan, D., and B. Hassani. 2013b. Using a time series approach to correct serial correlation in operational risk capital calculation. *The Journal of Operational Risk* 8 (3): 31–56.

Guégan, D., und B. Hassani. 2014. Distortion risk measure or the transformation of unimodal distributions into multimodal functions. In *Future perspectives in risk models and finance*, Alain Bensoussan, et al. (eds). Springer International Publisher, Switzerland.

Guégan, D., and P.-A. Maugis. 2010. Note on new prospects on vines. *Insurance Markets and Companies: Analyses and Actuarial Computations* 1 (1): 15–22.

Guégan, D., and P.-A. Maugis. 2011. An econometric study for vine copulas. *International Journal of Economics and Finance* 2 (1): 2–14.

Hassani, B. K., and A. Renaudin. 2013. The cascade bayesian approach for a controlled integration of internal data, external data and scenarios. Working Paper, Université Paris 1, ISSN : 1955-611X (halshs-00795046-version 1).

Huggenberger, M., and T. Klett. 2009. A g-and-h copula approach to risk measurement in multivariate financial models. Preprint, University of Mannheim, Germany.

Joe, H. 1997a. *Multivariate models and dependence concepts* (Monographs on statistics and applied probability). London: Chapman and Hall.

Joe, H. 1997b. *Multivariate models and dependence concepts* (Monographs on statistics and applied probability). London: Chapman and Hall.

Joe, H. 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis* 94:401–419.

Koehler, K. and J. Symanowski. 1995. Constructing multivariate distributions with specific marginal distributions. *Journal Multivariate Anaysis* 55 (2): 261–282.

Leadbetter, M. 1983. Extreme and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 65 (2): 291–306.

Lorenz, E. N. 2010. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20 (2): 130–141.

Majnoni, G., M. S. Martinez-Peria, W. Blaschke, and M. T. Jones. 2001. Stress testing of financial systems: An overview of issues, methodologies and fsap experiences. Working paper, IMF.

McCulloch, J. 1996. On the parametrization of the afocal stable distributions. *Bulletin of the London Mathematical Society* 28:651–655.

Mendes, B., E. de Melo and R. Nelsen. 2007. Robust fits for copula models. *Communications in Statistics-Simulation and Computation* 36(5).

Merton, R. C. 1972. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29:449–470.

Pearson, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5* 50 (302): 157–175.

Pesola, J. 2007. Financial fragility, macroeconomic shocks and banks' loan losses: evidence from europe. Working Paper, Bank of Finland Research.

Quagliariello, M. 2009. *Stress-testing the banking system: Methodologies and applications*. London: Cambridge University Press.

Riskmetrics. 1993. Var. JP Morgan. Introduction to creditmetrics. Technical Document.

Rockafellar, R., and S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2 (3): 21–41.

Rockafellar, R., and S. Uryasev. 2002. Conditional value at risk for general loss distributions. *Journal of Banking and Finance* 26 (7): 1443–1471.

Rodriguez, J.-C. 2007. Measuring financial contagion: A copula approach. *Journal of Empirical Finance* 14:401–423.

Said, S. and D. Dickey. 1984. Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika* 71 (3): 599–607.

Sereda, E., E. Bronshtein, S. Rachev, F. Fabozzi, W. Sun, and S. Stoyanov. 2010. *Distortion risk measures in portfolio optimisation*, volume 3 of Business and Economics. Springer US.

Siddique, A. and I. Hasan. 2013. *Stress testing: Approaches, methods and applications*. London: Risk Books.

Silverman, B. W. 1986. *Density Estimation for statistics and data analysis*. London: Chapman and Hall.

Sklar, A. 1959. Fonctions de répartition Ã n dimensions et leurs marges. *Publications De, l'Institut de Statistique de Paris* 8:229–231.

Taleb, N. 2010. *The Black Swan: the impact of highly improbable*. 2nd ed. New York: Random House.

Taqqu, M., and G. Samorodnisky. 1994. *Stable non-Gaussian random processes*. New York: Chapman and Hall.

Wang, S. S. 2000. A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* 67 (1): 15–36.

Weiss, G. 2010. Copula parameter estimation-numerical considerations and implications for risk management. *Journal of Risk* 13 (1): 17–53.

The *Skin in the Game* as a Risk Filter

Nassim N. Taleb and Constantine Sandis

A longer version of this chapter has appeared in the Review of Behavioral Economics, inaugural issue, January 2014.
[*Ethics/Epistemology/Risk Management/Probability*]

1 Agency Problems and Tail Probabilities

The literature in risk, insurance, and contracts has amply dealt with the notion of information asymmetry (see Ross 1973; Grossman and Hart 1983a, 1983b; Tirole 1988; Stiglitz 1988), but not with the consequences of deeper information opacity (in spite of getting close, as in Hölmstrom 1979), by which tail events are impossible to figure out from watching time series and external signs: in short, in the "real world" (Taleb 2013), the law of large numbers works very slowly, or does not work at all in the time horizon for operators, hence statistical properties involving tail events are completely opaque to the observer. And the central problem that is missing behind the abundant research on moral hazard and information asymmetry is that these rare, unobservable events represent the bulk of the properties in some domains. We define a fat tailed domain as follows: a large share of the statistical properties come from the extremum; for a time series involving n observations, as n becomes large, the maximum or minimum observation will be of the same order as the sum. Excursions from the center of the distributions happen brutally and violently; the rare event dominates. And economic variables are extremely fat tailed (Mandelbrot 1997). Further, standard economic theory makes an allowance for the agency problem, but not for the combination of agency problem, informational opacity, and fat-tailedness. It has not yet caught up that tails events are not predictable, not measurable statistically unless one is *causing* them, or involved in increasing their probability by engaging in a certain class of actions with small upside and large downside. (Both parties may not be able to gauge probabilities in the tails of the distribution, but the agent knows which tail events do not affect him.) redSadly, the economics literature's treatment of

C. Sandis (✉)
School of Philosophy,
Oxford Brooks University, Oxford, U.K.

N. N. Taleb
New York University School of Engineering,
Department of Finance and Risk Engineering, Brooklyn, New York, USA

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,* International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_4,

tail risks, or "peso problems" has been to see them as outliers to mention *en passant* but hide under the rug, or remove from analysis, rather than a core center of the modeling and decision-making, or to think in terms of robustness and sensitivity to unpredictable events. Indeed, this pushing under the rug the determining statistical properties explains the failures of economics in mapping the real world, as witnessed by the inability of the economics establishment to see the accumulation of tail risks leading up to the financial crisis of 2008 (Taleb 2009). The parts of the risk and insurance literature that have focused on tail events and extreme value theory, such as Embrechts (1997), accepts the large role of the tails, but then the users of these theories (in the applications) fall for the logical insonsistency of assuming that they can be figured out somehow: naively, since they are rare what do we know about them? The law of large numbers cannot be of help. Nor do theories have the required robustness. Alarmingly, very little has been done to make the leap that small calibration errors in models can change the probabilities (such as those involving the risks taken in Fukushima's nuclear project) from 1 in 10^6 to 1 in 50.

Add to the fat-tailedness the asymmetry (or skewness) of the distribution, by which a random variable can take very large values on one side, but not the other. An operator who wants to hide risk from others can exploit skewness by creating a situation in which he has a small or bounded harm to him, and exposing others to large harm; thus exposing others to the bad side of the distributions by fooling them with the tail properties.

Finally, the economic literature focuses on incentives as encouragement or deterrent, redbut not on disincentives as potent filters that remove incompetent and nefarious risk takers from the system. Consider that the symmetry of risks incurred on the road causes the bad driver to eventually exit the system and stop killing others. An unskilled forecaster with skin-in-the-game would eventually go bankrupt or out of business. Shielded from potentially (financially) harmful exposure, he would continue contributing to the buildup of risks in the system.[1]

Hence there is no possible risk management method that can replace skin in the game in cases where informational opacity is compounded by informational asymmetry viz. the principal-agent problem that arises when those who gain the upside resulting from actions performed under some degree of uncertainty are not the same as those who incur the downside of those same acts[2]. For example, bankers and corporate managers get bonuses for positive "performance", but do not have to pay out reverse bonuses for negative performance. This gives them an incentive to bury risks in the tails of the distribution, particularly the left tail, thereby delaying blowups.

The ancients were fully aware of this incentive to hide tail risks, and implemented very simple but potent heuristics (for the effectiveness and applicability of fast and

[1] The core of the problem is as follows. There are two effects: "crooks of randomness" and "fooled of randomness" (Nicolas Tabardel, private communication). Skin in the game eliminates the first effect in the short term (standard agency problem), the second one in the long term by forcing a certain class of harmful risk takers to exit from the game.

[2] Note that Pigovian mechanisms fail when, owing to opacity, the person causing the harm is not easy to identify.

frugal heuristics both in general and in the moral domain, see Gigerenzer 2010). But we find the genesis of both moral philosophy and risk management concentrated within the same rule[3] . About 3,800 years ago, Hammurabi's code specified that if a builder builds a house and the house collapses and causes the death of the owner of the house, that builder shall be put to death. This is the best risk-management rule ever.

What the ancients understood very well was that the builder will always know more about the risks than the client, and can hide sources of fragility and improve his profitability by cutting corners. The foundation is the best place to hide such things. The builder can also fool the inspector, for the person hiding risk has a large informational advantage over the one who has to find it. The same absence of personal risk is what motivates people to only appear to be doing good, rather than to actually do it.

Note that Hammurabi's law is not necessarily literal: damages can be "converted" into monetary compensation. Hammurabi's law is at the origin of the *lex talonis* ("eye for eye", discussed further down) which, contrary to what appears at first glance, it is not literal. *Tractate Bava Kama* in the Babylonian Talmud[4], builds a consensus that "eye for eye" has to be figurative: what if the perpetrator of an eye injury were blind? Would he have to be released of all obligations on grounds that the injury has already been inflicted? Wouldn't this lead him to inflict damage to other people's eyesight with total impunity? Likewise, the Quran's interpretation, equally, gives the option of the injured party to pardon or alter the punishment[5]. This nonliteral aspect of the law solves many problems of asymmetry under specialization of labor, as the deliverer of a service is not required to have the same exposure in kind, but incur risks that are costly enough to be a disincentive.

The problems and remedies are as follows:

First, consider policy makers and politicians. In a decentralized system, say municipalities, these people are typically kept in check by feelings of shame upon harming others with their mistakes. In a large centralized system, the sources of error are not so visible. Spreadsheets do not make people feel shame. The penalty of shame is a factor that counts in favour of governments (and businesses) that are small, local, personal, and decentralized versus ones that are large, national or multi-national, anonymous, and centralised. When the latter fail, everybody except the culprit ends up paying the cost, leading to national and international measures of endebtment against future generations or "austerity"[6].These points against "big government" models should not be confused with the standard libertarian argument

[3] Economics seems to be born out of moral philosophy (mutating into the philosophy of action via decision theory) to which was added naive and improper 19th C. statistics (Taleb 2007, 2013). We are trying to go back to its moral philosophy roots, to which we add more sophisticated probability theory and risk management.

[4] *Tractate Bava Kama*, 84a, Jerusalem: Koren Publishers, 2013.

[5] Quran, *Surat Al-Ma'idat*, 45: "Then, whoever proves charitable and gives up on his right for reciprocation, it will be an atonement for him." (our translation).

[6] See McQuillan (2013) and Orr (2013); cf. the "many hands" problem discussed by Thompson (1987).

against states securing the welfare of their citizens, but only against doing so in a centralized fashion that enables people to hide behind bureaucratic anonymity. Much better to have a communitarian municipal approach:in situations in which we cannot enforce skin-in-the game we should change the system to lower the consequences of errors.

Second, we misunderstand the incentive structure of corporate managers. Counter to public perception, corporate managers are not entrepreneurs. They are not what one could call agents of capitalism. Between 2000 and 2010, in the United States, the stock market lost (depending how one measures it) up to two trillion dollars for investors, compared to leaving their funds in cash or treasury bills. It is tempting to think that since managers are paid on incentive, they would be incurring losses. Not at all: there is an irrational and unethical asymmetry. Because of the embedded option in their profession, managers received more than four hundred billion dollars in compensation. The manager who loses money does not return his bonus or incur a negative one[7].The built-in optionality in the compensation of corporate managers can only be removed by forcing them to eat some of the losses[8].

Third, there is a problem with applied and academic economists, quantitative modellers, and policy wonks. The reason economic models do not fit reality (fat-tailed reality) is that economists have no disincentive and are never penalized for their errors. So long as they please the journal editors, or produce cosmetically sound "scientific" papers, their work is fine. So we end up using models such as portfolio theory and similar methods without any remote empirical or mathematical reason. The solution is to prevent economists from teaching practitioners, simply because they have no mechanism to exit the system in the event of causing risks that harm others. Again this brings us to decentralization by a system where policy is decided at a local level by smaller units and hence in no need for economists[9].

---

[7] There can be situations of overconfidence by which the CEOs of companies bear a disproportionately large amount of risk, by investing in their companies, as shown by Malmendier and Tate (2008, 2009), and end up taking more risk because they have skin in the game. But it remains that CEOs have optionality, as shown by the numbers above. Further, the heuristic we propose is necessary, but may not be sufficient to reduce risk, although CEOs with a poor understanding of risk have an increased probability of personal ruin.

[8] We define "optionality" as an option-like situation by which an agent has a convex payoff, that is, has more to gain than to lose from a random variable, and thus has a positive sensitivity to the scale of the distribution, that is, can benefit from volatility and dispersion of outcomes.

[9] *A destructive combination of false rigor and lack of skin in the game*. The disease of formalism in the application of probability to real life by people who are not harmed by their mistakes can be illustrated as follows, with a very sad case study. One of the most "cited" documents in risk and quantitative methods about "coherent measures of risk" set strong principles on how to compute the "value at risk" and other methods. Initially circulating in 1997, the measures of tail risk -while coherent -have proven to be underestimating risk at least 500 million times (sic, the number is not a typo). We have had a few blowups since, including Long Term Capital Management; and we had a few blowups before, but departments of mathematical probability were not informed of them. As we are writing these lines, it was announced that J.-P. Morgan made a loss that should have happened every ten billion years. The firms employing these "risk minds" behind the "seminal" paper blew up and ended up bailed out by the taxpayers. But we now know about a "coherent measure of risk".

Fourth, the predictors. Predictions in socioeconomic domains don't work. Predictors are rarely harmed by their predictions. Yet we know that people take more risks after they see a numerical prediction. The solution is to ask—and only take into account—what the predictor has done (what he has in his portfolio), or is committed to doing in the future. It is unethical to drag people into exposures without incurring losses. Further, predictors work with binary variables (Taleb and Tetlock 2013), that is, "true" or "false" and play with the general public misunderstanding of tail events. They have the incentives to be right more often than wrong, whereas people who have skin in the game do not mind being wrong more often than they are right, provided the wins are large enough. In other words, predictors have an incentive to play the skewness game (more on the problem in Sect. 2). The simple solution is as follows: predictors should be exposed to the variables they are predicting and should be subjected to the dictum "do not tell people what you think, tell them what you have in your portfolio" (Taleb 2012, p. 386). Clearly predictions are harmful to people as, by the psychological mechanism of anchoring, they increases risk taking.

Fifth, to deal with warmongers, Ralph Nader has rightly proposed that those who vote in favor of war should subject themselves (or their own kin) to the draft.

We believe *Skin in the game* is a heuristic for a safe and just society. It is even more necessary under fat tailed environments. Opposed to this is the unethical practice of taking all the praise and benefits of good fortune whilst disassociating oneself from the results of bad luck or miscalculation. We situate our view within the framework of ethical debates relating to the moral significance of actions whose effects result from ignorance and luck. We shall demonstrate how the idea of skin in the game can effectively resolve debates about (a) moral luck and (b) egoism vs. altruism, while successfully bypassing (c) debates between subjectivist and objectivist norms of action under uncertainty, by showing how their concerns are of no pragmatic concern.

Reputational Costs in Opaque Systems Note that our analysis includes costs of reputation as skin in the game, with future earnings lowered as the result of a mistake, as with surgeons and people subjected to visible malpractice and have to live with the consequences. So our concern is situations in which cost hiding is effective over and above potential costs of reputation, either because the gains are too large with respect to these costs, or because these reputation costs can be "arbitraged", by shifting blame or escaping it altogether, because harm is not directly visible. The latter category includes bureaucrats in non-repeat environments where the delayed harm is not directly attributable to them. Note that in many domains the payoff can be large enough to offset reputational costs, or, as in finance and government, reputations do not seem to be aligned with effective track record. (To use an evolutionary argument,

---

This would be the equivalent of risk managing an airplane flight by spending resources making sure the pilot uses proper grammar when communicating with the flight attendants, in order to "prevent incoherence". Clearly the problem is that tail events are very opaque computationally, and that such misplaced precision leads to confusion. The "seminal" paper: Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3), 203–228.

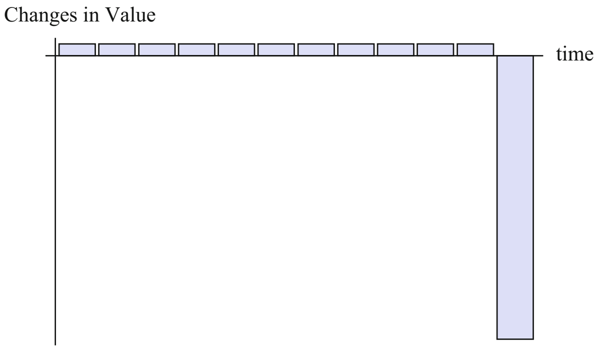


Fig. 1 The most effective way to maximize the expected payoff to the agent at the expense of the principal

we need to avoid a system in which those who make mistakes stay in the gene pool, but throw others out of it.)

Application of The Heuristic The heuristic implies that one should be the first consumer of one's product, a cook should test his own food, helicopter repairpersons should be ready to take random flights on the rotorcraft that they maintain, hedge fund managers should be maximally invested in their funds. But it does not naively imply that one should always be using one's product: a barber cannot cut his own hair, the maker of a cancer drug should not be a user of his product unless he is ill. So one should use one's products *conditionally* on being called to use them. However the rule is far more rigid in matters entailing sytemic risks: simply some decisions should never be taken by a certain class of people.

Heuristic vs Regulation A heuristic, unlike a regulation, does not require state intervention for implementation. It is simple contract between willing individuals: "I buy your goods if you use them", or "I will listen to your forecast if you are exposed to losses if you are wrong" and would not require the legal system any more than simple commercial transaction. It is bottom-up. (The ancients and more-or-less ancients effectively understood the contingency and probabilistic aspect in contract law, and asymmetry under opacity, as reflected in the works of Pierre de Jean Olivi. Also note that the foundation of maritime law has resided in skin-the-game unconditional sharing of losses, even as far in the past as 800 B.C. with the *Lex Rhodia*, which stipulates that all parties involved in a transaction have skin in the game and share losses in the event of damage. The rule dates back to the Phoenician commerce and caravan trades among Semitic people. The idea is still present in Islamic finance commercial law, see Wardé 2010.)

The rest of this essay is organized as follows. First we present the epistemological dimension of the hidden payoff, expressed using the mathematics of probability, showing the gravity of the problem of hidden consequences. We conclude with the notion of heuristic as simple "convex" rule, simple in its application (Fig. 1).

2 Payoff Skewness and Lack of Skin-in-the-Game

This section will analyze the probabilistic mismatch or tail risks and returns in the presence of a principal-agent problem.

Transfer of Harm *If an agent has the upside of the payoff of the random variable, with no downside, and is judged solely on the basis of past performance, then the incentive is to hide risks in the left tail using a negatively skewed (or more generally, asymmetric) distribution for the performance. This can be generalized to any payoff for which one does not bear the full risks and negative consequences of one's actions.*

Let $P(K, M)$ be the payoff for the operator over M incentive periods

$$P(K, M) \equiv \gamma \sum_{i=1}^{M} q_{t+(i-1)\Delta \mathrm{t}} \left(x^{j}_{t+i\Delta t} - K \right)^{+} \mathbf{1}_{\Delta} \mathrm{t}(i-1) + t < \tau \tag{1}$$

with $X^j = \left(x^{j}_{t+i\Delta t} \right)_{i=1}^{M} \in \mathbb{R}$, i.i.d. random variables representing the distribution of profits over a certain period $[t, t + i\Delta \mathrm{t}]$, $i \in \mathbb{N}$, $\Delta t \in \mathbb{R}^{+}$ and K is a "hurdle", $\tau = \inf\left\{ s : \left(\sum_{z \leq s} x_z \right) < x_{\min} \right\}$ is an indicator of stopping time when past performance conditions are not satisfied (namely, the condition of having a certain performance in a certain number of the previous years, otherwise the stream of payoffs terminates, the game ends and the number of positive incentives stops). The constant $\gamma \in (0,1)$ is an "agent payoff", or compensation rate from the performance, which does not have to be monetary (as long as it can be quantified as "benefit"). The quantity $q_{t+(i-1)\Delta \mathrm{t}} \in [1,\infty)$ indicates the size of the exposure at times $t+(i\text{-}1)\ \Delta t$(because of an Ito lag, as the performance at period s is determined by q at a a strictly earlier period $<s$)

Let $\{ f_j \}$ be the family of probability measures f_j of X^j, $j \in \mathbb{N}$. Each measure corresponds to certain mean/skewness characteristics, and we can split their properties in half on both sides of a "centrality" parameter K, as the "upper" and "lower" distributions. With some inconsequential abuse of notation we write $dF_j(x)$ as $f_j(x)\,\mathrm{d}x$, so $F_j^{+} = \int_K^{\infty} f_j(x)\ \mathrm{d}x$ and $F_j^{-} = \int_{-\infty}^{K} f_j(x)\ \mathrm{d}x$, the "upper" and "lower" distributions, each corresponding to certain conditional expectation $\mathbb{E}_j^{+} \equiv \frac{\int_K^{\infty} x f_j(x)\,\mathrm{d}x}{\int_K^{\infty} f_j(x)\ \mathrm{d}x}$ and $\mathbb{E}_j^{-} \equiv \frac{\int_{-\infty}^{K} x f_j(x)\,\mathrm{d}x}{\int_{-\infty}^{K} f_j(x)\ \mathrm{d}x}$.

Now define $\nu \in \mathbb{R}^{+}$as a K-centered nonparametric measure of asymmetry, $\nu_j \equiv \frac{F_j^{-}}{F_j^{+}}$, with values >1 for positive asymmetry, and <1 for negative ones. Intuitively, skewness has probabilities and expectations moving in opposite directions: the larger the negative payoff, the smaller the probability to compensate.

We do not assume a "fair game", that is, with unbounded returns $m \in (-\infty,\infty)$, $F_j^{+}\ \mathbb{E}_j^{+} + F_j^{-}\ \mathbb{E}_j^{-} = m$, which we can write as

$$m^{+} + m^{-} = m.$$

2.1 Simple Assumptions of Constant q and Simple-Condition Stopping Time

Assume q constant, q=1 and simplify the stopping time condition as having no loss larger than $-K$ in the previous periods, $\tau = \inf\{(t + i\Delta\text{t})\colon x_{\Delta}\text{t}(i-1)+t < K\}$, which leads to

$$\mathbb{E}(P(K,M)) = \gamma\ \mathbb{E}_j^+ \times \mathbb{E}\left(\sum_{i=1}^{M} \mathbf{1}_{t+i\Delta t<\tau}\right) \tag{2}$$

Since assuming independent and identically distributed agent's payoffs, the expectation at stopping time corresponds to the expectation of stopping time multiplied by the expected compensation to the agent $\gamma\ \mathbb{E}_j{}^+$. And $\mathbb{E}\left(\sum_{i=1}^{M} \mathbf{1}_{\Delta\text{t}(i-1)+t<\tau}\right) = \mathbb{E}\left(\left(\sum_{i=1}^{M} \mathbf{1}_{\Delta\text{t}(i-1)+t<\tau}\right) \wedge M\right)$.

The expectation of stopping time can be written as the probability of success under the condition of no previous loss:

$$\mathbb{E}\left(\sum_{i=1}^{M} \mathbf{1}_{t+i\Delta t<\tau}\right) = \sum_{i=1}^{M} F_j^+\ \mathbb{E}(\mathbf{1}_{x_{\Delta t(i-1)+t}>K}).$$

We can express the stopping time condition in terms of uninterrupted success runs. Let $\sum$ be the ordered set of consecutive success runs $\sum \equiv \{\{F\},\{\text{SF}\},\{\text{SSF}\},\ \ldots\ ,\{(M-1)\ \text{consecutive}S, F\}\}$, where S is success and F is failure over period t, with associated corresponding probabilities $\{(1-F_j^+), F_j^+\left(1-F_j^+\right), F_j^{+2}\left(1-F_j^+\right),\ \ldots\ , F_j^{+M-1}\left(1-F_j^+\right)\}$,

$$\sum_{i=1}^{M} F_j^{+(i-1)}\left(1-F_j^+\right) = 1 - F_j^{+M} \simeq 1 \tag{3}$$

For M large, since $F_j^+ \in (0,1)$ we can treat the previous as almost an equality, hence:

$$\mathbb{E}\left(\sum_{i=1}^{M} \mathbf{1}_{t+(i-1)\Delta\text{t}<\tau}\right) = \sum_{i=1}^{M} (i-1)\ F_j^{+(i-1)}\left(1-F_j^+\right) \simeq \frac{F_j^+}{1-F_j^+}.$$

Finally, the expected payoff for the agent:

$$\mathbb{E}\left(P(K,M)\right) \simeq \gamma\ \mathbb{E}_j^+ \frac{F_j^+}{1-F_j^+},$$

which increases by (i) increasing $\mathbb{E}_j^+$, (ii) minimizing the probability of the loss F_j^-, but, and that's the core point, even if (i) and (ii) take place at the expense of m the total expectation from the package.

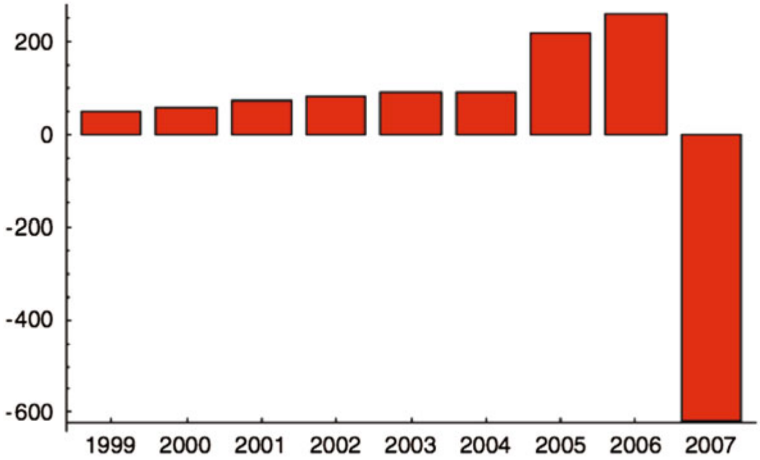


Fig. 2 Indy Mac, a failed firm during the subprime crisis (from Taleb 2009). It is a representative of risks that keep increasing in the absence of losses, until the explosive blowup

Alarmingly, since $\mathbb{E}_j^+ = \frac{m-m^-}{F_j^+}$, the agent doesn't care about a degradation of the total expected return m if it comes from the left side of the distribution, m^-. Seen in skewness space, the expected agent payoff maximizes under the distribution $j$with the lowest value of ν_j (maximal negative asymmetry). The total expectation of the positive-incentive without-skin-in-the-game depends on negative skewness, not on m (Fig. 2).

2.2 *Multiplicative q and the Explosivity of Blowups*

Now, if there is a positive correlation between q and past performance, or survival length, then the effect becomes multiplicative. The negative payoff becomes explosive if the allocation $q$increases with visible profitability, as seen in Fig. 2 with the story of IndyMac, whose risk kept growing until the blowup[10]. Consider that "successful" people get more attention, more funds, more promotion. Having "beaten the odds" imparts a certain credibility. In finance we often see fund managers experience a geometric explosion of funds under management after perceived "steady" returns. Forecasters with steady strings of successes become gods. And companies that have

[10] The following sad anecdote illustrate the problem with banks. It was announces that "JPMorgan Joins BofA With Perfect Trading Record in Quarter" (Dawn Kopecki and Hugh Son—Bloomberg News, May 9, 2013). Yet banks while "steady earners" go through long profitable periods followed by blowups; they end up losing back all cumulative profits in short episodes, just in 2008 they lost around 4.7 trillion U.S. dollars before government bailouts. The same took place in 1982–1983 and in the Savings and Loans crisis of 1991, see Taleb (2009).

Table 1 Multiplicative effect of skewness

	F = .6	0.7	0.8	0.9
r = 0	1.5	2.32	3.72	5.47
0.1	2.57	4.8	10.07	19.59
0.2	4.93	12.05	34.55	86.53
0.3	11.09	38.15	147.57	445.59

hidden risks tend to outperform others in small samples, their executives see higher compensation. So in place of a constant exposure q, consider a variable one:

$$q_{\Delta \mathrm{t}(i-1)+t} = q\ \omega(i),$$

where $\omega(i)$ is a multiplier that increases with time, and of course naturally collapses upon blowup.

Equation 1 becomes:

$$P(K, M) \equiv \gamma \sum_{i=1}^{M} q\ \omega(i) \left(x^{j}_{t+i\Delta \mathrm{t}} - K \right)^{+} \mathbf{1}_{t+(i-1)\Delta \mathrm{t}<\tau}\ , \tag{4}$$

and the expectation, assuming the numbers of periods, M is large enough

$$\mathbb{E}(P(K, M)) = \gamma\ \mathbb{E}^{+}_{j}\ q\ \mathbb{E}\left(\sum_{i=1}^{M} \omega(i)\ \mathbf{1}_{\Delta \mathrm{t}(i-1)+t<\tau} \right). \tag{5}$$

Assuming the rate of conditional growth is a constant $r \in [0,\infty)$, and making the replacement $\omega(\mathrm{i}) \equiv e^{ri}$, we can call the last term in Eq. 2 the multiplier of the expected return to the agent:

$$\mathbb{E}\left(\sum_{i=1}^{M} e^{ir} \mathbf{1}_{\Delta \mathrm{t}(i-1)+t<\tau} \right) = \sum_{i=1}^{M} (i-1)\ F_j{}^{+} e^{ir} \mathbb{E}(\mathbf{1}_{x_{\Delta t(i-1)+t}>K}) \tag{6}$$

$$= \frac{\left(F^{+}-1\right)\left(\left(F^{+}\right)^{M}\left(M e^{(M+1)r} - F^{+}(M-1)e^{(M+2)r}\right) - F^{+}e^{2r}\right)}{(F^{+}e^{r}-1)^{2}} \tag{7}$$

We can get the table of sensitivities for the "multiplier" of the payoff (Table 1):

2.3 Explaining why Skewed Distributions Conceal the Mean

Note that skewed distributions conceal their mean quite well, with $P(X < \mathbb{E}(x)) < \frac{1}{2}$ in the presence of negative skewness. And such effect increases with fat-tailedness. Consider a negatively skewed power law distribution, say the mirror image of a standard Pareto distribution, with maximum value $x_{\min}$, and domain $(-\infty, x_{\min}]$, with exceedance probability $P(X > x) = -x^{-\alpha} x_{\min}^{\alpha}$, and mean $-\frac{\alpha x_{\min}}{\alpha - 1}$, with $\alpha > 1$, have a proportion of $1 - \frac{\alpha - 1}{\alpha}$ of its realizations rosier than the true mean. Note that fat-tailedness increases at lower values of α. The popular "eighty-twenty", with tail exponent $\alpha = 1.15$, has >90 % of observations above the true mean[11]. Likewise, to consider a thinner tailed skewed distribution, for a Lognormal distribution with domain $(-\infty, 0)$, with mean $m = -e^{\mu + \frac{\sigma^2}{2}}$, the probability of exceeding the mean is $P(X > m = \frac{1}{2}\text{erfc}\left(-\frac{\sigma}{2\sqrt{2}}\right)$, which for $\sigma = 1$ is at 69 %, and for $\sigma = 2$ is at 84 %.

2.4 Forecasters

We can see how forecasters who do not have skin in the game have the incentive of betting on the low-impact high probability event, and ignoring the lower probability ones, even if these are high impact. There is a confusion between "digital payoffs" $\int f_j(x)\,\mathrm{d}x$ and full distribution, called "vanilla payoffs", $\int x f_j(x)\,\mathrm{d}x$, see Taleb and Tetlock (2013)[12].

Acknowledgments Jon Elster, David Chambliss Johnson, Gur Huberman, Raphael Douady, Robert Shaw, Ralph Nader, Barkley Rosser, James Franklin, Russell Roberts, Marc Abrahams, Andreas Lind and Elias Korosis.

References

1. Embrechts, P. 1997. *Modelling extremal events: For insurance and finance*. Springer.
2. Gigerenzer, G. 2010. Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science* 2 (2010): 528–554.
3. Grossman, S. J., and O. D. Hart. 1983a. An analysis of the principal-agent problem. *Econometrica* 51 (1): 7–45.
4. Grossman, S. J., and O. D. Hart. 1983b. Implicit contracts under asymmetric information. *The Quarterly Journal of Economics* 98 123–156.
5. Hölmstrom, B. 1979. Moral hazard and observability. *The Bell Journal of Economics* 10 (1): 74–91.

[11] This discussion of a warped probabilistic incentive corresponds to what John Kay has called the "Taleb distribution", John Kay "A strategy for hedge funds and dangerous drivers", Financial Times, 16 January 2003.

[12] Money managers do not have enough skin in the game unless they are so heavily invested in their funds that they can end up in a net negative form the event. The problem is that they are judged on frequency, not payoff, and tend to cluster together in packs to mitigate losses by making them look like "industry event". Many fund managers beat the odds by selling tails, say covered writes, by which one can increase the probability of gains but possibly lower the expectation. They also have the optionality of multi-time series; they can manage to hide losing funds in the event of failure. Many fund companies bury hundreds of losing funds away, in the "cemetery of history" (Taleb 2007).

6. Malmendier, U., and G. Tate. 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89 (1): 20–43.
7. Malmendier, U., and G. Tate. 2009. Superstar CEOs. *Quarterly Journal of Economics* 124 (4): 1593–1638.
8. Mandelbrot, Benoit B. 1997. *Fractals and scaling in finance: Discontinuity, concentration, risk*. New York: Springer-Verlag.
9. Ross, S. A. 1973. The economic theory of agency: The principal's problem. *The American Economic Review* 63 (2): 134–139.
10. Slote, M. A. 1964. An empirical basis for psychological egoism. *Journal of Philosophy* 61: 530–537.
11. Stiglitz, J. E. 1988. *Principal and agent. In the new palgrave: A dictionary of economics*, 3 vol. London: Macmillan.
12. Taleb, N. N. 2007. *The black swan*. Penguin and Random House.
13. Taleb, N. N. 2009. Errors, robustness, and the fourth quadrant. *International Journal of Forecasting* 25 (4): 744–759.
14. Taleb, N.N. 2012. *Antifragile: Things that gain from disorder*. Penguin and Random House.
15. Taleb, N.N. 2013. *Probability and risk in the real world, vol 1: Fat Tails*.
16. Taleb, N.N., and P. E Tetlock. 2013. On the difference between binary prediction and true exposure with implications for forecasting tournaments and decision making research. SSRN. http://ssrn.com/abstract=2284964 or http://dx.doi.org/10.2139/ssrn.2284964. Accessed 25 June 2013.
17. Tirole, J. 1988. *The theory of industrial organization*. Cambridge: MIT press.
18. Wardé, I. 2010. *Islamic finance in the global economy*. Edinburgh University Press.

Capital Adequacy, Pro-cyclicality and Systemic Risk

Raphael Douady

1 Reasons and Problems of the Current Capital-Adequacy

The current Capital Adequacy rule, as stated by Basel II agreement, computes the "regulatory capital" of a financial institution as a multiple of its "Value-at-Risk" (VaR), itself the sum of several risk sources: market risk, credit risk, counterparty risk, operational risk, etc.

The recommendations for computing the "economic capital" follows the same guidelines.

The VaR depends on two parameters: the horizon h (usually 5 days) and a percentile q (usually 99 %). It is the amount V such that the probability that the institution loss over the horizon h exceeds V is equal to $1-q$. An abundant literature has been published on the various methods for computing the Value-at-Risk.

The required capital to operate on markets is $k \times$ VaR, where the multiplier k is a number between 3 and 10, which depends on the "quality" of the VaR computation. The assessment of the "quality" is the result of two verifications—and as the result, the worst of both:

- a qualitative assessment of the process
- a quantitative back-test counting the number of exceptions, i.e. where the loss exceeds the VaR, and comparing this number with the declared frequency $1-q$.

The good thing in this setting is that the Regulator lets the institution compute its risks, assuming that it has a better knowledge of its details to better track the pitfalls of its own risk evaluation. It only acts as a *verifier* who checks afterwards that the risk has been correctly computed. If the risk was underestimated, the sanction is a higher ratio, hence a higher cost in regulatory capital for further operations.

It also represents a substantial economy for the Regulator, who leaves the burden of computing risks to the institutions. It is indeed a massive distribution of the burden

R. Douady (✉)
Labex ReFi, University Paris 1, 106 Bd de l'Hôpital, 75647 Paris Cedex 13, France
e-mail: rdouady@univ-paris1.fr

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,* International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_5,

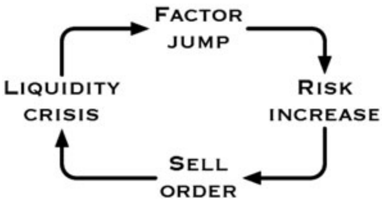


Fig. 1 Procyclicality of *reactive* VaR measures

across all the institutions, each of them taking care of its own risk computation. The verification task is less costly than the computation itself by orders of magnitude.

However, this approach leads to numerous problems, which became stringent throughout the 2008 crisis:

1. The rationale for the economic capital is to avoid bankruptcy, hence the loss should *never* exceed $k \times$ VaR. This trigger is surprisingly enough never tested.
2. The risk measure is 1-*dimensional* and neither tells the exact risk source, nor the market scenario it corresponds to. As a consequence, the Regulator cannot realistically require that the loss *never* exceeds the declared risk.
3. The most serious problem is *pro-cyclicality*: in a market downturn, the risk measure increases, leading most market participants to sell out positions in order to meet capital adequacy, adding to the market turmoil (Fig. 1).

2 Should a Risk Measure be Reactive or Anticipative?

2.1 Why VaR Reactivity Is Dangerous

It is often thought that a reactive VaR measure is a good one. We will argue the contrary. Reactivity of a VaR measure means that upon a sudden market event, for example a jump in one of the factors relevant to the risk of a fund, the risk measure immediately increases. It is perceived that such a risk measure will let the manager react quickly to changing market conditions and adjust her/his positions accordingly. However, quite the contrary is true. Reactivity of a VaR measure only shows that it does *not* in fact reveal all the hidden risks of an investment and, therefore, leads to a very dangerous circle.

A reactive measure of the VaR creates a mechanism that leads to a dangerous cyclical cascade of market events as shown in the figure above:

Factor jump ⇒ Risk increase ⇒ Sell order ⇒ Liquidity crisis ⇒ Factor jumps further

This is the typical risk adjustment chain that induced the 1987 crisis, as well as a number of other crises, including the recent 2008 credit squeeze.

2.2 Reactivity and Backtesting

On a wider scale this pro-cyclical nature of a reactive VaR creates the conditions for a deepening of a crisis. For a manager, it induces the cancelation of risky positions in illiquid times that are least favorable. This phenomenon is obvious, although seemingly counterintuitive, since most of the time, one will find risk measurers advertising their reactivity. The reason for leaning towards a reactive measure lies in the backtesting principles used by Regulators to assess the risk measure soundness. The Basel committee recommends that two properties of the VaR be tested:

1. **Frequency of out-of-sample exceptions:** an "out-of-sample" exception of the VaR with percentile q (e.g. 99 %) and horizon h (e.g. 10 days) occurs when the portfolio loss between date t and date $t+h$ exceeds the VaR computed at date t. Such exceptions should not occur with a frequency higher than $1-q$ (i.e. 1 % of the time if $q = 99\,\%$.)
2. **Independence of exceptions:** the occurrence of exceptions at dates separated by more than the horizon should not be correlated. In other words, if $n \geq h$, then the frequency of exceptions should not be significantly impacted if we condition it by the fact that there was or there wasn't an exception n days before.

It appears that when a VaR measure is purely based on past returns, the more the measure is reactive, i.e. relies on the recent past and ignores the deep past, the better conditions 1 and 2 are satisfied. Other properties of the VaR which, in our view, are as important as the two above, are being left aside. Taking them into account would force risk methodologies to look deeper in the past. These properties are:

3. **Independence on market regime:** exceptions should not occur more frequently in periods of high volatility or in periods of low volatility. By definition, a risk is a surprise. If a bias is observed with respect to some measure of the environment, it should be incorporated in the models.
4. **Inclusion of anticipated extreme events:** if an extreme event, a sector or geographic crisis is announced as a possible event, it should not been taken as an excuse for an exception, because its impact should be incorporated in the VaR measure. Risk managers should be equally surprised by its occurrence and by its "non-occurrence".
5. **Exception size:** exceeding the VaR by a small amount is not as serious as exceeding it by a large amount. This is essential as capital adequacy is meant to protect institutions from failure. If capital adequacy is a multiple of the VaR, then exceptions that are larger than this multiple are lethal and should truly be avoided. Counting exceptions that exceed certain multiples of the VaR is therefore a necessity. One could say that a sound risk measure is such that the portfolio loss, counted in units of risk measures, should have thin tails: high multiples are really rare events!

2.3 Linear Models, Even with Fat-Tails, Are Reactive

In most factor risk models, a fund, a portfolio is modelled by a function that is typically linear in the factor returns:

$$\text{Fund/Portfolio return} = \text{Model (Factor return)} = k_1 F_1 + \ldots + k_n F_n$$

In this case, the risk is determined by two inputs:

- the joint density of factor returns F_i, and
- the coefficients k_i

Following a jump of one of the factors, two things typically occur:

- The distribution of the factor is changed. This may possibly be a substantial change since the factor distributions are estimated by deliberately overweighting the recent past. This is done in order to produce better percentages of exceptions in backtesting.
- The correlation between the fund and the factor changes, typically increasing the coefficients k_i. This is sometimes referred to as a correlation break, where in effect it is simply a sign of nonlinearities which the linear model is not able to capture.

Both of these reactions will increase the risk estimate, and sometimes, when acting jointly, by a substantial amount. Back-filling the fund history in order to produce longer term statistics does not remedy the issues and is just an intermediary step resulting ultimately in same mechanism of risk reaction to a factor jump.

3 Naïve Delusion of Fat Tails

Fat tailed models have been introduced in risk management to overcome the insufficiencies of the long-lasted "normal model", based on Gaussian distributions. But *risk management* is not simply *risk measurement*. The principle of fat tails is to enhance the probability of large events (5–10 standard deviations) from truly negligible to one or a few percents, in order to enter into VaR calculations.

However, this approach, based on stretching the shape of distributions in order to force them to incorporate observed occurrences that don't fit into the "normal" model, miss the essential of risk estimation, and therefore of its management: the actual behaviour of markets during a crisis is far different from what can be observed in business as usual situations. Like a crowd in panic rushing through the door when the alarm rings has nothing to do with the same crowd, calmly exiting the room at the end of the show. Yet, the panic behaviour can be anticipated, not by "stretching" the normal one, but by observing the same or other crowds under panic.

When financial markets enter a crisis, a certain number of well-known features are observed: different asset classes, which usually are uncorrelated, become correlated, alternative investments, which have been precisely chosen for delivering *alpha* without *beta*, suddenly exhibit *beta* and no *alpha* at all, etc. Fat tail models, which are mostly calibrated on business-as-usual periods, completely miss these particular features. So-called "robust calibration" is even more of a flawed patch to the problem, as it minimizes the weight of large events in the calibration, when one should, on the contrary, increase it.

The only possible approach to anticipate crises (even small ones) and provide meaningful hedging or risk mitigation recommendations is to use models in which, in order to simulate extreme events, the calibration specifically focuses on the most agitated periods and, more particularly, on those extreme correlations one can observe during these periods.

4 Proposed Solution for an Non Pro-Cyclical Capital-Adequacy Rule

Pro-cyclicality results from the fact that only a global risk measure is considered (see point 2 above) and from the *reactivity*, rather than *pro-activity* of the risk measure. Preventing a risk measure from being pro-cyclical is not an easy task. It requires the regulator to anticipate market crises, using its knowledge of the financial and economic situation. It also requires verifying that financial institutions have a correct monitoring of their extreme exposures. Indeed, any type of rule that forces institutions to act in the middle of the turmoil will mechanically aggravate the liquidity crisis and add to the turmoil.

4.1 General Principle

We here propose to include *stress tests* in the measure of capital adequacy in such a way that we respect the following 3 *golden rules*:

1. The Regulator defines which stress to apply to which indices. This will ensure that economic research is unbiased to anticipate potential market shifts.
2. The Institution computes itself the impact of stress scenarios on its activity. It is free to add other scenarios that the Regulator didn't think of for its particular case, either by stressing other risk factors or by increasing the stress size given by the Regulator.
3. The Regulator verifies that losses incurred by the Institution (if ever) do not exceed what could have been anticipated given the declared stress tests and the actual market moves. In other words, the Institution is responsible for correctly anticipating the impact of markets on its activity, but not for the moves of markets itself.

The required operating capital is proportional to the worst declared stress test (as of 2). The initial multiplier value is 1. In case of a violation, the impact on the multiplier depends on the amount of the violation. Minor violations have a minor impact, while large violations severely impact the multiplier. If there is no violation, the multiplier is progressively brought down back to the value 1.

We mean to exclude all notions of probability in this framework. Institutions should be responsible for the amounts *they declare, not for the* probability *of such or such event.*

4.2 Details of the Regulatory Capital Computation

4.2.1 Stress Tests Defined by the Regulator

Each period (e.g. month, but can be more frequent if necessary), the Regulator issues a list of market indices to stress with, for each index, a list of amounts by which it should be stressed.

This means a table of the following kind:

Index	Stress ++	Stress +	Stress 0	Stress −	Stress −
S&P500	+ 20 %	+ 10 %	0 %	– 10 %	– 20 %
TB yield 10Y	+ 200 bp	+ 100 bp	0 bp	– 100 bp	– 200 bp
BAA credit spread	+ 500 bp	+ 200 bp	− 10 bp	– 100 bp	– 200 bp
...					

Each entry of this table corresponds to a shift of an index I_i by an amount $\Delta\mathrm{I}_{ij}$. Ideally, each $\Delta\mathrm{I}_{ij}$ corresponds to some percentile of the anticipated probability distribution of the index shift over the next coming month. For instance:

Stress ++ = 99 % percentile up
Stress + = 84 % percentile up
Stress 0 = Median
Stress − = 84 % percentile down
Stress − = 99 % percentile down

We here gave an example with round figures, but the regulator is free to apply any quantitative model leading to the values of the shifts $\Delta\mathrm{I}_{ij}$. It is somehow recommended that these figures be rather stable through time, especially the most extreme ones, as capital requirements will be direct functions of them. In particular, rather than keeping them strictly constant for some time, then re-adjusting them suddenly, one should estimate them in the most anticipative manner, in order to temper down the probability of a large jump.

A sequence of 5 stresses for each index seems reasonable but this number can be subject to discussions. More important is the list of basic indices to be stressed. These should cover all asset classes (equity, fixed-income, credit, volatility, FX, emerging markets, commodities, real estate, etc.), as well as the most important drivers of majorly traded securities.

It is extremely important that the Regulator makes all efforts to anticipate the distribution of possible forward moves, and does not simply rely on the past volatility of each index.

4.2.2 Computation of Stress Scenarios by Institution

Each ΔI_{ij} must be seen by the institution as a full market scenario, not just a single shift of a single asset class. For instance, if the scenario represents a 15 % increase of the oil price, the impact of such a scenario on other asset classes, such as energy related stocks or whatever market should be accounted for. The importance of this point will be clear when we shall describe how the Regulator will verify the accuracy of risk computations.

First, the Institution can, on a fully customary basis, decide to divide itself into "divisions" $D_1, \ldots, D_n$ which correspond to business units mostly exposed one dominant source of risk. Then, every reporting period (e.g. every week), each division D_k produces a Risk report that contains its own P/L estimate in case of scenario ΔI_{ij}, which we denote L_{ijk}.

This P/L estimate is supposed to be the *lower bound* of a confidence interval of the impact of the scenario. Specifying the probability to which this confidence interval corresponds to is not necessary. It is in the interest of the Institution not to overpass this lower estimate, or by a limited amount, as we shall see.

The institution is free to add more scenarios ΔI_{ij} either by adding other risk factors I_i or other shifts ΔI_{ij} for existing factors. It is in the interest of the institution to be as exhaustive as possible in the declaration of its risk sources, in order to avoid violations which may increase its multiplicative ratios. For instance, would one of the divisions be particularly exposed to Kazakhstan equities, a risk factor not listed by the Regulator, it is in the interest of the Institution to report a potential exposure to this risk factor, in order to avoid a violation of declared risks in case of a pure Kazakh crisis.

The regulatory capital C is computed as follows, where λ represents the multiplier:

$$C = \lambda \sum_k \max_{i,j} L_{ijk}$$

In other words, the Institution's capital is the sum of that of its divisions and, for each division, it is the maximum potential loss stemming from one of the declared scenarios.

4.2.3 Back-testing Stress Scenarios

The key point of this regulatory framework is the ability for the Regulator to back-test the accuracy and completeness of risk reports by financial institutions. Risk reports are established at dates t for a horizon h. Institutions provide, for each market scenario ΔI_{ij} the possible loss $L_{ijk}(t)$ of each division D_k. At the end of the period $t+h$, the

Regulator observes the actual shift of each index $\Delta \mathrm{I}_{i\,\mathrm{act}}(t) = \mathrm{I}_i(t+h) - \mathrm{I}_i(t)$. By a simple linear interpolation, or extrapolation if this shift happens to exceed the range of published $\Delta \mathrm{I}_{ij}$, $j = 1...5$, the Regulator computes what should have been the impact of this shift on each division D_k:

$$\mathrm{L}_{ik,\mathrm{impact}}(t) = \text{Inter/Extrapolation of } \mathrm{L}_{ijk} \text{ for } \Delta \mathrm{I}_{i\,\mathrm{act}}(t)$$

Note that, in case $\Delta \mathrm{I}_{i\,\mathrm{act}}(t)$ is outside the bounds of the 5 values $\Delta \mathrm{I}_{ij}$, this computed loss may exceed the maximum declared loss. Then the *Maximum Accepted Loss* of the division is computed:

$$MAL(D_k, t) = \max_i L_{ik,\mathrm{impact}}(t)$$

The Institution *Maximum Accepted Loss* is the sum of that of its divisions:

$$MAL(t) = \sum_k MAL(D_k, t)$$

A violation is incurred when the Institution actual loss $\mathrm{L}_{\mathrm{act}}(t)$ over the period $[t, t+h]$ exceeds $\mathrm{MAL}(t)$. In this case, the *violation ratio* is simply:

$$\mathrm{V}(t) = \max(1,\ \mathrm{L}_{\mathrm{act}}(t)/\mathrm{MAL}(t))$$

4.2.4 Computation of the Multiplier

The multiplier $\lambda(t)$ is re-computed at every reporting period, according to 2 rules:

1. If there is no violation, it is reduced in order to converge after a reasonable period of time to the value 1.
2. If there is a violation, it is increased by an amount depending on the severity of the violation.

The formula proposed here to compute $\lambda(t)$ from its previous value $\lambda(t-1)$ and from the violation ratio is quite simple. First we compute a natural dampening of $\lambda(t-1)$:

$$\bar{\lambda}(t) = \max(1, \min((1-\alpha)\lambda(t), \lambda(t) - \varepsilon))$$

In this formula, $1-\alpha$ is a dampening factor, whose role is to tame down the multiplier. If the multiplier is already rather close to 1, the reduction is at least ε, without possibility to drop below 1. Parameters α and ε depend on the reporting frequency and should be set in such a way that, approximately after a year without violation, the multiplier is set back to 1. For instance, for a weekly reporting, $\alpha = 1\%$, $\varepsilon = 2\%$ ($\approx 1/52$).

In words, every week, if there is no violation, the multiplier is reduced by 1 % of its value, the reduction not being less than 0.02 until the value 1 is reached. If $\lambda(t-1) > 2$, then $\bar{\lambda}(t) = 0.99\lambda(t-1)$, if $1.02 < \lambda(t-1) \le 2$ then $\bar{\lambda}(t) = \lambda(t-1) - 0.02$ and if $1 \le \lambda(t-1) \le 1.02$, then $\bar{\lambda}(t) = 1$.

In case of a violation, i.e. $V(t) > 1$, the multiplier is simply multiplied by the violation ratio:

$$\lambda(t) = \bar{\lambda}(t)V(t)$$

This way, the penalty for violating the MAL is strictly proportional to the size of the violation.

4.3 Eliminating Pro-Cyclicality

The key point to prevent cyclicality is the de-correlation between violations of risk measures and market events. This de-correlation will be achieved under two conditions:

1. The Regulator decides the stresses to apply, hence is in a position to smoothly impose deleveraging before it becomes an unsolvable problem. This is why the regulator must have an anticipative measure of factor risks and, in particular, of systemic risk.
2. Risk reporting is not a figure, but a function of markets hence violations are not due to markets swings but to misreporting of extreme risks. If institutions correctly report their extreme exposures, there is no reason why they would more violate their assessment during a crisis than during normal periods.

There is a chance that, even with such a setting, capital adequacy constraints still remain pro-cyclical if one of the following occurs:

a. The Regulator fails to anticipate systemic risks,
b. Institutions fail to correctly estimate their exposures to extreme market conditions

For these reasons, it is of utmost importance that the Regulator puts in place appropriate tools and analyses to cleverly monitor and update the list of imposed stress tests.

4.4 Other Risk Sources

It is in the interest of the Institution to foresee other risk sources within its risk reporting. These will add to the regulatory capital, but will avoid costly violations. The custom definition of "division" allows institutions to easily include extra risk sources in their reporting. For instance, in order to include Operational Risk, one can simply add a division supposed to entirely support this risk. The strength of this framework is to allow reporting not only a maximum amount, but an amount that may depend on external risk factors in a purely customary way. Let us here examine a few examples.

4.4.1 Operational Risk

The simplest way to include such a risk is to create a specific division D_{op}. Assuming operational risk cannot be related to any market factor, one will declare a fixed amount L_{op} which will be added to the overall MAL.

Operational losses, which are included in the actual loss $L_{act}(t)$, will be compared to this extra buffer in the MAL.

4.4.2 Counterparty Risk

Each major counterparty can be made a division. Losses stemming from the default of a counterparty are directly related to market events, in two ways: first by the amount of the engagement, second by the probability of default, which naturally depends on market conditions. It is in the interest of the Institution to estimate the reliability of its counterparty and to optimize the risk of violation and its cost in capital vs. its probability of occurrence. Using hedges, such as credit derivatives, or diversification strategies across several counterparties will be a direct consequence of this optimization.

4.4.3 Liquidity Risk

When the Institution deals with illiquid assets or even with assets exposed to liquidity risk, that is, assets that may suddenly become illiquid, such events materialize by a sharp price drop, which can be anticipated through nonlinear (optional) modelling. It is in the interest of the institution to anticipate such events and the market scenarios that may trigger them. If such scenarios are outside the bounds specified by the regulator, it might be optimal to include more extreme scenarios in the reporting, or to include scenarios rather specific to such or such asset class which were not in the Regulator's list.

4.4.4 Default Risk

The default risk on an asset is directly identifiable with a price drop. It is quite easy to anticipate as the impact of some market scenario and to include it into the risk computation.

5 Counterparty Risk: Monitoring "Too Connected to Fail"

One important teaching of the recent crisis is that market risk netting at the level of an entire institution is missing an important issue: counterparty risk. In fact, each major counterparty of the institution should be considered as a separate portfolio as, in case of default, the firm-wide netting will be destroyed.

For this reason, institutions should report their exposure to major counterparties in the same format as they do it for their own risk. In other words, a risk report of the same nature as that described above should be produced for each counterparty that accounts for a significant portion of the firm business. The “negative” part of the engagement with the counterparty (i.e. when the institution owes to the counterparty) is a simple market risk, to be aggregated with other market risks in the global market risk report. However, the “positive” part represents counterparty risk and should be subject to reporting in the form of stress tests as above.

The amount at risk, given by the worst case scenario, should be considered as a “loan” to the counterparty and be regulated as such with prudential rules. This being a risk figure and not a foreseeable amount with certainty, the Cooke ratio does not need to be applied in full. However, monitoring this potential risk with full knowledge of the market scenario in which it materializes will be crucial for the regulator to anticipate possible cascade effects. In fact, with such information, the Regulator will be able to run simulations and identify institutions that are “too connected to fail” as opposed to “too big to fail”.

6 A New Macroeconomic Lever

In the traditional Keynesian economy, the Government basically holds two macroeconomic levers in order to optimally monitor the country growth: short term interest rates and government spending. The third one—printing paper money—is not available to all governments (e.g. EU countries) and, generally speaking, is subject to constraints and must be used with extreme care.

Moreover, all these levers jointly act on the “financial world” and on the so-called “real economy”. Implicitly, the “financial world” is supposed to be in line with the “real economy”. Unfortunately, recent economic history has shown that the development of financial technology—both computerization and new financial instruments—allow strong discrepancies between the two. These are precisely those discrepancies which are targeted by regulation on capital adequacy.

The setting described here offers a new type of command lever to the Regulator. By allowing more or less regulatory capital to financial institutions, it can accurately monitor the general leverage of the system and discriminatively act on the financial world without touching the “real economy”, that is the industrial corporations, hence preventing too large discrepancies as observed during speculative bubbles.

7 Conclusion

The methodology presented here for capital adequacy rules is a strong framework for regulators to improve the current Basel II principles and, in particular, avoid pro-cyclicality without incurring uncontrollable costs for the Regulator to monitor the risks of financial institutions. This is achieved by applying the following principles:

1. The Regulator defines which market scenarios should be anticipated and tested. By its *anticipative* action, it avoids pro-cyclicality.
2. The Institutions provide their own risk estimate in each of these scenarios. The Regulator does not need to control the details of each institution, avoiding the corresponding costs.
3. The regulator checks that risks are correctly reported by comparing actual losses to an amount which depends on both the risk report of the institution and the actual market move. Institutions must correctly declare the potential impact of markets, but are not required to actually anticipate markets themselves. This task is left to the Regulator.
4. The Regulator penalizes institutions proportionally to the *amount* of the violation and not with respect to their frequency only. This puts a responsibility on the Institutions to provide *readable* risk figures and not abstract numbers potentially not in relation with actual losses.

Thanks to these rules, one can establish a healthy operating framework in which Institutions and Regulators keep confidence in each other. Institutions are naturally led to report risks as exhaustively as possible and to avoid "putting the dust under the carpet". Moreover, the extra safety gained by the approach will probably allow a reduction of capital for a number of healthy institutions that were penalized by the hazardous activity of others, which forced the Regulator to increase margins when this was not absolutely necessary.

Finally, the definition of stress scenarios used for computing the regulatory capital is a new command lever for preventing speculative bubbles, by accurately acting on the appropriate asset class and, therefore, avoiding systemic risk.

Acknowledgments Ilija Zovko, Marc Gross, Riskdata team, Myron Scholes for contribution and fruitful discussions

Financial Modelling and Memory: Mathematical System

Charles S. Tapiero and Pierre Vallois

1 Introduction

Data based search for patterns are of concern to almost all fields of studies and research. To do so, parametric mathematical models are constructed to explain observed data sets, forecast future prices, etc. The models we use are extremely varied, seeking to be reliable, robust and explanatory. Financial models for example have for the most part assumed models that are based on drift and randomness to construct models of asset prices. Typical examples include random walks and stochastic (Brownian motion) differential equations (such as a lognormal process) as well as Poisson and Jump stochastic processes. These latter models are based on events that occur at random and independent times. They assume also a number of simplifying assumptions, all of which are either explicit or implicit. For example, stochastic differential equations assume that data can be expressed in terms of two statistical properties: the data drift and its volatility, both of which are defined in terms of underlying (Brownian motion) normally and statistically distributed events. By the same token, Poisson-like jump processes are defined in terms of a memory-less process (with independent inter-events time distributions). To circumvent some of their assumptions, more complex and multi-variable models are used to account for observations that such models fail to explain. For the most part these models are based on both a rationality and an experience acquired based on theoretical constructs and on data analyses. These models are extremely useful, and provide an ex-ante interpretation for the behavior of data sets as well define the statistical properties of observed financial variables. Ex-post, however, these models may default since all

C. S. Tapiero (✉)
Department of Finance and Risk Engineering, New York University Polytechnic School of Engineering, Brooklyn, New York, USA
e-mail: cst262@nyu.edu

P. Vallois
Institut de Mathematiques Elie Cartan—UMR7502, University of Lorraine, CNRS, INRIA, B. P. 239, 54506 Vandoeuvre-les-Nancy Cedex, France
e-mail: vallois@iecn.u-nancy.fr

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance*,
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_6,

models are merely an educated hypothesis of underlying processes. Modeling financial processes is therefore a work in process, in search for coherent and complete mathematical systems that can on the one hand be justified theoretically and on the other account far more precisely for observed data sets.

The first and fundamental mathematical system underlying financial dundamental theories is the definition of a pricing probability measures that defines a Martingale on the basis of which future prices can be priced based on observable current prices. Technically such approaches are based on the translation of the underlying "noise" of the financial process. The assumptions of such models are based on economic rationalities. Their extensions to more complex and generalized mathematical systems include fractional calculus as well as other systems where both the time scale of underlying stochastic processes as well underlying "noise processes" differ from the Brownian motion. These elements provide both theoretical challenges to economic and financial rationalities and at the same time provide a greater number of parameters we use to explain the bahavior of data sets.

This paper provides an elementary introduction to such systems for financial engineers seeking to expand financial modeling. Necessarily, these models are based on mathematical assumptions which will be defined explicitly. Financial modeling consists then in seeking an economic rationality based on the least number of parameters which is cogent with observed financial facts and financial agents behaviors.

For example, the traditional assumption that stocks' rates of returns increments $\Delta R(t)$ and $\Delta R(s)$ are stationary and independent have far reaching implications to pricing models. Their simplicity and limitations are compensated however by their usefulness in defining and "pricing" financial assets that agents are accepting as reference prices to trade with. Such an assumption has specific and important implications. Explicitly, consider the mathematical function $f(t) = E(R(t)) - E(R(0))$ at two instants of time t and s, then independence implies a functional relationship: $f(t+s) = f(t) + f(s)$ where $f(t+s) = E(R(t+s) - R(t)) + E(R(t) - R(0))$ and therefore, it equals $E(R(s) - R(0)) + E(R(t) - R(0)) = f(t) + f(s)$. This functional equation has a unique time linear solution $f(t) = tf(1)$ which proves that such models have time linear expectations. A similar analysis for its second moment provides a similar time linearity. However, such assumptions are not always confirmed by assets time series even though they underlie fundamental financial models. For example, efficient markets, defined by Fama 1970, as markets in which information is instantly reflected in the market price and reflect "all the information" that is relevant to that stock—past and future, presume such a linearity (see also Fama's earlier papers, 1963, 1965a, 1965b). In this framework, the future is model based on a predictable future of state and their equilibrium prices. Each of these predictable prices is implicit in traders exchanges occurring at a present time and result in an equilbrium for all (known) future state prices. Their current expectation (with respect to the probability measure that define their prices), thus defines (imply) the current future price. A weak efficiency, however, is defined when the current price reflects past prices only. This means that, provided all the past information is used (which we will denote for the moment and for simplicity by the filtration $\Im(t)$) accounting for what is known at time t, a market is weak efficient if its expected price conditioned by this information equals the current price. Thus a price at time time t, $p(t)$, for a future asset price at time $t+T$, $p(t+T)$, obeys with respect to a probability measure Q, given by a financial model definition by:

$$p(t) = E^Q(p(t+1)\,|\Im(t)) = \cdots = E^Q(p(t+T)\,|\Im(t))$$

Markets efficiency (assuming that it exists) has assumed an extremely important role in theoretical finance. When markets are complete, the Martingale is a mathematical property that underlies financial asset pricing models. Data as well as complex financial markets point out however that markets may often be incomplete. In economics, interest in such market models arose after the publication of a paper by Charles Nelson and Charles Plosser 1982 on “Trends and Random Walks in Macroeconomic Time Series”. Based on the study of a very large number of time series, they were not able to confirm that “noise” is normal (which is often stated as the unit root hypothesis). This means either that economic processes are inherently non-stationary and thus resist explanations by linear models such as ARIMA or related models, or that there are long range memory effects to economic time series which are not accounted for by simple time linear Markov models. These observations still contribute to our search for time series models that can better explain business cycles, underlying volatility processes and in general better explain the evolution of prices.

Expectations, forecasting, data extrapolation, intrapolation, memory, filtering and re-modelling with far more complex models (multi-variables, nonlinear, etc.) are thus sought to reconcile theoretical models and observed data. For example, past and current data are used to estimate volatility processes based on both historical data (to estimate the historical volatility) as well as on implied future prices (to estimate an implied volatility). By the same token, numerous dynamical models are based on a model of memory. For example, consider a mean reverting stochastic model (an Ornstein-Uhlenbeck process) with an initial condition $x(0) = 0$, where $\{x, t > 0\}$ is a non-adaptive stochastic process deviating from its mean evolution and $W(\tau)$ is a standard Brownian motion, or:

$$dx = -ax\,dt + \sigma dW(t), x(0) = 0 \text{ and its integral } x(t) = \sigma \int_0^t e^{-a(t-\tau)} dW(\tau)$$

This process has a (statistical) exponential decay memory, weighting more importantly recent random deviations from their mean. Alternatively, it is an exponentially decreasing function of past “normal noises”. The weight of memory is then model based. Generally,

$$x(t) = x(0)e^{-at} + \sigma \int_0^t e^{-a(t-\tau)} dW(\tau), x(0) \neq 0$$

In this case,

$$E(x(t)) = E(x(0))e^{-at}$$

$$\operatorname{var}(x(t)) = E\left\{(x(0) - E(x(0)))e^{-at} + \sigma \int_0^t e^{-a(t-\tau)} dW(\tau)\right\}^2$$

If the initial condition is not anticipating as well (or equivalently, not adapted), it is possible to write:

$$\operatorname{var}(x(t)) = \operatorname{var}(x(0))e^{-at} + \sigma^2 \int_0^t e^{-2a(t-\tau)} d\tau = \frac{\sigma^2}{2a} + \left(\operatorname{var}(x(0)) - \frac{\sigma^2}{2a}\right) e^{-at}$$

This elementary memory model can be extended in numerous ways and functionally transformed may lead to an interpretation of memory and its uses in different ways. As we shall subsequently see, such a memory may be conceived and defined by probability models such as Gamma and Weibull distributions as well as long run memory trends with fat tails distributions and a memory defined by a calculus based on the Riemann-Liouville derivative definition of a fractional derivative. These models on the one hand question the precision of using limit arguments of very small time intervals between events (can such intervals be parameterized to be greater than the limit dt?). On the other, are noise processes fuelled by Brownian motion or by some more elaborate processes? The particularity of these questions underlies the development of alternative mathematical systems to better appreciate and manage financial processes.

A preliminary extension of this model leads to financial models where the weighting of "past noises is some function $m(t-\tau)$, while others are based on the definition of the stochastic differential equation and its integral and still others are based on the definition of the "noise" itself. With a normal (Brownian) noise, a premium drift added to the Brownian motion will provide as we shall see, a probability measure that defines a pricing model. A variety of mathematical systems are thus possible, each extension providing a generalization for the fundamental OU normal process. For example, consider the following model

$$\tilde{x}(t) = x_0 m(t) + \sigma \int_0^t m(t-\tau) dW(\tau) \quad \text{with} \quad m(0) = 1$$

And let the stochastic integral equality below be defined by an application of Ito's definition of a stochastic integral:

$$\int_0^t h(t,\tau) dW(\tau) = h(t,t)W(t) - \int_0^t \frac{\partial h(t,\tau)}{\partial \tau} W(\tau) d\tau$$

And therefore,

$$\tilde{x}(t) = x_0 m(t) + \sigma \left(m(0)W(t) - \int_0^t \frac{\partial m(t-\tau)}{\partial \tau} W(\tau) d\tau \right)$$

Of course if $m(t) = e^{-\alpha t}$ then $m(t) = e^{-\alpha t}$ and $\frac{\partial m(t-\tau)}{\partial \tau} = \frac{\partial}{\partial \tau} e^{-\alpha(t-\tau)} = \alpha e^{-\alpha(t-\tau)}$ and therefore,

$$\tilde{x}(t) = x_0 e^{-\alpha t} + \sigma \left(W(t) - \alpha \int_0^t e^{-\alpha(t-\tau)} W(\tau) d\tau \right)$$

1.1 The Langevin Representation

Consider again, the Ornstein-Uhlenbeck process, equivalently written as follows:

$$dx = -axdt + \sigma w(t)\sqrt{dt}, x(0) = 0 \text{ and } w(t)\sqrt{dt} = dW(t)$$

Note that since $\operatorname{var}(dW(t)) = dt, \operatorname{var}(w(t)\sqrt{dt}) = dt \operatorname{var}(w(t))$ and therefore, $\operatorname{var}(w(t)) = 1$. The equation is due to Langevin, which we can write as follows:

$$dx = -axdt + \sigma \left(w(t)(dt)^{\frac{1}{2}} \right), \; x(0) \text{ given}$$

Such a formulation renders explicit the different time scales of the drift and the noise component due to the special considerations one has to maintain when calculating a stochastic integral. Explicitly since $W(t)$ is a normally distributed random variable, then integration of $\Delta W(t)$ resulting in $W(t)$ is defined computationally by its integration choice, its time interval and how the stochastic integral is numerically defined.

These comments are in fact the elements that distinguish between the Ito stochastic calculus and the Stratonovich (in fact Riemanian) calculus. Further, the choice of the limit time interval Δt in the stochastic calculus we choose to apply, defines as well financial models that may have different interpretations. For example, consider the Ornstein-Uhlenbeck model and assume that the time interval is parameterized to $\Delta t = (\Delta t)^H$ where H is a parameter. Of course if $H = 1$, that would be equivalent to a Langevin equation (or a Brownian motion). Namely,

$$\Delta^H x = -ax(\Delta t)^H + \sigma w(t)(\Delta t)^{H/2}, \; x(0) \text{ given}$$

And at the limit:

$$d^H x = -ax(dt)^H + \sigma w(t)(dt)^{H/2}, \; x(0) \text{ given}$$

In other words, by rescaling the stochastic differential calculation we have created another mathematical system with its own rules and its own analytical procedures. The integration of such an equation does not lead to the same result its Ito calculus model indicates, although we shall show later on that there is an equivalence between these two formulations. In fact, we can interpret the H fractional formulation as a generalization of the OU process since for H = 1, we have the standard OU Brownian mean reverting process. In general, we shall see that a fractional calculus is based on a mathematical operators based on a Cauchy-Riemann-Liouville equation as well as on the use of a Mittag-Leffler function (a generalization of an exponential function) as we shall see subsequently. These operators, while mathematically consistent and inclusive of the Riemann calculus alter some of the calculations we use to interpret financial models based on Ito stochastic calculus as well as based on the use of the Brownian motion as an essential source of risk-noise.

1.2 A Poisson Noise

Let noise be defined by a Poisson probability distribution (as we shall subsequently see), then such a process will generate another family of financial models. It is then defined by *stationary independent* (memory-less) *increments* which are *identically distributed as a Poisson probability distribution*. That is, for increment Δx_t, we have $E(\Delta x_t) = \lambda \Delta t$ where λ is the mean rate of a Poisson distribution:

$$P(\Delta x_t = n) = \exp(-\lambda \Delta t)(\lambda \Delta t)^n / n!, \mathrm{n} = 0, 1, 2, 3, ...$$

The probability distribution is in this case a function of the time interval Δt. That is, setting, $(\Delta t)^\alpha$, we have then: $P((\Delta x_t^\alpha) = n) = \exp(-\lambda_\alpha (\Delta t)^\alpha)(\lambda_\alpha^n (\Delta t)^{\alpha n})/n!, \mathrm{n} = 0, 1, 2, 3, ...$ where we note explicitly that in a time interval $(\Delta t)^\alpha$, events rates associated to this time interval λ_α, define a probability for the number of events within that time interval. In this case,

$$P((\Delta x_t^\alpha) = 0) = \exp(-\lambda_\alpha (\Delta t)^\alpha) \approx 1 - \lambda_\alpha (\Delta t)^\alpha + \frac{1}{2}(\lambda_\alpha)^2 (\Delta t)^{2\alpha} + ...$$

If $\alpha > 1$ then of course, second order terms are negligible if the time interval is very small. If $(\Delta t) > 1$, then evidently, second order terms will be increasing providing a greater importance to our accounting for these terms, etc. In other words, the time scale of a a model (whether for counting insurance or financial events) does alter the underlying assumption we make regarding such a model. Of course, if $\alpha = 1$ we recuperate the common Poisson distribution which acts in this case as an anchor distribution with respect to which computational and theoretical results implied by the model and the time scale we use can be evaluated.

An alternative representation of a common Poisson process as a time continuous stochastic differential equations consists in the following:

$$\Delta x_t = \lambda \Delta t + \sigma \tilde{\mu}(\Delta t) \;\; \text{or} \;\; x_t = \lambda t + \sigma \tilde{N}_q(t)$$

Where $\tilde{N}_q(t)$ is the sum of a standardized Poisson events, with $\tilde{\mu}(\Delta t)$ in a time interval (0, t). Such standardization allows us to construct an approximation to that of a continuous state (Brownian process). If $\lambda = \sigma^2$, the mean equals the variance and at least the moments of such a process are equivalent to that of a Poisson process (although, these are not generated in the same way and therefore they may not be a Poisson process but its approximation). To standardize $\tilde{\mu}(\Delta t)$ as a Poisson noise approximating a Brownian motion model, we let $\tilde{\mu}(\Delta t)$ be a zero mean and unit variance process:

$$\tilde{\mu}(\Delta t) = \mu(\Delta t) - \Delta t \text{ with}$$

$E(\widetilde{\mu}(\Delta t)) = 0$ and $\mu(\Delta t) = \Delta t$ and var $(\widetilde{\mu}(\Delta t)) = \text{var}\,(\widetilde{\mu}(\Delta t)) = \Delta t$. $\mu(\Delta t)$ is thus a random variable with parameter Δt with $P(\mu(\Delta t) = 1) = \Delta t + o(\Delta t)$ and $P(\mu(\Delta t) = 0) = 1 - \Delta t + o(\Delta t)$ while all other values have negligible probabilities,

$$\mu(\Delta t) = \begin{cases} \text{Prob}[\mu(\Delta t) = 0] = 1 - \Delta t + 0(\Delta t) \\ \text{Prob}[\mu(\Delta t) = 1] = \Delta t + 0(\Delta t) \\ \text{Prob}[\mu(\Delta t) \geq 2] = 0(\Delta t) \end{cases}$$

with $E\{\mu(\Delta t)\} = 0$ and var $\{\mu(\Delta t)\} = \Delta t$. When Δt becomes very small, the stochastic Poisson process is formally defined by a stochastic differential equation with noise $\tilde{\mu}(dt)$;

$$dx_t = \lambda dt + \sigma \tilde{\mu}(dt); \sigma^2 = \lambda$$

Evidently, $\tilde{\mu}(dt)$ represents now a discrete state process, taking on values of zero and one only. If $\sigma^2 \neq \lambda$ the process defined above may approximate a hyper or super Poisson process in a mean-variance sense. Again, in a mean-variance sense, it may be approximated by the continuous Wiener process, which assumes the standard form:

$$dx_t = \lambda dt + \sigma dW(t), x_0$$

Our use of the underlying Bernoulli "noise" process over a given time interval assuming binary values can of course be generalized in many different ways based on its moments. However, changing the time scale, the assumption that only one event may occur in a given time interval (however small) can be misleading. Noise models resulting from the development of "Bernoulli" models may be co-dependent (as Bernoulli events can in fact be statistically dependent, be causal, etc.). For example, does an event, producing "information" useful for financial purposes alter the probability of a future event (information). Do such events have memory in the sense that their time record may alter financial decisions and thereby lead to "other" evolutions of market price processes sensitive to such memory-information.

The representation of a discrete (Bernoulli) random process by a continuous one (Poisson, diffusion approximation) is necessarily a mathematical convenience, which is a model of a far more complex reality. In this sense, financial models are essentially "simple models" of a financial reality while the real evolution of financial prices is far more complex than presumed by such models. For example, as indicated above, redefining the time scale to say $\tilde{\mu}((dt)^{\alpha})$ we will necessarily redefine the underlying process-model, we shall call for convenience, fractional model due to the change effected on the model time scale. In this case, note that if $\alpha < 1$, then as the limit of Δt, tending to zero, has $(\Delta t)^{\alpha}$ tend to such a limit is a slower manner (and therefore mathematically, limit arguments have to be treated more carefully). For example, let $\alpha = 0.1$, then: $(\Delta t)^{k\alpha}$ is a term greater than Δt for k = 1,2,3,... 9 and therefore neglecting higher order terms may be more difficult to neglect. In this case, as stated above, consider terms of order up to $2\alpha < 1$, and note that:

$$
\begin{aligned}
&P(\Delta x_t = 0) = \exp\left(-\lambda(\Delta t)^{\alpha}\right) = 1 - \lambda(\Delta t)^{\alpha} + \frac{1}{2}\lambda^2(\Delta t)^{2\alpha} \\
&P(\Delta x_t = 1) = \left(1 - \lambda(\Delta t)^{\alpha} + \frac{1}{2}\lambda^2(\Delta t)^{2\alpha}\right)(\lambda(\Delta t)^{\alpha}) = \lambda(\Delta t)^{\alpha} - \lambda^2(\Delta t)^{2\alpha} \\
&P(\Delta x_t = 2) = \lambda^2(\Delta t)^{2\alpha}
\end{aligned}
$$

with $P(\Delta x_t = 0) + P(\Delta x_t = 1) + P(\Delta x_t = 2) = 1 + \frac{1}{2}\lambda^2(\Delta t)^{2\alpha}$. However if $1 - \lambda(\Delta t)^{\alpha} >> \frac{1}{2}\lambda^2(\Delta t)^{2\alpha}$ then,

$$
\begin{aligned}
&P(\Delta x_t = 0) = 1 - \lambda(\Delta t)^{\alpha} \\
&P(\Delta x_t = 1) = \lambda(\Delta t)^{\alpha} - \lambda^2(\Delta t)^{2\alpha} \\
&P(\Delta x_t = 2) = \lambda^2(\Delta t)^{2\alpha}
\end{aligned}
$$

And their sum equals one, which yields a process:

$$
\mu\left(\Delta t^{\alpha}\right) = \begin{array}{l}
\text{Prob}\left[\mu\left(\Delta t^{\alpha}\right) = 0\right] = 1 - \lambda(\Delta t)^{\alpha} + 0\left(\Delta t^{k\alpha}\right), k > 3 \\
\text{Prob}\left[\mu\left(\Delta t^{\alpha}\right) = 1\right] = \lambda(\Delta t)^{\alpha} - \lambda^2(\Delta t)^{2\alpha} + 0\left(\Delta t^{k\alpha}\right), k > 3 \\
\text{Prob}\left[\mu\left(\Delta t^{\alpha}\right) = 2\right] = \lambda^2(\Delta t)^{2\alpha} + 0\left(\Delta t^{k\alpha}\right), k > 3 \\
\text{Prob}\left[\mu\left(\Delta t^{\alpha}\right) \geq 3\right] = 0\left(\Delta t^{k\alpha}\right), k > 3
\end{array}
$$

which leads to a fractional Poisson model. Physically, it means that when we increase the time interval the number of events we have to consider is increasingly larger and further, the fundamental assumption of events independence (which is generated by considering one event at a time with independent probabilities) may be misleading. We shall subsequently see that this is the case when we consider fractional Poisson processes.

Other financial systems do exist, each expressing various rationalities or parametric generalizations of specific reference models. For example, maximizing entropy (a measure-metric for disorder) is used in some cases to justify the adoption of noise models (or explicit probability distributions that meet certain informational assumptions) that in some specific cases include the standard (and financial model-conventional) normal probability distribution. In this sense, informational departures from what leads to such a distribution may be understood to provide a measurement for the effects of information that leads to selecting instead of the normal, another maximum entropy consistent probability distribution. By the same token, instead of using and "informational rationality", one may use a behavioral rationality based on assumptions regarding human preferences and risk attitudes (as it is the case when using a utility theory to define a financial process through say, the CPM and the CCAPM type models).

Time, temporal scale, memory, an underlying informational and "behavioural" rationality, noise and their mathematical representations and treatments thus encompass a broad approach to financial modelling. These approaches result for example, in the recognition that mean-variance models may be limiting (such as CAPM models or lognormal stochastic processes based on rates of returns and volatility statistics). Thus, leading a search for financial stochastic models to be far more coherent with skew, kurtosis and other statistical characteristics observed in real data. Such a pursuit, leads necessarily to a departure from the tenets of fundamental finance theories such as the Arrow-Debreu framework for financial assets pricing. Such a theoretical framework assumes in fact (among various assumptions regarding informational and financial rationalities) that future financial state preferences are completely known (and therefore it is called complete markets finance). Extending financial models to be amenable to empirical testing and yet be expressed in terms of an Arrow-Debreu theory of a deterministic financial equilibrium financial rationality underlies complete markets financial modeling. Deviating from any of its assumptions is defined in terms of ill-defined market prices disequilibrium, which we call globally, incomplete markets. Even though, each departure from such a theory provides in fact a different market "incompleteness".

In this sense, Brownian noise based on prevalent mean and volatility processes may be too restrictive. Financial data as stated above often exhibit higher order moments pointing to skewness (asymmetric distribution) as well as to kurtosis (fat tails). In this case, fractional (time scaling) noise processes (whether normal based or Poisson based) alter in a specific way the manner with which information is defined and the values financial models assume as a function of the time scale the financial model uses. These elements, implied in both the manner we construct models and the manners in which we calculate the evolution of models, calculate derivatives, stochastic models and stochastic integrals, etc. alter financial calculus.

For example, we commonly use a Brownian (normal) noise process to represent independent noise events, lacking any underlying rationality and purpose, occurring randomly and then proceed to assess their temporal effects by integrating their past random outcomes into an integrated whole. Such models are non-anticipating, in the sense that future "noises" are not integrated and are independent of current noise. This is of course, justified by the fact that "fundamentally", the future "noise" of prices is deemed "unpredictable".

Computational finance is also, necessarily, models dependent. In financial engineering, much use is made of financial models based on Ito's calculus (rather than, say Stratonovich stochastic calculus, both differentiated by their approach to computing stochastic integrals as we shall subsequently see). I believe that this is due to finance's risk measurements to be particularly focused on returns and their volatility, defined by Brownian motion models and less with other moments (although in practice, these are elements of increased importance). This is changing however, as financial models are increasingly more sensitive to returns skewness (distributions' asymmetry) and to their kurtosis (fat tails). Levy processes we shall outline briefly are attracting and increasing interest for financial modeling due to their expanding the moments they consider and yet maintain the (statistical) stability of the probability distributions it uses.

Fractional or "Long run memory" models, use instead a calculus based on the Cauchy-Riemann-Liouville equation and the Mittag-Leffler function. In this paper, some of these elements will be introduced based on an extensive bibliography and research spanning mathematics, physics, psychology and of course, economics and finance.

Finally, the manner in which we interpret data and relate it to a future time is also an important issue to reckon with. In particular we shall refer to "short memory" to account for the effects of a specific event on the subsequent probability of a future event. This memory is then the event itself or the past few events. A shift of model to model is then causal. It may be also be a statistical realization of the past which alters a financial (or any other) subsequent process causality is the a statistical event. When such a process occurs once and does not predictably recur, it is then a "bifurcation"—an event changing a process irrevocably. When an event conditionally alters a process in well-defined ways, then of course, each time such an event occurs, such a process bifurcates and change occurs predictably (it is not a bifurcation however). We shall call such processes short memory or persistent processes.

The empirical study and estimation of such models are evidently model-dependent. For example, learning (Bayesian or using other statistical tools) may be applied. These are issues that this paper will not for brevity, attend to. Generally, financial models are constructed to replicate a specific element or problem of real finance. Fundamental finance based on the Arrow-Debreu framework seeks to do by essentially defining a probability measure that ascribes to future prices a current price. Its usefulness is based on a transparent rationality that justifies our belief in its predictions. Its weaknesses, as with all models, are embedded in their assumptions. Their advantage are based (in finance at least) on their acceptance to support trade and investment decisions as well as their tractability. When its assumptions are met, markets are called complete (or efficient) and when they are not, they are called incomplete. In this latter case, financial models are sought that can enrich the basic complete markets framework to account for financial observations that better fit observed behaviors and financial time series. Mathematical "systems" may then be applied, each with characteristics that may differ. Complete markets pricing models do so by accounting for the risk premium of say a stock price process, which is used to "transform" the stock to be financially equivalent to another asset whose price is known for sure in both the present and in its future. In this sense, complete market pricing models are a relative price construct (rather than a real price) with specific mathematical properties that define the evolution of theoretical asset prices (defined in terms of a Martingale as we shall see subsequently).

Fractional (long-run) memory generalizes fundamental models (in the sense that it recoups the fundamental model as a special case) and at the same time desensitizes the application of its calculus to extreme limit times differentials (i.e. the time scale is parameterized). Its application is not obvious however, not for the mathematics it uses, but for the finance it implies. Its parametric generalization provides however a measure when deviating from a reference (complete market) financial model. By the same token, q-calculus proceeds in a similar manner, albeit with another generalization (see Jackson calculus as well as applications of maximum Boltzmann-Gibbs 1878 as well as Tsallis and other functional generalizations of entropy functions and their rationalities). Differences are also observed in the manner in which stochastic noise is defined, interpreted and calculated. These and many other approaches one might define provide foundations to analytical approaches to financial modelling and analysis.

The intent of this paper provides an overview of such models as well as uses problems to assess the effects of such models on some basic results in finance we take for granted. In this sense, it brings no new results, and replicate some results published by other authors including Jumarie. Vallois, Herrmann, Oren Tapiero, my own past publications and others, to which I am most grateful for their permission to use their publications. Numerous papers are listed in references to provide for the motivated reader a background and sources for further study. Finally, financial examples are used to highlight the relevance and the potential usefulness of various mathematical and memory-information systems for financial modelling. For simplicity, we use both fixed income-bond models as well as lognormal stochastic price models extended in different manners according to the mathematical system they use. The intent of these applications are to challenge both computational approaches in finance and fundamental financial theories.

2 Time and Memory

What is memory and what is time? These are two intrinsically and fundamental concepts that underlie mathematical modeling in general and financial modeling in particular. Time is often defined implicitly in terms of religious and cultural values to seek a relative line (a time line) to organize our beliefs and thoughts, our data and information to construct a model of memory and confirm theories. Dynamic models such as asset price processes are merely models sought to organize a hypothetical evolution of prices along a given time line and based on the interpretation—objective or subjective, statistical or punctual of past events—which we call memory. Extending this time line beyond a current time provides a means to forecast future events and future prices. Time is defined by fractioning a temporal perspective into a sequence of points, denoted by the year, the month, the week, the day, the hours, the seconds and the microseconds. In this context, time can be objective measuring time by a "clock" while it can be relative to a set of measured or memorized events. A time line is thus a scale against which we measure a theoretical or a "psychological" process. It may have different properties, explanatory and predictive powers and each embeds past events and by their memory they can affect attitudes, decisions and future predictions. The greater the time line interval, the more a memory is embedded (endogenized) in the evolution of the processes' parameters it seeks to represent and thus may not be apparent. For example, mathematical models in high frequency trading may be based on and account for many considerations that are dissipated when they are integrated (endonenized) in financial day data. Thus, models based on monthly data, day data or on High-Frequency data can differ appreciably when a stochastic model is constructed, yet they may be in some rational or statistical way be dependent—one partly explaining the other and vice versa. Co-dependence of temporal events may be integrated in parameters that define a process trend by the effects that one has over the other. Similarly, predicted future prices may be endogenous to a current price (based on the assumption that some expectations are self-fulfilling). These elements are of course challenging when estimating and interpreting financial stochastic models. For example, in fundamental finance option prices may be defined based on historical estimations of volatility (and therefore, memory based) or based on a future expectation of future state prices (under a specific probability measure) that are merely a current manifestations of prices traded with a future intent. Validation of such an approach presumes that such estimates are independent of past data and therefore past memory. The past, however, may or may not be relevant to such expectations, requiring that such models (hypotheses) be carefully assessed. Further, in some models (as with short memory), we can show that memory, while based on specific events may subsequently be implicit in processes trends and volatility. Modeling memory and defining its time scales is therefore both a challenging modeling and analytical problem. In other words, real-information may alter future trajectories and therefore the models we a-priori hypothesize.

Models' challenges may be circumvented by mathematical transformation of the data we have, by a dimensional expansion of the underlying model and by defining more specifically what memory we "mean". How future states and expectations are applied to define their present consequence, is the underlying function of pricing models.

"Memory", unlike time (but temporally dependent), represents quantitatively the effects of past states on current and future ones. Memory can assume then many forms. For example, it may be interpreted as statistical estimates of past and future events, often defined in terms of probabilities that we use to elucidate future events (as it is the case in Bayesian probability models that adapt probability estimates by a Bayesian learning model). In such cases, memory, current data and its reliability are used to learn and exert an "information transfer" to infer a future probability. These assume that the past and the future are "temporally dependent"—either in fact as observed by data or by the definition of a model we construct. Namely, since all models are merely a partial representation of reality, defined by design or implied in data sets we use to construct such models, time and memory may simply be tools we use to augment the comfort of our predictions. Technically, we then use terms such as temporal dependence, auto-correlations as well as short and long run memory to define the elements a process is based on with its parameters estimated as information unfolds.

"Independence" as opposed to "dependence" is used to indicate that events vary statistically and independently of one another. Namely, observing or modeling an underlying set of events that are independent will presume that they are "timeless"—i.e. the occurrence of an event at one time is independent of another at another time. Suppose that the underlying statistical elements of a process are used to represent an evolution which is temporally independent. Then temporality is a term assumed a-priori to build a statistical model. Time and memory (in its broader sense, of past and future) are thus intrinsically related.

Examples abound. Financial models use profusely the"Markovian" property, where future events are defined conditionally on a process current state, ("representing a point memory of everything that has happened before"). For example, momentum trading strategies are based on the belief that a stock price increase may imply a future increase, and vice versa. Mean reversion is then a momentum strategy but expressed in terms of deviation from an underlying long run trend. When above the trend, the tendency would be to decline and return to be in the trend and vice versa. Charting in finance practice is of course memory based as are many trading techniques on empirical experience and traders' beliefs.

Fundamental finance theory is also based on memory, but formally representing all future expected results in terms of a filtration with respect to a theoretical and sets of past data. There are further, many financial models seeking to depart from the Markovian assumption in order to explain contagion processes (of banks of default loans, etc.) or using complex models where the underlying source of risk is a random (or Markovian) noise. Information or "its manifestation in memory" is a transfer from a past to a future which assumes of course many forms and in particular unpredictable ones. A rolling ball is hit by another—its trajectory will necessarily be altered and therefore valuation models adapt to such changes. Similarly, when interests are changed by the Central Bank from say 3–3.1 %, prices, future predictions of interest rates are altered. By the same token, interventions by Sovereign agents on FX markets produce information regarding States intents' which also alters future expectations and the evolution of FX rates. etc. Memory may also linger over long periods of time or may be curtailed, with only recent events and information dominating the evolution of financial processes. Both long (with significant autocorrelation) and short memory assume then many forms, some expressed by a nonlinearly growing volatility over time to point out the fact, that we can hardly predict the future, while others alter future trajectories (or Short Run Memory Models). Both Long and Short Run memory models are manifest in and out of finance (Baddeley 1994, 2003; Atkinson and Shiffrin1968; Miller 1956; Tapiero 1988, 2004, 2010, 2012a, Tapiero on adverting, its content and forgetting (1975, 1978, 1979, 1982, 1983, 2004, 2005).

In car insurance –the claim history of an insured driver has an effect on the probability of the driver having future accidents (since an accident may be learned by the driver that in the future will be more careful and seek to prevent such accidents). In intraday financial data, as well as in High Frequency Trading (HFT), algorithms are constructed to profit from memory which may or may not always exist. Momentum trading strategies for example, assume explicitly that an increasing stock price may be self-enforcing and thereby lead to future growth in its price, and inversely. Memory was shown on the S&P, IBM and other securities to have a memory effect on the future probabilities of prices lasting 1–5 min in intraday data. Over longer periods of time, the memory is abated rapidly. In fact, a theoretical model we constructed (Vallois and Tapiero 2007, 2008) has pointed out explicitly, that lingering memory effects are embedded in our estimate of process trends and its volatility. Lacking appropriate models to represent the "information transfer from past to future", memory is embedded in an uncertainty defined by distributions' volatility, skewness and in general probability processes that are "nonlinear" and more complex than normal distributions.

For financial theories, a fundamental question is: Do stocks have memory? If yes, then theoretical models we use to price financial assets (based on Martingale price processes) are misleading (although a Martingale is a mathematical construct). The fundamental theory of finance answers this question by a categorical no. Yet, forecasts are based on past data and past experience as well as current future prices (of forward and options' prices) that are implicitly used in models with no memory. Such a categorical no is further challenged by the fact that financial statistics point out to skewness, kurtosis (fat tails), to stochastic volatility as well as to jumps—some of which are known in probability and some are not known. The existence of memory is further compounded by its definition and its manifestations as stated for example below.

- Memory is a "model" to construct a temporal causality that defines the occurrence of future events (stock prices) and their likelihood-probabilities (prices, interest rates etc.). Thus, better understand and use past information and future expectations to predict the evolution of prices.
- The physical fact that we are always in a present has motivated our quest to define a virtual past dictating both our understanding of the present and attitudes towards the future. These underlie our acts in the face of a future which may be predictable or unpredictable. To frame a future in a cognitive manner, "expectations-models" and scenarios generating approaches are constructed based on experience, information, needs and attitudes—these result in a memory.

These are the factors that define our predictions and expectations of future events and prices.

In some cases, autocorrelation, expressing the dependence of stock prices over time, presume that memory is embedded in trends, abated by a forgetting rate or summarized by a statistical reckoning of past meaningful events. Some of these events are assessed by their deviation from normal expectations (for example outliers) or may be statistical summaries of statistical events. The existence of memory and how it manifests itself in underlying financial models is therefore important for both financial theoretical and practical reasons as well financial statistical and econometric reasons.

Financial risk models in fact are thus and often are, art rather than just a technique that consists in defining a meaningful and tractable model that reconciles best our understanding and observed data sets. For example, the mean reversion model is based (as its integral solution indicates) on the record of past and random events weighted exponentially. By the same token option prices are based on a model of future and rational expectation of future prices. On theoretical grounds, the mathematical validation of a financial theory based on complete markets presumes both the predictability of future state prices defined by a (memory-less) martingale (defined with respect to an appropriate probability measure). Such processes were suggested and used by Bachelier who defined an underlying random walk (and Brownian motion) price process (Bachelier 1900, 1901; Einstein 1905). These processes have special characteristics consisting of independent increments, independently and identically distributed as Gaussian (normal) random variables with mean zero and variance t leading time linear volatility growth models.

In all cases, financial models are present models. Namely, they are defined relative to one instant of time, with the past and the future integrated into a present instant. This idea has been stated clearly by Saint Augustine (Confessions, Book XI, xx.):

> … Yet perchance it might be properly said, "there be three times; a present of things past, a present of things present, and a present of things future." For these three do exist in some sort, in the soul, but otherwise I do not see them; present of things past, memory; present of things present, sight; present of things future, expectation.

Financial and quantitative memory models are therefore always in the present, explicitly stated in financial models by the conditionality of prices and estimates taken with respect to a filtration at the time predictions and estimates are made.

Our ability to relate the past and the future to one another other and vice versa—i.e. make sense of temporal change as well as make sense of current option prices on a future price volatility. For example, "remembering that stock markets behave cyclically" might induce a cyclical behavior of prices (which need not, of course, be the case). "Remembering" i.e. recording the claims history of an insured over the last years may be used to determine a premium payment schedule and therefore the probabilities of future claims. The "health" history of a patient might provide important clues to determining the probabilities of his survival. However, while in the past such models were constructed on limited and memorized data, leading to an extensive and competing number of models, currently, clouds, big data, data streams and diverse data sources and richness, etc. provide new opportunities to reassess using new and practical models, uses of extremely large and deep data, and the memory they provide.

Memory models are also based on psychology. Persons do hear and learn and remember some, but not all. Some is forgotten almost immediately and some is remembered "forever". James Miller in 1956 (the Magical Number 7 plus or Minus 3), for example emphasized that the human capacity to process information and remember it is small and therefore, information is necessarily lost (see also Baddeley 1994, 2003). A system model distinguishing between short term memory, long term memory and sensory memory was suggested by Atkinson and Sheffrin (1968) and expanded further in numerous studies (for example, Baddeley 2003). These studies are particularly important in marketing where consumers' recall of past advertising messages as well as their forgetting are important (see Tapiero's advertising and repeat purchases related papers, 1975–2005).

Potential models we can consider are many and varied. Mean reversion and Log-normal models profusely used in different forms in financial modeling are however representative of how we might begin. While such models are mathematical stochastic process, prices are the outcomes of an exchanges between traders, investors and financial agents in general. This model (which is in fact non-stochastic in a conventional way) therefore, results in a pricing model that under very specific assumption conforms to a fundamental economic rationality as defined by a general equilibrium theory in economics and extended to future markets by Arrow and Debreu.

3 Models of Memory and Brownian Mathematical Systems

3.1 The Lognormal Price Process

Assume that a stock price process is defined by a lognormal stochastic differential equation defined by:

$$\frac{d\tilde{S}(t)}{\tilde{S}(t)} = \lambda dt + \sigma d\tilde{W}(t), S(0) > 0$$

where $\{\tilde{W}(t), t > 0\}$ is a standard (non-adapted) Brownian motion (normal probability) process with zero mean and variance t. This is a "noise", with no memory, recurring as an independent and stochastic process. The derived rate of returns process $d\tilde{R}(t) = d\tilde{S}(t)/\tilde{S}(t) = d\ln(\tilde{S}(t))$ has normal probability distribution defined by an application of Ito's calculus to the variable transformation $\tilde{R}(t) = \ln(\tilde{S}(t))$, leading to:

$$d\tilde{R}(t) = \mu dt + \sigma d\tilde{W}(t), R(0)\,\text{given}, \mu = \lambda - \frac{1}{2}\sigma^2$$

The "noise process" $\{\tilde{W}(t), t > 0\}$ is then a "historical random process" rather than defining a probability measure we define to construct a predictable pricing process. Explicitly, if λ is a mean rate of returns, if R_f is a risk free rate R_f, then $\lambda - R_f$ is a risk premium a financial agent would pay to "remove the risk" from expected returns from the lognormal pricing model. A fundamental financial model does so, by amending the historical (probability measure) "Brownian noise" $\{\tilde{W}(t), t > 0\}$ to that of and expected riskless one. Namely, paying a risk premium for the risk implied by such a model reduces the effective rate of returns to be riskless. Explicitly, adding and removing the risk premium in the lognormal model, we have:

$$\frac{d\tilde{S}(t)}{\tilde{S}(t)} = \lambda dt + \sigma d\tilde{W}(t) + (\lambda - R_f)dt - (\lambda - R_f)dt, S(0) > 0$$

And defining the probability measure $d\tilde{W}^Q(t) = d\tilde{W}(t) + \frac{\alpha - R_f}{\sigma}dt$ reduces the lognormal process to a process whose expected rate of returns is risk free (without accounting for its volatility since its risk effects were annulled by the risk premium). In this case, a risk neutral pricing model is defined by the model:

$$\frac{d\tilde{S}(t)}{\tilde{S}(t)} = R_f dt + \sigma d\tilde{W}^Q(t), S(0) > 0 \quad \text{where} \quad d\tilde{W}^Q(t) = d\tilde{W}(t) + \frac{\lambda - R_f}{\sigma} dt$$

In this case, a solution of this equation is:

$$\tilde{S}(t) = S(0) \exp\left\{R_f t - \frac{1}{2}\sigma^2 t + \sigma d\tilde{W}^Q(t)\right\}$$

Whose expectation is the fundamental (simplified) pricing model in financial assets pricing,

$$S(0) = e^{-R_f t} E^Q\left(\tilde{S}(t)\right)$$

where $E^Q(.)$ indicates an expectation is taken with respect to the Q probability measure. This property provides a rational (pricing) expectation of prices based on the probability measure Q. Further note that, such a pricing model is defined by a Martingale since:

$$S(t_0) = e^{-R_f(t-t_0)} E^Q\left(\tilde{S}(t)\right) \text{ or } S(t_0)e^{-R_f t_0} = e^{-R_f t} E^Q(\tilde{S}(t))$$

A pricing process such as the one above presumes that markets are complete, defined by a Martingale. The inverse is not true however, namely, a Martingale need not define a pricing process.

3.2 Mean Reverting Models

Such models, assume that rates of returns have a normal probability distribution. Empirical and statistical analyses of various sorts may indicate rates of returns we may model otherwise. These models are often used to model interest rate processes that underlie bond prices. For example, say that a stock rates of returns or interest rates have long run trends $\bar{R}(t)$, then:

$$d\tilde{R}(t) = -\beta(\tilde{R}(t) - \bar{R}(t))dt + \sigma d\tilde{W}(t), R(0) = R_0$$

And thus the stock (or Bond with stochastic interest rates) price is: $S(t) = S(0) \exp\{\tilde{R}(t)\}$. Generally, for linear models of this sort, we may hypothesize various forms and volatility functions $\sigma(\xi, \tilde{R}(t))$, or:

$$d\tilde{R}(t) = -\beta(\tilde{R}(t) - \bar{R})dt + \sigma(\xi, \tilde{R}(t))d\tilde{W}(t), R(0) = R_0$$

For example, the following volatility functions are often used: $\sigma(\xi, \tilde{R}(t)) = \sigma\tilde{R}(t)$, $\sigma(\xi, \tilde{R}(t)) = \sigma\sqrt{\tilde{R}(t)}$, $\sigma(\xi, \tilde{R}(t)) = \xi\tilde{R}(t) + \sigma$ as well as more complex models such as a mean reverting model $\sigma(\xi, \tilde{R}(t)) = \xi(\tilde{\sigma}(t) - \bar{\sigma})$, etc. Such models are embedded in stochastic volatility models with their historical estimates based on ARCH-GARCH type statistical estimation techniques that capture memory trends—namely, the evolution of rates of returns based on past data models. These models assume however that the "noise", fueling models' risk is memory free. From a financial viewpoint, these issues are in some cases discarded when an appropriate pricing Martingale is found reducing expected future prices to their current value which are in such cases observed (and therefore riskless) and traded by financial agents and traders agreeing to exchange money for a stock share. Such models assume an exponential memory, with past noise "forgotten" at an exponential rate. More elaborate memory models can be constructed. For example consider a Gamma memory model with probability distribution:

$$f(t) = \frac{t^{k-1} e^{-t/\lambda}}{\lambda^k \Gamma(k)}.$$

Consider the stochastic integral

$$\tilde{y}(t,k) = \frac{1}{\lambda^k \Gamma(k)} \int_0^t (t-\tau)^{k-1} e^{-(t-\tau)/\lambda} dW(\tau)$$

When k = 1, we have the exponential model: $\tilde{y}(t,1) = \frac{1}{\lambda} \int_0^t e^{-(t-\tau)/\lambda} dW(\tau)$. For k an integer, consider $h(\tau) = (t-\tau)^{k-1} e^{-(t-\tau)/\lambda}$ and the integral defined by Ito calculus:

$$\int_0^t h(\tau) dW(\tau) = h(t)W(t) - \int_0^t \frac{\partial h(\tau)}{\partial \tau} W(\tau) d\tau$$

where $h(t) = 0$, $\frac{\partial h(\tau)}{\partial \tau} = \left(-(k-1) + \frac{1}{\lambda}(t-\tau)\right)(t-\tau)^{k-2} e^{-(t-\tau)/\lambda}$ and therefore, setting,

$$Y(t,k-1) = \frac{1}{\lambda^{k-1}\Gamma(k-1)} \int_0^t (t-\tau)^{k-2} e^{-(t-\tau)/\lambda} W(\tau) d\tau$$

We have:

$$\tilde{y}(t,k) = \frac{1}{\lambda}(Y(t,k) - Y(t,k-1))$$

which leads to a multivariate models of memory if k is an integer. We shall see subsequently, that such models are generalized further by extending them to a long run memory. A typical example is the continuous time fractional ARMA process (for example, see Viano et al. 1994) defined by:

$$y(t\,|k\,) = \int_{-\infty}^t f(t-\tau) W(\tau) d\tau$$

Where $f(.)$ is a continuous, twice differentiable function with Laplace Transform:

$$F^*(s) = E\left(e^{-st} f(t)\right) = \prod_1^K (s - a_\ell)^{\delta_\ell}$$

The Gamma distribution $f(t\,|\lambda\;, k) = \frac{1}{\lambda^k \Gamma(k)} t^{k-1} e^{-t/\lambda}$ is of particular interest when extending the Ornstein-Uhlenbeck exponential process. First note that for the exponential (OU) memory model (see also Wopert and Taqqu 2005):

$$\tilde{x}^{(1)}(t) = \sigma\sqrt{2\beta} \int_{-\infty}^{t} e^{-\beta(t-\tau)} dW(\tau), x_0 = 0$$

With autocorrelation:

$$\rho^{(1)}(t) = E\left(\tilde{x}^{(1)}(0)\tilde{x}^{(1)}(t)\right) = 2\sigma^2\beta \int_{-\infty}^{0\wedge t} e^{-\beta(t-\tau)} e^{-\beta(0-\tau)} d\tau = \sigma^2 e^{-\beta|t|}$$

Now define the process:

$$\tilde{x}^{(2)}(t) = \int_{-\infty}^{t} \beta e^{-\beta(t-\tau)} \tilde{x}^{(1)}(\tau) d\tau \quad \text{or} \quad \tilde{x}^{(2)}(t) = \sigma\sqrt{2\beta} \int_{-\infty}^{t} \beta(t-u) e^{-\beta(t-u)} dW(u)$$

which leads to a recursive equation whose solution for k integer, is a Gamma "memory" model as outlined above.

$$\tilde{x}^{(k)}(t) = \int_{-\infty}^{t} \beta e^{-\beta(t-\tau)} \tilde{x}^{(k-1)}(t) d\tau = \frac{\sigma\sqrt{2\beta}}{\Gamma(k)} \int_{-\infty}^{t} \beta^{k-1} (t-\tau)^{k-1} e^{-\beta(t-\tau)} dW(\tau)$$

However, if k is not an integer, say an arbitrary parameter $\alpha > 1/2$ replacing k, we obtain:

$$\tilde{X}^{(\alpha)}(t) = \frac{\sigma\beta^{\alpha-1}\sqrt{2\beta}}{\Gamma(\alpha)} \int_{-\infty}^{t} (t-\tau)^{\alpha-1} e^{-\beta(t-\tau)} dW(\tau)$$

with a covariance function:

$$\rho^{(\alpha)}(t) = E\left(\tilde{x}^{(\alpha)}(0)\tilde{x}^{(\alpha)}(t)\right) = \frac{2\sigma^2 e^{-\beta|t|}}{[\Gamma(\alpha)]^2} \int_0^{\infty} (\beta\,|t| + \tau)^{\alpha-1} (\tau)^{\alpha-1} e^{-2\tau} d\tau$$

Comparing it to the covariance of an exponential memory $\sigma^2 e^{-\beta|t|}$, it provides a functional evaluation for the relative volatility that each of these models assumes. Finally note that we assumed an initial condition and therefore the integral in the time interval (0,t) yields:

$$\tilde{X}(t) = X(0) + \sigma\beta^{\alpha-1}\sqrt{2\beta} \int_0^t \frac{(t-\tau)^{\alpha-1} e^{-\beta(t-\tau)}}{\Gamma(\alpha)} W(\tau) d\tau$$

3.3 Fractional Models: Preliminary introduction

There are numerous generalizations to random walks that generate various approaches to their mathematical treatment (Montroll and Weiss 1965; Montroll and West 1979; Samorodnistky and Taqqu 1994; Taqqu 1986, 2003; Kac and Cheung 2002; Fitouhi et al. 2005) and many others. Below we provide a preliminary appreciation of fractional calculus and its definition of stochastic differential equations. In particular, say that the Ornstein-Uhlenbeck equation is to be in the following form due to Langevin:

$$d\tilde{x}(t) = -\beta\tilde{x}dt + \sigma w(t)(dt)^{\frac{1}{2}}, \tilde{x}(0) = R_0 - \bar{R} = 0$$

where $w(t)$ is a "White Noise" (a Brownian motion derivative) of mean 0 and variance 1). Thus, in expectation, $E(d\tilde{x}(t)) = -\beta E(\tilde{x})dt$ while. $\operatorname{var}(d\tilde{x}(t)) = (\sigma\sqrt{dt})^2 \operatorname{var}(w(t)) = \sigma^2 dt$ that are similar to that defined by an Ito stochastic differential equation. Note that in this formulation that there are two time scales, on for the drift and the other for the "noise". Consider instead a time scale defined by $(dt)^\alpha$ with $\alpha < 1$. If $\alpha = 1$ the variance growth is:

$$dX(t) = \sigma^2 dt \text{ which means } X(t) - X(t - dt) = \sigma^2 dt$$

Now consider time intervals $(dt)^\alpha$ with a differential $d^\alpha X(t)$ defined by:

$$d^\alpha X(t) = X(t) - X(t - (dt)^\alpha) = \sigma^2 (dt)^\alpha.$$

Of course for $\alpha = 1$ these two equations are identical. If not, how different or similar are these two equations if they were measured in terms of the time intervals dt? To answer this question, we consider two differential equations and will show that these two equations have the same solutions if $\Psi = \Gamma(1 + \alpha)$:

$$\frac{d^\alpha X(t)}{(dt)^\alpha} = \Psi\sigma^2 \text{ and } \frac{dX(t)}{dt} = \alpha t^\alpha \sigma^2$$

First consider their Laplace Transforms:

$$L^*\left(\frac{d^\alpha X(t)}{(dt)^\alpha}\right) = p^\alpha X^*(p) - p^{\alpha-1}X(0) = \frac{\Psi\sigma^2}{p}$$
$$L^*\left(\frac{dX(t)}{dt}\right) = pX^*(p) - X(0) = \sigma^2\frac{\alpha\Gamma(\alpha)}{p^\alpha} = \sigma^2\frac{\Gamma(1+\alpha)}{p^\alpha}$$

Equating these equations:

$$p^{\alpha+1}X^*(p) - p^\alpha X(0) = \Psi\sigma^2 \text{ and } p^{\alpha+1}X^*(p) - p^\alpha X(0) = \sigma^2\Gamma(1+\alpha)$$

we obtain $\Psi = \Gamma(1+\alpha)$. Setting $d^\alpha X(t) = \Psi dX(t) = \Gamma(1+\alpha)dX(t)$ and replacing the time scale to be $(dt)^\alpha$, $d^\alpha X(t) = \Gamma(1+\alpha)dX(t)$, we have therefore a fractional model:

$$d^\alpha X(t) = \Gamma(1+\alpha)\sigma^2(dt)^\alpha \quad \text{or} \quad \frac{d^\alpha X(t)}{(dt)^\alpha} = X^\alpha(t) = \Gamma(1+\alpha)\sigma^2$$

with $X^{\alpha}(t)$ denoting the fractional derivative. These approaches to financial modeling will be considered in greater detail with financial applications and references used to direct the reader to related and prior studies. The equivalence between these equations is however important. First it confirms that a time scale does not change the basic assumptions of a financial model based on fractional calculus or on the Riemanian calculus and second, it indicates how one can be obtained form the other. From a financial viewpoint, these are two important approaches to financial modeling and both have to be understood.

Additional references and paper we may refer to for a motivated reader include for example, Miller and Ross 1973; Oldham and Spanier 1974; Mandelbrot and Taqqu 1979; Mandelbrot and Van Ness 1968; Belair 1987; Beran 1992; Jumarie 1993, 2004, 2005, 2005, 2008, 2010, 2012a, 2012b and his recent book, 2013, Hu and Oksendahl 2003; Osler 1971; Fox and Taqqu 1985; Granger and Joyeux 1980; Graf 1983; Duncan et al. 2000. On fractional Poisson processes, there is an expanding list of references including among others, Laskin 2000, 2003, 2011; Wolpert and Taqqu 2005; Pincibono and Bendjaballah 2006; Orsingher et al. 2012 and others (see references in this text). Applications in finance and in the study of time series are also extensive which we refer to in various sections of this paper. These include and extensive list of papers of Mandelbrot but also Gray et al. 1989; Gewehe and Woodward 1984; Granger and Joyeux 1980; Willinger and Paxson 1998 (on internet application); Samko et al. 1987. In financial modeling there are numerous research papers and books such as Cont and Tankov 2004; Rostek and Schobel 2013 as well as many applications we shall use and developed by Jumarie (see his extensive list of papers).

3.4 Fat Tails Memory (Weibull Memory)

Consider at present a "fat tail memory" defined by a weighting function of past observations $e^{-\beta(t-\tau)^{\alpha}}$ (note that it is the exponential in a Weibull probability distribution). Namely, inserting in the mean reversion model:

$$\tilde{x}(t) = x_0 e^{-\beta t^{\alpha}} + \int_0^t e^{-\beta(t-\tau)^{\alpha}} dW(\tau), \alpha > 1$$

By Ito's differential rule (and assuming $x_0 = 0$):

$$\tilde{y}(t) = \int_0^t e^{-\beta(t-\tau)^{\alpha}} dW(\tau) = W(t) + \beta^{1-\alpha} \int_0^t \alpha\beta^{\alpha}(t-\tau)^{\alpha-1} e^{-\beta(t-\tau)^{\alpha}} W(\tau) d\tau$$

Or

$$\tilde{y}(t) = W(t) + \beta^{1-\alpha} \int_0^t f(t-\tau) W(\tau) d\tau, f(u) = \alpha\beta^{\alpha} u^{\alpha-1} e^{-\beta u^{\alpha}}$$

Where $f(u)$ is a Weibull probability distribution. The stochastic integral is a convolution integral and its generating function can be calculated and its moments derived.

3.5 Generalized Pareto Distributions

These distributions have a regular variation at infinity, or:

$$\underset{t\to\infty}{Lim}\frac{1-F(tx)}{1-F(t)}=x^{-\xi},\xi>0$$

For example, a particularly useful case that satisfies this property is the Pareto distribution given by:

$$f(t)=\alpha s^{\alpha}t^{-\alpha-1},\ F(t)=1-s^{\alpha}t^{-\alpha}\ \text{and}$$

$$\underset{t\to\infty}{Lim}\frac{1-F(tx)}{1-F(t)}=\underset{t\to\infty}{Lim}\frac{s^{\alpha}(xt)^{-\alpha}}{s^{\alpha}t^{-\alpha}}=x^{-\alpha},\alpha>0.$$

In this case, assuming a Pareto distribution for the memory of past "noises", we have:

$$\tilde{y}(t)=\int_0^t f(t-\tau)dW(\tau)=\alpha s^{\alpha}\int_0^t(t-\tau)^{-\alpha-1}dW(\tau)$$

which points out to a fractional integral.

Pareto Stable models have been popular in the Finance modeling literature ever since Fama's thesis supervised by Mandelbrot (1963). A standard form of the Pareto distribution is given by:

$$f(x)=\alpha\theta^{\alpha}(x+\theta)^{-(1+\alpha)},x\geq 0.$$

whose hazard rate is: $h(x)=\alpha/(x+\theta)$. Generally, Pareto Stable Distributions (PSD) are defined stable as follows: If say two random variables x_1 and x_2 are distributed as a_1u+b_1 and a_2u+b_2, their sum $x=x_1+x_2$ is also of the form $au+b$. These probability distributions are characterized by general (natural logarithm) ln characteristic functions $\varphi(z)=E(e^{ixz}),i=\sqrt{-1}$ with 3 parameters m,c,α given by (see Feller 1957–1966 for example):

$$\ln\varphi(z)=imz-c|z|^{\alpha}\left\{1-\frac{iz}{|z|}\text{tang}\left(\frac{\alpha\pi}{2}\right)\right\},i=\sqrt{-1}$$

These distributions have a field of attraction, thus the distribution of their sums differ only by location parameters. Second, they conform to some time series that exhibit the fat tail property. An explicit (albeit computationally intensive formulation) expression for Pareto Stable distributions is given by Feller (1957–1966) by:

$$\text{For } x>0, 0<\alpha<1: f(x:\alpha,\gamma)=\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(\alpha k+1)}{k!}(-x)^{\alpha}\sin\left(\frac{k\pi}{2}(\gamma-\alpha)\right)$$

$$\text{For } x > 0, 1 < \alpha < 2 : f(x:\alpha,\gamma) = \frac{1}{\pi}\sum_{k=1}^{\infty}\frac{\Gamma(\alpha^{-1}k+1)}{k!}(-x)^k \sin\left(\frac{k\pi}{2\alpha}(\gamma-\alpha)\right)$$

$$\text{For } x < 0 : f(-x:\alpha,\gamma) = f(x:\alpha,-\gamma) \text{ with } |\gamma| \leq \begin{cases} \alpha & \text{if } 0 < \alpha < 1 \\ 2-\alpha & \text{if } 1 < \alpha < 2 \end{cases}$$

Some special and well known cases include: The Normal Probability Distribution; The Generalized Pareto Distribution indicated above; The Cauchy (Infinite variance; Probability Distribution; Levy Processes and others (there are in fact an infinite number of such distributions)

Consider the Normal and Pareto Levy Stable Distributions. The Normal Distribution has, as seen earlier a finite variance (and thin tails). Assume that a stock rate of returns noise $\{\varepsilon_i\}, i = 1{,}2\cdots, n$ are i.i.d. random variables with known mean μ_ε and known finite variance σ_ε^2. The standardized random variable $\tilde{R}_n$ of a sample of i.i.d. random variables $\left\{R_\varepsilon^{(i)}\right\}, i = 1{,}2\cdots, n$ is defined by:

$$\tilde{R}_\varepsilon^{(n)} = \frac{1}{\sigma\sqrt{n}}\left\{\sum_{j=1}^{n} R_\varepsilon^{(j)} - n\mu_\varepsilon\right\}$$

It converges, by the central limit theorem, to the standard normal distribution. It is both stable and reproducible (since the sum of normally distributed events has also a normal distribution). In other words if a time series has events that are all normally distributed, their sums and averages are normally distributed. Their memory is then summarized by their parameters. The proof of such an assertion is easily demonstrated by using the Characteristic Function given by: $\Psi_1(\lambda) = E(e^{i\lambda x})$ and calculating the characteristic function of the sum of independent distributions. Letting samples of a sum of random variables be iid, we obtain the sample characteristic function equals the product of such characteristic functions, or:

$$\Psi_n(\lambda) = (\Psi_1(\lambda))^n \quad \text{as well as} \quad \Psi_{nk}(\lambda) = (\Psi_n(\lambda))^k = (\Psi_1(\lambda))^{nk}.$$

In such cases, the characteristic function has the same functional form—giving its name as stable distribution. Functionally such distributions are infinitely divisible. The Central limit theorem is then proved by showing that:

$$\underset{n\to\infty}{Lim}\Psi_n(\lambda) = \underset{n\to\infty}{Lim}(\Psi_1(\lambda))^n = \Psi_{Normal}(\lambda)$$

A slight generalization leads to the definition of Pareto-Levy processes. Let $\{R_i\}, i = 1{,}2\cdots, n$ be a random event at time i with characteristic function $\Phi_1(\lambda)$ And let $R_n = \sum_{j=1}^{n} R_j$, have a characteristic function $\Phi_n(\lambda)$. Unlike the normal probability case, we would like to find a characteristic function that meets the following conditions:

$$\Phi_n(\lambda) = [\Phi_1(\lambda)]^n \quad \text{and} \quad [\Phi_1(\alpha(n)\lambda)] = [\Phi_1(\lambda)]^n$$

In other words, both the sum and the product of the random variable multiplied by some parameters (a function of the summing parameter) are distribution invariant. In this case, we can proceed as follows (see Seville Nanjana,.MIT Lecture Notes, April, 2006). First note that:

$$[\Phi_1(\alpha(1)\lambda)] = [\Phi_1(\lambda)]^1, \Phi_1(\alpha(1)\lambda) = \Phi_1(\lambda) \text{ where } \alpha(1) = 1.$$

Next let the ln characteristic function be $\Lambda(\lambda) = \ln[\Phi_1(\lambda)]$. Thus, $\ln[\Phi_1(\alpha(n)\lambda)] = n\ln[\Phi_1(\lambda)]$ implies $\Lambda(\alpha(n)\lambda) = n\Lambda(\lambda)$. Deriving with respect to n we have:

$$\Lambda(u) = n\Lambda(\lambda),\ u = \alpha(n)\lambda \quad \text{and} \quad \frac{d\alpha(n)}{dn}\frac{d\Lambda(u)}{du}\lambda\Lambda(u) = \Lambda(\lambda),\ u = \alpha(n)\lambda,$$

$$\text{or at} \quad n = 1:$$

$$\lambda\frac{d\alpha(1)}{dn}\frac{d\Lambda(\alpha(1)\lambda)}{du} = \Lambda(\lambda) \quad \text{and} \quad \frac{d\Lambda(u)}{du} = \frac{d\Lambda(\lambda)}{d\lambda} = \frac{\Lambda(\lambda)}{\lambda\frac{d\alpha(1)}{dn}}$$

A solution to this differential equation is then of the type:

$$\Lambda(\lambda) = \begin{cases} a|\lambda|^\alpha & \lambda > 0 \\ b|\lambda|^\alpha & \lambda < 0 \end{cases} \quad \text{And thus by symmetry } \Phi(\lambda) = \begin{cases} e^{a|\lambda|^\alpha} & \lambda > 0 \\ e^{b|\lambda|^\alpha} & \lambda < 0 \end{cases}$$

Since, $\Phi(-\lambda) = \Phi(\lambda)^*$, the following characteristic function results,

$$\Phi(\lambda) = \exp\left\{(c_1 + ic_2 \text{ sgn } (\lambda))\left|\lambda\right|^\alpha\right\} \text{ with } \ c_1, c_2 \in \Re$$

The characteristic function above is a special case of the Levy-Kintchine formula for finite mean distributions for Pareto Stable distributions with the following characteristic function:

$$\Phi(\lambda) = \exp\left\{-a\left|\lambda\right|^\alpha\left(1 - i\beta\tan\left(\frac{\alpha\pi}{2}\right)\text{ sgn }(\lambda)\right)\right\}$$

Where β relates to the distribution's skewness. If the distribution is symmetric, it is then a Levy distribution which is given as stated above by:

$$\beta = 0 \text{ and } \Phi_L(\lambda) = \exp\{-a|\lambda|^\alpha\}$$

The motivation for Levy processes arises because of our own concern that a two moments distribution does not capture the richness that data indicates. In particular, data skewness, contradicting the assumption of normally distributed data. Levy processes that replace the Brownian motion process thus provide an opportunity to characterize skewness in such stable processes.

4 q-Calculus and Long Run Trends

Consider the Bernoulli differential equation:

$$\frac{d\bar{S}}{\bar{S}^q} = \bar{\mu}dt, \bar{S} = \bar{S}(0)$$

whose solution is given by: $\bar{S}_q(t) = \bar{S}(0)(1+(1-q)\bar{\mu}t)^{\frac{1}{1-q}}$ with $\underset{q\to 1}{Lim}\bar{S}_q(t) = \bar{S}(0)e^{\hat{\mu}t}$ and $\underset{q\to 1}{Lim}\, d\bar{S}_q(t)/S_q = \hat{\mu}dt$. Or, $\bar{S}_{q=1}(t) = \bar{S}(0)e^{\bar{\mu}t}$ or $(1-(1-q)\bar{\mu}t)^{\frac{1}{1-q}} \underset{\to q=1}{=} e^{-\mu t}$. The expression $(1+(1-q)\bar{\mu}t)^{\frac{1}{1-q}}$ provides a two parameters functional generalization of the exponential distribution. Let the parameters estimates be $\hat{q}$ and $\hat{\mu}$, where $\hat{\mu}$ is a long run constant corresponding to q = 1. When $q \neq 1$we have thus:

$$\frac{d\bar{S}_q(t)}{dt} = \bar{S}(0)\hat{\mu}\left((1+(1-q)\hat{\mu}t)^{\frac{q}{1-\hat{q}}}\right) \quad \text{or} \quad \frac{d\bar{S}^q(t)}{\bar{S}_q(t)} = \frac{\hat{\mu}}{(1+(1-q)\hat{\mu}t)}dt$$

As a result, a model of the form:

$$x(t) = \sigma \int_{-\infty}^{t} (1-(1-q)\bar{\mu}(t-\tau))^{\frac{1}{1-q}} dW(\tau)$$

provides a sort of extension of the OU process based on the following defined by:

$$e_q^x = \{1+(1-q)x\}_+^{\frac{1}{1-q}} = \begin{cases} 0 & if\ q < 1, x < -1/(1-q) \\ (1+(1-q)x)^{\frac{1}{1-q}} & if\ q < 1, x > -1/(1-q) \\ e^x & if\ q = 1 \\ (1+(1-q)x)^{\frac{1}{1-q}} & if\ q > 1, x < -1/(1-q) \end{cases}$$

These elements define a q-calculus (or Jackson calculus) (see for example Jackson 1910; Borges and Roditi 1998, Stephen Oney, May, 19, 2007, Oren Tapiero 2012). Some of the following relationships then hold:

$$\ln_q(x) = \frac{x^{1-q}-1}{1-q}, x > 0, q \in \Re \text{ and } \ln_q(x) = x^{1-q}\ln_{2-q}(x)$$

$$e_q^x e_{2-q}^{-x} = 1 \text{ or } \left(e_q^x\right)^q e_{1/q}^{-qx} \text{ and } e_q^{x+y+(1-q)xy} = e_q^x e_q^y$$

Importantly, the differential and the derivative are defined by:

$$d_q f(x) = f(qx) - f(x) \text{ and } \frac{d_q f(x)}{d_q x} = \frac{f(qx)-f(x)}{(q-1)x}$$

By the same token its integral (called in this case, the anti-derivative) is:

$$F(x) = \int f(x) d_q f(x)$$

For example

$$F(x) = \int_a^b f(x) d_q f(x) = \int_0^b f(x) d_q f(x) - \int_0^a f(x) d_q f(x)$$

Where

$$\int_0^a f(x)d_q f(x) = (1-q)b\sum_{j=0}^{\infty} q^j f(q^j b)$$

The financial justification for this calculus and the resultant probability distribution it provides for asset prices distributions is based on Physics distinction between extensiveness and non-extensiveness. The former is equivalent to a complete definition of all states (and therefore corresponding to complete markets) and the other to a partial definition of all states and therefore corresponding to incomplete state preferences in finance. In other words, a future price assuming 6 potential futures when only 4 are defined, corresponds to two unspecified futures and thereby to an incomplete accounting of future states. To account for future states uncertainties, Tsallis 1988, Tsallis et al. 2003, Tapiero (2013a, b, c, d) suggested that we extended the principle of maximum entropy to non-extensive systems and thus generates distributions for future states compatible with the information we have. Explicitly, for an extensive system, maximizing the Boltzmann-Gibbs entropy which is an information measures of disorder subject to a known mean and known variance, the normal distribution results. When information regarding future state preferences is partial (and thus the system is non-extensive), distributions are derived by an application of entropy functions that generalize functionally the Boltzmann-Gibbs entropy (for example, Tsallis, Renyi and others entropy measures).

For example, finding a bounded distribution that maximizes the Boltzmann-Gibbs entropy provides a uniform distribution, with all events in its bounds being of equal probability. Its rationality is a measure of disorder that defines the distribution "the most disordered distribution" that meet a set of constraints the distribution is limited to. The measure of disorder—the entropy (an informational measure), then simply the expected ln of the distribution. For a distribution whose only information is its mean, its maximum entropy distribution is exponential. If its mean and volatility are known, its maximum Boltzmann-Gibbs entropy is a normal distribution (Tapiero 2013a, b, c, d). These results concord the Laplace of Insufficient Reason, that states that (in an entropy sense), these are distributions that assume the least and yet meet the constraints imposed on their definitions. A functional generalization of the entropy provides other distributions that by definition, and for an appropriate set of parameters will also indicate the normal probability distribution. In this sense, alternative mathematical systems based on principles of maximization and constraints (a proxy for information) underlie probability models that may be used hypothetically to construct financial models. For example, the first order condition to maximizing the Tsallis entropy based on its first two moments yields a generalized function for the normal probability distribution. Setting,

$$e_q^{-\beta^* x^2} = \left\{1-(1-q)\beta^* x^2\right\}_+^{\frac{1}{1-q}}, \beta^* = \beta / \int p^q(x)dx,$$

we have the probability distribution (which reduces to a normal distribution for q = 1)

$$p(x) = \frac{e_q^{-\beta^* x^2}}{Z_q}, Z_q = \int_{-\infty}^{+\infty} e_q^{-\beta^* x^2} dx = \frac{\sqrt{\pi}\Gamma\left(\frac{1}{q-1} - \frac{1}{2}\right)}{\sqrt{\beta^*}\Gamma\left(\frac{1}{q-1}\right)\sqrt{q-1}}$$

Applying the moment constraints, we solve for β^*: $\beta^* = \frac{1}{5-3q}$. Note that for $q \to 1$, we recover the standard normal distribution (see also Wikipedia for a summary of results for the q-Gaussian distribution). Explicitly, the q-Gaussian distribution is then defined by:

$$f(x) = \frac{\sqrt{\beta}}{C_q} e_q^{-\beta x^2} \text{ with } \begin{cases} C_q = \dfrac{2\sqrt{\pi}\Gamma\left(\dfrac{1}{1-q}\right)}{(3-q)\sqrt{1-q}\Gamma\left(\dfrac{3-q}{2(1-q)}\right)} & for -\infty < q < 1 \\ C_q\sqrt{\pi} & for\ q = 1 \\ C_q = \dfrac{2\sqrt{\pi}\Gamma\left(\dfrac{1}{1-q}\right)}{(3-q)\sqrt{1-q}\Gamma\left(\dfrac{3-q}{2(1-q)}\right)} & for\ 1 < q < 3 \end{cases}$$

The parameter 'q' can be, estimated providing an index of departure from the normal probability distribution (say, by means of maximum likelihood). By the same token, assuming that the information on hand is the price of a call option, the, following distribution results:

$$p(S_T) = \left(\frac{q}{1-q}\right)^{\frac{1}{1-q}} \left(e_q^{-\sum \beta_i (S_T - K_i)^+}\right)^{\frac{1}{1-q}}$$

$$p^q(S_T) = \frac{\left(e_q^{-\sum \beta_i (S_T - K_i)^+}\right)^{\frac{1}{1-q}}}{Z(S_T : \beta, q)}, Z(S_T : \beta, q) = \int \left(e_q^{-\sum \beta_i (S_T - K_i)^+}\right)^{\frac{1}{1-q}} dS_T$$

These are power law distributions (and therefore with fat tails) enriched by the additional parameter q whose meaning is interpreted by Tsallis as a measure of physical non extensiveness (Tsallis 2009, book as well as Juniper 2007 and related by Oren Tapiero 2012, 2013a, b, c, d to Knights definition of uncertainty expressed by the incomplete definition of future state price preferences). In Oren Tapiero thesis, (2012), the parameter q is interpreted as a measure on financial incompleteness with a q smaller than 1 pointing out to a sum of future state probabilities to be incomplete, although renormalized to sum to 1. In this sense, a q-calculus provides a measure of sensitivity to complete state preferences—measured with respect to Knight's uncertainty or the definition of incomplete state preferences.

5 Fractional Models and Long Run Memory

Models such as the Orntein-Uhlenbeck mean reversion, the normal rates of returns model and so many others are based on the assumption that the time interval dt as well as state increments (in temporal processes, etc.) dS are very small. Now say that we "parameterize" time intervals to be defined by $(dt)^{\alpha}$. For example, say that $dt = 0.5$ then necessarily if $\alpha > 1$, say $\alpha > 2$, then the time interval $(dt)^{\alpha}$ is smaller than (dt). However, if $\alpha < 1$, say $\alpha = 1/2$, then $(dt)^{\alpha} = 0.7071$ is greater. In this case, the time clock of two measurements is greater than 0.5. Explicitly, say that we have price at time t, $S(t)$ and consider the price previous time $S(t - dt)$. This price would be a price 0.5 time units before while the price $S(t - (dt)^{\alpha})$ would be "older price". Memory in this case, is defined by the elapsed times during which events were recorded and that affect current prices. Jumarie (2006) presents a number of simple examples. Consider again the example:

$$dS(t) = \alpha\rho t^{\alpha-1}dt, S(0) > 0 \text{ and } dS(t) = \rho(dt)^{\alpha}, S(0) > 0$$

Both prices have the same solutions $S(t) = S(0) + \rho t^{\alpha}$. Yet, they are not the same equation. The proof of their equality is important to highlight the particularities of their computations. We consider again the Laplace Transform of $dS(t) = \alpha\rho t^{\alpha-1}dt$ which is:

$$pS^*(p) - S(0+) = \rho\alpha\Gamma(\alpha)p^{-\alpha} \text{ or } S^*(p) = \frac{S(0)}{p} + \rho\Gamma(1+\alpha)p^{-\alpha}$$

Now consider $dS(t) = \rho(dt)^{\alpha}$. A first transformation of the fractionalized equation (to have both the right and the left sides of the equation to be measured along a similar time scale) is (as noted earlier):

$$d^{\alpha}S(t) = \Gamma(1+\alpha)\rho(dt)^{\alpha} \text{ or } S^{(\alpha)}(t) = \frac{d^{\alpha}S(t)}{(dt)^{\alpha}} = \Gamma(1+\alpha)\rho$$

Using Laplace Transform, we have:

$$L^*\left(S^{(\alpha)}(t)\right) = p^{\alpha}S^*(p) - p^{\alpha-1}S(0) = L^*(\Gamma(1+\alpha)\rho)$$

However, this latter term is the derivative of $\Gamma(1+\alpha)\rho(dt)^{\alpha}$ and therefore,

$$L^*(\Gamma(1+\alpha)\rho) = p^{-1}\Gamma(1+\alpha)\rho$$

Or,

$$p^{\alpha}S^*(p) - p^{\alpha-1}S(0) = p^{-1}\Gamma(1+\alpha)\rho \quad \text{or,} \quad S^*(p) = \frac{S(0)}{p} + \Gamma(1+\alpha)\rho p^{\alpha-1}$$

which is identical to the previous equation

The equation $S(t+dt) = S(t) + \alpha\rho t^{\alpha-1}dt$ relates $S(t)$ to its next record by the small (infinitesimal) time interval dt. The latter, relates to a time interval which his greater if α is smaller than 1, since it increases the time interval. In this sense, the future state arises from the observation of an event that has occurred further down the "past". By multiplying by $\Gamma(1+\alpha)$ we have transformed the calculus of this equation to be uniformly defined in terms of $(dt)^\alpha$ and therefore:

$$S^\alpha(t+(dt)^\alpha) = S^\alpha(t) + \Gamma(1+\alpha)\rho(dt)^\alpha$$

Or

$$dS(t) = \rho(dt)^\alpha \text{ relates to } d^\alpha S(t) = \Gamma(1+\alpha)\rho(dt)^\alpha \text{ or } d^\alpha S(t) = \Gamma(1+\alpha)dS(t)$$

Thus, decreasing with α, the past memory increases. In financial models, data availability will thus dictate the type of models we can use and the results we may predict. A High Frequency Trader (HFT) using a continuous time equivalent with instant memory will of course neglect the importance of longer run prices. By the same token, using an exponential weighting function of past events, may provide a model to account (at least in theory) for the effects of past prices on a current price.

In some cases, one reduces the frequency of data records linearly, say by a time transformation $t' = \lambda t$ with $\lambda < 1$ slow and $\lambda > 1$ fast. In this case, $dS = f(S)(\lambda dt)$. Long run memory models consider instead time intervals $(dt)^\alpha$, and therefore uses differential models $dS = f(S)(dt)^\alpha$. To maintain the fractional consistency of this equation, memory is introduced by setting (as shown above):

$$d^\alpha S = \Gamma(1+\alpha)f(S)(dt)^\alpha.$$

And therefore, to a fractional derivative:

$$S^\alpha(t) = \frac{d^\alpha S(t)}{(dt)^\alpha} = \Gamma(1+\alpha)f(S)$$

These transformations and manipulations require a mathematical system to maintain their coherence. Numerous contributions have contributed to both the development of such models and their application. A seminal paper by Tyrone Duncan et al. (2000) has provided a series of rules that are parallel to Ito's calculus that account for fractional stochastic models. Guy Jumarie, Mandelbrot and Taqqu, Laskin and many others (see the extensive list of papers and books in references) have provided both a set of rules that simplify a coherent application of fractional calculus as well as numerous examples and applications in several fields, including finance**.**

Fractional calculus uses intensively two function we shall use repeatedly. These are the Mittag-Leffler and the Riemann-Liouville functions.

5.1 The Mittag-Leffler Function (Mittag Leffler 1903–1905)

The Mittag-Leffler function is defined by the infinite series:

$$E_\alpha(h) = \sum_{k=0}^{\alpha} \frac{h^k}{\Gamma(1+\alpha k)}$$

And is a generalization of exponential models in the sense that for $\alpha = 1$:

$$E_1(h) = \sum_{k=0}^{\infty} \frac{1}{\Gamma(1+k)} h^k = e^h$$

For example, setting $h = \lambda t^\alpha$ we have $E_\alpha(\lambda t^\alpha) = \sum_{k=0}^{\alpha} \frac{1}{\Gamma(1+\alpha k)} (\lambda t^\alpha)^k$ whose Laplace Transform can be shown to be: $L^*(E_\alpha(\lambda t^\alpha)) = \frac{p^{\alpha-1}}{p^\alpha - \lambda}$. Laplace and other transforms may be used profitably as they provide simpler and treatable formulations of fractional models. For example, if the Mittag-Leffler function is a generalization of the exponential and its Laplace Transform is as stated above, the convolution of m such functions will have a Laplace Transform given by $\left(\frac{p^{\alpha-1}}{p^\alpha - \lambda}\right)^m$. Further, if we set: $\Phi_\alpha(t) = E_\alpha(\Omega_\alpha(t))$ then the following ln for a fractional α holds: $\Omega_\alpha(t) = \ln_\alpha(\Phi_\alpha(t))$

5.2 *The Cauchy-Riemann-Liouville function (Liouville 1832)*

To better appreciate the role of the Riemann-Liouville equation, it is useful to establish its relationship with the Cauchy equation which has established for integer values a recursive integration formula (for a review of this approach and its relationship to fractional calculus, the reader is referred to Wikipedia).

Consider first the basic definition of a derivative, which we express for convenience by $f^{(1)}(t)$:

$$\underset{\Delta t \to 0}{Lim} \frac{\Delta f(t)}{\Delta t} = \underset{\Delta t \to 0}{Lim} \frac{f(t+\Delta t) - f(t)}{\Delta t} = \frac{df(t)}{dt} \equiv f^{(1)}(t)$$

Generally, for the nth integer derivative, we have:

$$\underset{\Delta t \to 0}{Lim} \frac{\Delta^n f(t)}{\Delta t^n} = \frac{d^n f(t)}{dt^n} \equiv f^{(n)}(t)$$

Cauchy's repeated integration of such derivative will then proceed as follows:

$$\frac{d^{n-1} f(t)}{dt^{n-1}} \equiv \int_0^t f^{(n)}(\tau) d\tau \text{ or } f^{(n-1)}(\tau) \equiv \int_0^t f^{(n)}(t) d\tau$$

Thus,

$$f^{(n-2)}(t) \equiv \int_0^t f^{(n-1)}(\tau) d\tau = \int_0^t \left(\int_0^\tau f^{(n-1)}(s) ds \right) d\tau$$

And recursively, Cauchy's integration equation is:

$$f^{(n-k)}(t) \equiv \int_0^t f^{(n-(k-1))}(\tau) d\tau = \frac{1}{(n-(k-1))!} \int_0^t (\tau - t)^{n-(k-1)} f^{(n)}(\tau) d\tau$$

Where $f^{(n-k)}(t)$ is the kth integration of the nth derivative $f^{n}(t)$ which we may rewrite conveniently (as it is done commonly) by setting $f(t) \equiv f^{n}(t)$ and therefore, $f^{(k)}(t) \equiv f^{(n-k)}(t)$, or:

$$f^{(k)}(t) \equiv = \frac{1}{(k-1)!} \int_{0}^{t} (\tau - t)^{(k-1)} f(\tau) d\tau$$

This equation can be verified by noting that:

$$f^{(1)}(t) = \int_{0}^{t} f(\tau) d\tau, f^{(2)}(t) = \int_{0}^{t} (\tau - t) f(\tau) d\tau, \text{etc.}$$

The integration operator thus indicates a relationship that will underlie, as we shall see below the fractional calculus:

$$f^{(k+m)}(t) \equiv f^{(k)}(t) f^{(m)}(t) \text{ with } f^{(k)}(t) f^{(m)}(t) = f^{(m)}(t) f^{(k)}(t)$$

The latter condition being commutative (which does not hold for a fractional model).

Fractional calculus is concerned with fractional integrations (or derivatives) of the Cauchy equation, defined by non-integers (although, they include as special cases integers). In particular, let α be a real value and set:

$$f^{(\alpha)}(t) \equiv = \frac{1}{\Gamma(\alpha)} \int_{0}^{t} (\tau - t)^{\alpha - 1} f(\tau) d\tau$$

This is also called the Riemann-Liouville equation which is a fractional reformulation of Cauchy's integration formula. The product operator above is maintained as indicated below, although it is no longer commutative:

$$f^{(\alpha+\beta)}(t) \equiv f^{(\alpha)}(t) f^{(\beta)}(t)$$

The proof of this relationship is led out in both Wikipedia and in Jumarie (2013). For educational purposes, we repeat the proof below with an application a derivative $f^{(\beta)}(\tau)$ and subsequently to that of a derivative $f^{(\alpha)}(\tau)$. By definition:

$$\begin{aligned} f^{(\alpha+\beta)}(t) &= \frac{1}{\Gamma(\alpha)} \int_{0}^{t} (\tau - t)^{\alpha - 1} f^{(\beta)}(\tau) d\tau \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_{0}^{t} f(u) \int_{0}^{\tau} (t - \tau)^{\alpha - 1} (\tau - u)^{\beta - 1} du d\tau \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_{0}^{t} \int_{u}^{t} (t - \tau)^{\alpha - 1} (\tau - u)^{\beta - 1} f(u) du d\tau \end{aligned}$$

$$= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int\limits_0^t f(u) \left(\int\limits_u^t (t-\tau)^{\alpha-1}(\tau-u)^{\beta-1} f(u) d\tau \right) du$$

Introduce the following change of variables $\tau = u + (t-u)r$,then

$$f^{(\alpha+\beta)}(t) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^t (t-u)^{\alpha+\beta-1} f(u) \left(\int_0^1 (1-r)^{\alpha-1} r^{\beta-1} dr \right) du$$

Since for the Beta integral we have:

$$\int\limits_0^1 (1-r)^{\alpha-1} r^{\beta-1} dr d\tau = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

We have,

$$f^{(\alpha+\beta)}(t) = \frac{1}{\Gamma(\alpha+\beta)} \int\limits_0^t (t-\tau)^{\alpha+\beta-1} f(\tau) d\tau = f^{(\alpha)}(t) f^{(\beta)}(t)$$

For example say that $f(t) = t^k$ and therefore in a Riemannian calculus,

$$f^{(1)}(t) = kt^{k-1}, f^{(2)}(t) = k(k-1)t^{k-2} \text{ and } f^{(n)}(t) = \frac{k!}{(k-n)!} t^{k-n}$$

In a fractional calculus, we have instead:

$$f^{(\alpha)}(t) = \frac{d^\alpha (t^k)}{dt^\alpha} = \frac{\Gamma(k+1)}{\Gamma(k-\alpha+1)} t^{k-\alpha}$$

In particular, for $k = 1$ and $\alpha = 1/2$, we have:

$$f^{(\frac{1}{2})}(t) = \frac{d^{\frac{1}{2}}(t)}{dt^{\frac{1}{2}}} = \frac{\Gamma(1+1)}{\Gamma(1-\alpha+1)} t^{1-\frac{1}{2}} = \frac{2}{\Gamma(3/2)} t^{\frac{1}{2}} = \frac{2\sqrt{t}}{\sqrt{\pi}}$$

Next assume that $\alpha = \frac{3}{2} > 1$ which we rewrite as follows: $\alpha = \frac{3}{2} = 1 + \frac{1}{2}$. In this case,

$$f^{(\frac{3}{2})}(t) = \frac{d^{\frac{1}{2}}(t)}{dt^{\frac{1}{2}}} = f^{(\frac{1}{2})}(t) f^{(1)}(t)$$

Since, $f^{(1)}(t) = \frac{df(t)}{dt}$ and if $f(t) = t$ or $f^{(1)}(t) = 1$ we have:

$$f^{(\frac{3}{2})}(t) = f^{(\frac{1}{2})} f^{(1)}(t) = f^{(\frac{1}{2})}(1) \text{ and therefore,}$$

$$f^{(\frac{1}{2})} f^{(1)}(t) = f^{(\frac{1}{2})}(1) = \frac{1}{\Gamma\left(\frac{1}{2}\right)} \int\limits_0^t (\tau-t)^{\frac{1}{2}-1} d\tau = \frac{1}{\Gamma\left(\frac{1}{2}\right)} \int\limits_0^t (\tau-t)^{-\frac{1}{2}} d\tau$$

Calculations such as these are often easier when we use Laplace Transforms. Define a function $f(t)$, then it Laplace Transforms is:

$$f^*(s) = \int_0^\infty e^{-s\tau} f(\tau)d\tau \text{ and therefore } f^{*(1)}(s) = \int_0^\infty e^{-s\tau} \int_0^\tau f(u)dud\tau = \frac{1}{s} f^*(s)$$

For a double integral, we have: $f^{*(2)}(s) = \frac{1}{s^2} f^*(s)$ and generally, for an integer n, $f^{*(n)}(s) = \frac{1}{s^n} f^*(s)$. However, for a fractional α, we have:

$$f^{*(\alpha)}(s) = f^{-1*}\left(\frac{1}{s^\alpha} f^*(s)\right) \text{ where } f^{-1*} \text{ is an inverse transform.}$$

For example,

$$f^{(\alpha)}(t^k) = f^{-1^*}\left(\frac{\Gamma(k+1)}{s^{\alpha+k+1}}\right) = \frac{\Gamma(k+1)}{\Gamma(k+\alpha+1)} t^{\alpha+k+1}$$

By the same token, the Laplace Transform of a convolution integral has a Laplace Transform which equals the product of their Laplace Transforms.

The Riemann-Liouville function as indicated above, provides a functional model relating fractional models (in $(dt)^\alpha$) into Riemann's calculus. The fractional calculus is then explicitly defined for a function $f(x)$ which need not be continuous (since these are evaluated by integrals). Let its forward value be $f(x+h)$, with h a constant. A fractional derivative of order α of a function $f(x)$ is thus defined by $\Delta^{(\alpha)} f(x)$, where:

$$\Delta^{(\alpha)} f(x) = f^{(\alpha)}(x) \quad \text{or} \quad f^{(\alpha)}(x) = d^{(\alpha)} f(x)/(dx)^\alpha$$

For example, the fractional derivative of say a financial price $S^{(\alpha)}(t)$ which is derived from a fractional stochastic model whenever its limit exists, is then

$$S^{(\alpha)}(t) \propto \underset{\Delta t \to 0}{Lim}(\Delta^\alpha S(t)/(\Delta t)^\alpha) \text{ when it exists and is finite.}$$

This allows one to write: $\Delta S(t) = \sigma(\Delta t)^\alpha$ or $\Delta^\alpha S(t) = \Gamma(1+\alpha)\sigma(\Delta t)^\alpha$ and at the limit

$$S^{(\alpha)}(t) = \Delta^\alpha S(t)/(dt)^\alpha = \Gamma(1+\alpha)\sigma.$$

The solution of the fractional derivative for $S(t)$ is a function of the parameter α and given by Riemann-Liouville function, or:

$$S^{(\alpha)}(t) = \frac{d^{(\alpha)} S(t)}{(dt)^\alpha} = \begin{cases} \dfrac{1}{\Gamma(-\alpha)} \int_0^t (t-\tau)^{-\alpha-1} S(\tau)d\tau & if \quad \alpha < 0 \\ \dfrac{1}{\Gamma(1-\alpha)} \dfrac{d}{dt} \int_0^t (t-\tau)^{-\alpha}(S(\tau) - S(0))d\tau & if \quad 0 < \alpha < 1 \\ \left(S^{(\alpha-n)}(t)\right)^{(n)} & if \quad n \le \alpha < n+1, n \ge 1 \end{cases}$$

When $\alpha < 0$, the past weight of prices $(t-\tau)^{|-\alpha|-1}$ increases the more negative this parameter. In other words, past prices "are not forgotten", maintaining an "appreciable" effect on current prices. When $0 < \alpha < 1$, a time derivative yields:

$$\frac{d^{(\alpha)}S(t)}{(dt)^{\alpha}} = \frac{\alpha}{\Gamma(1-\alpha)}\int_0^t (t-\tau)^{-\alpha-1}(S(\tau)-S(0))d\tau$$
$$= \frac{1}{\Gamma(-\alpha)}\int_0^t (t-\tau)^{-\alpha-1}(S(\tau)-S(0))d\tau$$

And therefore, a memory effect is maintained but at the same time, it is less pronounced. The Laplace transform for $0 < \alpha < 1$ is then a simple convolution or:

$$L^*\left(S^{(\alpha)}(p)\right) = \frac{1}{\Gamma(-\alpha)}L^*\left(t^{-\alpha-1}\right)\left(L^*(S(t)) - L^*(S(0))\right) \text{ where}$$
$$L^*(t^{-\alpha-1}) = p^{\alpha}\Gamma(-\alpha)$$

As a result,

$$L^*\left(S^{(\alpha)}(p)\right) = p^{\alpha}\left(L^*(S(t)) - \frac{S(0)}{p}\right) = p^{\alpha}L^*(S(t)) - p^{\alpha-1}S(0),\ 0 < \alpha < 1$$

And

$$L^*\left(S^{(\alpha)}(p)\right) == p^{\alpha}S^*(p) - p^{\alpha-1}S(0) \text{ or } S^*(p) = \frac{L^*(S^{(\alpha)}(p))}{p^{\alpha}} + \frac{S(0)}{p}$$

Of course, if $\alpha = 1$ then $L^*(f^{(1)}(x)) = sL^*(f(x)) - f(0), \alpha = 1$ which correspond to:

$$L^*\left(\frac{dS(t)}{dt}\right) = pL^*(S(t)) - f(0), \alpha = 1$$

In this sense the Riemann-Liouville provides a functional approach to calculating fractional derivatives which reduce to the standard calculus when $\alpha = 1$.

The Mittag-Leffler equation however provides a functional approach to calculating Taylor series expansions of fractional order α. It is written as follows (where we maintain as well the kth derivative of order α):

$$S(t+h) = \sum_{k=0}^{\infty} \frac{\left(h^{\alpha k}\dfrac{d^{(k)}S^{\alpha}(t)}{(dt)^{(k)}}\right)}{\Gamma(1+\alpha k)}$$
$$= \sum_{k=0}^{\infty} \frac{\left(h^{\alpha k}\dfrac{d^{(k)}}{(dt)^{(k)}}S^{(\alpha)}(t)\right)}{\Gamma(1+\alpha k)} = E_{\alpha}\left(h^{\alpha}D^{\alpha}\left(S^{(\alpha)}(t)\right)\right), 0 < \alpha < 1$$

with the notation: $\frac{d^{(k)}S^{(\alpha)}(t)}{(dt)^{(k)}} = D^k(S^{(\alpha)}(t))$. In other words, a two terms series expansion of S(t) by a fractional calculus of order α yields:

$$S(t+h) = S(t) + \frac{h^\alpha}{\Gamma(1+\alpha)}\frac{d^{(\alpha)}S(t)}{(dt)^{(\alpha)}} + \frac{h^{2\alpha}}{\Gamma(1+2\alpha)}\frac{d^{(2\alpha)}S(t)}{(dt)^{(2\alpha)}} + \cdots ., 0 < \alpha < 1$$

Or

$$S(t+h) = S(t) + \frac{h^\alpha}{\Gamma(1+\alpha)}S^{(\alpha)}(t) + \frac{h^{2\alpha}}{\Gamma(1+2\alpha)}S^{(2\alpha)}(t) + \cdots ., 0 < \alpha < 1$$

Note however that, the order of integration is not symmetric. Namely, although:

$$S^{2\alpha}(t) = \frac{d^\alpha}{dt}S^\alpha(t), S^{3\alpha}(t) = \frac{d^\alpha}{dt}S^{2\alpha}(t) \neq \frac{d^{2\alpha}}{dt^2}S^\alpha(t)$$

As a result, the order in which fractional derivatives are calculated matters. Further, since the fractional derivatives depend on $\beta = \alpha k$ and since the definition of the derivative defined by the Riemann-Liouville function varies according to its parameter (in this case β replacing α), one has to apply correspondingly the appropriate derivative "transformation". In any case, a Taylor series development of the first α order yields:

$$S(t+h) = S^{(0)}(t) + S^{(\alpha)}(t)\frac{h^\alpha}{\Gamma(1+\alpha)}, 0 < \alpha < 1$$

While a second order approximation yields:

$$S(t+h) = S^{(0)}(t) + S^{(\alpha)}(t)\frac{h^\alpha}{\Gamma(1+\alpha)} + \frac{h^{2\alpha}}{\Gamma(1+2\alpha)}S^{2\alpha}(t)$$

where the derivative $S^{2\alpha}(t)$ is necessarily a function of whether α is smaller or greater than 1. Jumarie 2009, (p. 381), in particular proves a more general equation

$$S(t+h) = \sum_{k=0}^{\infty}\frac{h^k}{k!}S^{(k)}(t) + \sum_{k=1}^{\infty}\frac{h^{(\beta k+m)}}{\Gamma(1+m+\beta k)}S^{(\beta k+m)}(t), \beta = \alpha - m$$

where the order of a fractional derivative is important as indicated previously. Explicitly, consider a derivative of the α order first and then that of β order and vice versa, applying first a Laplace transform of β order first and then that of α order. We have then two formulas that differ due to the initial condition of their first derivatives, or:

$$L^*(D^{(\alpha+\beta)}S(t)) = L^*(D^{(\alpha)}S^{(\beta)}S(t)) = p^{\alpha+\beta}L^*(S(t)) - p^{\alpha+\beta-1}S(0) - p^{\beta-1}S^{(\alpha)}(0),$$

$$L^*(D^{(\alpha+\beta)}S(t)) = L^*(D^{(\beta)}f^{(\alpha)}f(t)) = p^{\alpha+\beta}L^*(S(t)) - p^{\alpha+\beta-1}S(0) - p^{\alpha-1}S^{(\beta)}(0),$$

To circumvent this issue, Jumarie proposes that derivatives ought to be in increasing order such that:

$$D^{(\alpha+\beta)}S(t) \equiv D^{\max(\alpha,\beta)}S(t)D^{\min(\alpha,\beta)}S(t)$$

Similar issues recur when we seek the following derivative: $D^{(2\alpha)}S(t)$. Should we replace α by 2α in the Riemann-Liouville equation or apply the derivative twice. In the first and second case we have:

$$L^*(D^{(2\alpha)}S(t)) = p^{2\alpha}L^*(S(t)) - p^{2\alpha-1}S(0)$$
$$L^*(D^{(\alpha)}D^{(\alpha)}S(t)) = p^{2\alpha}L^*(S(t)) - pS(0) - p^{\alpha-1}S^{(\alpha)}(0),$$

which of course they are not the same since $L^*(D^{(2\alpha)}S(t)) \neq L^*(D^{(\alpha)}D^{(\alpha)}S(t))$ as indicated previously since it is their initial conditions that have to carefully assessed (see Jumarie 2009, p. 380).

The MacLaurin expansion is of course a special case of a Taylor series expansion, leading to:

$$S(t) = \sum_{k=0}^{\infty} \frac{x^{\alpha k}}{\Gamma(1+\alpha k)} S^{(k\alpha)}(0), 0 < \alpha < 1$$

These functions lead to a calculus with which a trove of results can be obtained. Below, we summarize a number useful relationships that may be proved using the elements stated above:

5.3 *Some Useful Relationships with deterministic fractional calculus:*

$$d^{\alpha}S(t) \approx \Gamma(1+\alpha)dS(t), 0 < \alpha < 1$$
$$D^{\alpha}(t)^{\theta} \approx \frac{\Gamma(1+\theta)}{\Gamma(1+\theta-\alpha)}(t)^{\theta-\alpha}, 0 < \alpha \leq 1$$

Derivative operators (for $0 < \alpha < 1$):

$$\frac{d^{\alpha}}{dt^{\alpha}} \int_0^t S(\tau)(d\tau)^{\alpha} = \Gamma(1+\alpha)S(t)$$

$$\frac{d^{\alpha}}{dt^{\alpha}} \int_0^{U(t)} S(\tau)(d\tau)^{\alpha} = \Gamma(1+\alpha)f(U(t))(U'(t))^{\alpha}$$

Further, integration with respect to $(dt)^{\alpha}$ yields,

$$y(t) = \int_0^t S(\tau)(d\tau)^{\alpha} = \alpha \int_0^t (t-\tau)^{\alpha-1}S(\tau)d\tau, 0 < \alpha \leq 1$$

As well as

$$y(t) = \int_a^t S(\tau)(d\tau)^\alpha = \alpha \int_0^t (t-\tau)^{\alpha-1} S(\tau - a)d\tau, 0 < \alpha \leq 1$$

5.3.1 Other Calculations Examples:

$$\int_0^t \tau^\theta (d\tau)^\alpha = \frac{\Gamma(1+\alpha)\Gamma(1+\theta)}{\Gamma(1+\alpha+\theta)} t^{\alpha+\theta}, 0 < \alpha \leq 1$$

while

$$y(t) = \int_a^t \tau^\theta (d\tau)^\alpha = \frac{\Gamma(1+\alpha)\Gamma(1+\theta)}{\Gamma(1+\alpha+\theta)} (t-a)^{\alpha+\theta}, 0 < \alpha \leq 1$$

and

$$y(b\,|a\,) = \int_a^b (d\tau)^\alpha = (b-a)^\alpha.$$

For a non-continuous function, integration can be realized by segments:

$$\begin{aligned} y(b\big|a) &= \int_a^b f(\tau)(d\tau)^\alpha \\ &= \sum_{j=1}^{N} \int_{a_{j-1}}^{a_j} \left(\frac{b-\tau}{a_j - 1}\right)^{\alpha-1} f(\tau)(d\tau)^\alpha, a_0 = a < a_1 < a_2 < \cdots < a_N \end{aligned}$$

In addition, note that an integration with respect to the "double order integration" $(d\tau)^{\alpha+\beta}$ would integrate first with respect to $(d\tau)^\beta$ and then with respect to order $(d\tau)^\alpha$, or:

$$y(t) = \int_0^t f(\tau)(d\tau)^{\alpha+\beta} = \frac{\alpha}{\alpha+\beta} \int_0^t (t-\tau)^\beta f(\tau)(d\tau)^\alpha, 0 < \alpha + \beta \leq 1$$

And then,

$$\int_0^t (t-\tau)^\beta f(\tau)(d\tau)^\alpha = \alpha \int_0^t (t-\tau)^{\alpha+\beta-1} f(\tau)d\tau, 0 < \alpha \leq 1$$

And thereby,

$$y(t) = \int_0^t f(\tau)(d\tau)^{\alpha+\beta} = \frac{\alpha^2}{\alpha+\beta} \int_0^t (t-\tau)^{\alpha+\beta-1} f(\tau) d\tau, 0 < \alpha+\beta \le 1$$

Now say that $2 > \alpha + \beta > 1$, then,

$$y\,(t) = \int_0^t f\,(\tau)(d\tau)^{\alpha+\beta} = \int_0^t f\,(\tau)(d\tau)^{1+\lambda} \text{ where } 0 < \lambda = (1-\alpha-\beta) < 1$$

And therefore,

$$f^{(1+\lambda)}(t) = \frac{1}{\Gamma(\lambda)} \int_0^t (\tau - t)^{\lambda-1} \frac{d}{du} \int_0^\tau f(u) du d\tau$$

Thus, if $f(u) = e^{-\mu u}$,

$$f^{(1+\lambda)}(t) = \frac{1}{\mu\Gamma(\lambda)} \int_0^t (\tau - t)^{\lambda-1}(1 - e^{-\mu\tau}) d\tau$$

$$= \frac{1}{\mu\Gamma(1-\alpha-\beta)} \int_0^t (\tau - t)^{-(\alpha+\beta)}(1 - e^{-\mu\tau}) d\tau$$

5.3.2 Integration by Parts

Integration by parts yields the following results:

$$y(b\big|a) = \int_a^b u^{(\alpha)}(\tau)v(\tau)(d\tau)^\alpha = \Gamma(1+\alpha)\,[u(\tau)v(\tau)]_a^b - \int_a^b u(\tau)v^{(\alpha)}(\tau)(d\tau)^\alpha$$

The application models are deterministic approximations of mean reversion as well as the lognormal model. Their purpose if the highlight the effects of a fractional term on the evolution of the underlying process.

5.4 A risk free bond and the Vacicek (OU) model

A risk free bond model is defined as follows:

$dB(t) = R_f B(t) dt, B(0) > 0$ or $d \ln B(t) = R_f dt$ and $B(t) = B(0)e^{R_f t}$. The price of a risk free bond whose price at maturity is $B(T)$ is thus $B(0) = B(T)e^{-R_f T}$.

Say that time is fractional such that $dB(t) = R_f B(t)(dt)^\alpha$. In this case, in a fractional model, we have a bond price:

$$\frac{d^\alpha B(t)}{B(t)} = \Gamma(1+\alpha)R_f(dt)^\alpha \text{ or } \int_{B(0)}^{B(T)} \frac{d^\alpha B(t)}{B(t)} = \ln_\alpha B(T) = \Gamma(1+\alpha)R_f \int_0^T (dt)^\alpha$$

And inversely, $B(0)E_\alpha(\ln_\alpha(B(t))) = B(t)$. For a time varying risk free rate,

$$\int_{B(0)}^{B(T)} \frac{d^\alpha B(t)}{B(t)} = \ln_\alpha\left(\frac{B(T)}{B(0)}\right) = \Gamma(1+\alpha)\int_0^T R_f(t)(dt)^\alpha$$

Thus,

$$B(T) = B(0)\alpha E_\alpha\left(\Gamma(1+\alpha)\int_0^T (T-\tau)^{\alpha-1}R_f(\tau)d\tau\right)$$

Note that if $R_f(\tau)$ is constant, then:

$$\int_0^T (T-\tau)^{\alpha-1}R_f d\tau = -R_f\int_T^0 u^{\alpha-1}du = \frac{1}{\alpha}R_f T^\alpha$$

And therefore,

$$B(T) = B(0)E_\alpha(\Gamma(1+\alpha)R_f T^\alpha)$$

Where $E_\alpha(\Gamma(1+\alpha)R_f T^\alpha)$ is the Mittag-Leffler function (which is the exponential function when $\alpha = 1$). Explicitly,

$$E_\alpha(\Gamma(1+\alpha)R_f T^\alpha) = \sum_{k=0}^{\infty}\frac{(\Gamma(1+\alpha)R_f T^\alpha)^k}{\Gamma(1+\alpha k)}$$

The price of a bond in a fractional model is thus:

$$B(0) = (E_\alpha(\Gamma(1+\alpha)R_f T^\alpha))^{-1}B(T)$$

When the risk free rate is stochastic (say, driven by a Brownian motion), we then have a lognormal model.

A slight generalization consists in setting a mean rate of return trend in dt and a longer long run volatility trend with $R_f$and σ constants (for example, a bond with two time scales as note previously):

$$\frac{dB(t)}{B(t)} = R_f dt + \lambda(dt)^\alpha,\ B(0) = B_0 > 0, 0 < \alpha < 1$$

Consider

$$dB_1(t)/B_1(t) = R_f dt \text{ and } B_1(t) = B(0)e^{R_f t} \text{ as well}$$

$$dB_2(t)/B_2(t) = \lambda(dt)^\alpha \text{ and } d^\alpha B_2(t)/B_2(t) = \Gamma(1+\alpha)\lambda(dt)^\alpha \text{ or } \mathrm{B}_2(t) = E_\alpha(\Gamma(1+\alpha)\lambda t^\alpha)$$

and therefore,

$$B(t) = B_1(t)B_2(t) = B(0)e^{R_f t}E_\alpha(\Gamma(1+\alpha)\lambda t^\alpha)$$

And the price of a bond whose maturity is at T and whose nominal value is $B(T)$ is:

$$B(0) = e^{-R_f T}(E_\alpha(\Gamma(1+\alpha)\lambda T^\alpha))^{-1}B(T)$$

If $\alpha = 1$, we have then as expected:

$$B(t) = B(0)e^{R_f t}E_1(\Gamma(1+\alpha)\lambda t^\alpha) = B(0)e^{(R_f)t}e^{(\lambda)t} = B(0)e^{(R_f+\lambda)t}$$

Which verifies the solution of the differential equation $dB(t)/B(t) = (R_f + \lambda)\, dt$, $\mathrm{B}(0) = B_0$.

By the same token, consider a stock price whose mean evolution is:

$$\frac{dS(t)}{S(t)} = R_f dt - \frac{1}{2}\sigma^2(dt)^\alpha,\ S(0) = S_0 > 0, 0 < \alpha < 1$$

We have then the solution $S(t) = S(0)e^{R_f t}E_\alpha\left(-\frac{1}{2}\Gamma(1+\alpha)\sigma^2 t^\alpha\right)$. Thus, if we consider a data time series, then:

$$\ln S(t) = \ln S(0) + R_f t + \ln E_\alpha\left(-\frac{1}{2}\Gamma(1+\alpha)\sigma^2 t^\alpha\right)$$

And explicitly, since $E_\alpha\left(-\frac{1}{2}\Gamma(1+\alpha)\sigma^2 t^\alpha\right) = \sum_{k=0}^{\infty}\left(\frac{1}{2}\Gamma(1+\alpha)\right)^k \frac{(-\sigma^2 t^\alpha)^k}{\Gamma(1+\alpha k)}$, we have:

$$\ln S(t) = \ln S(0) + R_f t + \ln \sum_{k=0}^{\infty}\left(\frac{1}{2}\Gamma(1+\alpha)\right)^k \frac{(-\sigma^2 t^\alpha)^k}{\Gamma(1+\alpha k)}$$

Where

$$\sum_{k=0}^{\infty}\left(\frac{1}{2}\Gamma(1+\alpha)\right)^k \frac{\left(-\sigma^2 t^\alpha\right)^k}{\Gamma(1+\alpha k)} = 1 - \frac{1}{2}\frac{\left(\sigma^2 t^\alpha\right)}{\Gamma(1+\alpha)} + \left(\frac{(\Gamma(1+\alpha))^2}{4\Gamma(1+2\alpha)}\right)(\sigma^4 t^{2\alpha}) - \ldots.$$

If $\sigma^4 << \sigma^2$, then:

$$\ln S(t) = \ln S(0) + R_f t + \ln\left(1 - \frac{1}{2}\frac{(\sigma^2 t^\alpha)}{\Gamma(1+\alpha)}\right)$$

And therefore, $\ln\left(1-\frac{1}{2}\frac{(\sigma^2 t^\alpha)}{\Gamma(1+\alpha)}\right) < 0$. Say that approximately, $\ln\left(1-\frac{1}{2}\frac{(\sigma^2 t^\alpha)}{\Gamma(1+\alpha)}\right) = -\delta t$, thus:

$$\ln S(t) \approx \ln S(0) + (R_f - \delta)t \text{ or } S(0) = e^{-(R_f-\delta)t} S(t)$$

In which case, a fractional risk free bond with the characteristics indicated will value more the, future payout of the bond compared to the risk free. Say, that the bond payout is in 5 years, and let its payout be 1 dollar. Further, let $\sigma^2 = 0.05$ and $\alpha = 0.8$, then

$$0 < \delta = -\frac{1}{5}\ln\left(1 - \frac{1}{2}\frac{((0.05)5^{.8})}{0.8\Gamma(0.8)}\right) = -\frac{1}{5}\ln\left(1 - \frac{0.11324}{\Gamma(0.8)}\right)$$

Which points out a greater price for the bond since its effective price is larger. However if we interpret R_f as a truly risk free bond and σ^2 as a factor due to random factors, then the bond price is smaller the larger δ. In the special case above, $S(0) = e^{-5(R_f-\delta)}$.

5.4.1 The Deterministic Two Time Scales Vacicek Interest Rates (Deterministic) Model

Consider again a two time scales interest rates:

$$dx(t) = -\beta x dt + \sigma(dt)^\alpha, \tilde{x}(0) = R_0 - \bar{R} = 0$$

There are two time intervals $(dt, (dt)^\alpha)$ in the same equation. Namely, mean reversion "is fast" as it is of order dt while there is a "slow deterministic" (and thus constant) perturbation when $\alpha < 1$. This is a deterministic Vacicek model and is used for demonstration purposes. Its' solution may be approached in different ways. First, consider $\beta x dt$ and $\sigma(dt)^\alpha$ separately. Letting $y(t) = e^{\beta t}x(t)$ we have, $dy(t) = \beta e^{\beta t}x(t) + e^{\beta t}dx(t)$. As a result,

$$\left(e^{-\beta t}dy(t) - \beta x(t)dt\right) = -\beta x dt + \sigma(dt)^\alpha, \tilde{x}(0) = R_0 - \bar{R} = 0$$

Or

$$dy(t) = e^{\beta t}\sigma(dt)^\alpha, \tilde{x}(0) = R_0 - \bar{R} = 0 \text{ and } d^\alpha y(t) = \Gamma(1+\alpha)e^{\beta t}\sigma(dt)^\alpha,\ y(0) > 0$$

And therefore,

$$\frac{d^\alpha y(t)}{(dt)^\alpha} = y^\alpha(t) = \Gamma(1+\alpha)e^{\beta t}\sigma, y(0) > 0$$

Whose solution in terms of the Riemann-Liouville equation provides a relationship with the non-fractional in $y(t)$. In this case, $\Gamma(1+\alpha)e^{\beta t}\sigma$ is a first α order derivative whose integral is:

$$y(t) = y(0) + \alpha\sigma\int_0^t (t-\tau)^{\alpha-1}e^{\beta\tau}d\tau$$

Replacing $y(t)e^{-\beta t} = x(t)$, we obtain:

$$x(t) = x(0) + \alpha\sigma \int_0^t (t-\tau)^{\alpha-1} e^{-\beta(t-\tau)} d\tau$$

or equivalently

$$x(t) = x(0) + \sigma \int_0^t e^{-\beta(t-\tau)} (d\tau)^\alpha = x(0) + \alpha\sigma \int_0^t (t-\tau)^{\alpha-1} e^{-\beta(t-\tau)} d\tau, 0 < \alpha < 1$$

Of course, if $dx(t) = \sigma(dt)^\alpha$ (i.e. $\beta = 0$), then: $d^\alpha x(t) = \Gamma(1+\alpha)\sigma(dt)^\alpha$ and $x^\alpha(t) = \Gamma(1+\alpha)\sigma$ where $\Gamma(1+\alpha)\sigma$ is the result of a first order α derivative, and:

$$x(t) = x(0) + \alpha\sigma \int_0^t (t-\tau)^{\alpha-1} d\tau = x(0) + \alpha\sigma \int_0^t u^{\alpha-1} du = \alpha\sigma \frac{t^\alpha}{\alpha} = x(0) + \sigma t^\alpha$$

5.4.2 The Case $1 < \alpha < 2$

The case $1 < \alpha < 2$ corresponds to a shorter memory since $(dt)^\alpha < dt$. For example, if in a fractional model, the variance of stock prices increases at a linear rate $\sigma^2 dt$, then if it were defined by $\sigma w(t)(dt)^{\alpha/2}, \alpha/2 = 0.65 > 1/2$ then $\alpha = 1.2 > 1$. This case corresponds to a increasing nonlinear variance since it is equal to $\sigma^2(t)^\alpha = \sigma^2(t)^{1.2}$. However setting $\beta = \alpha - 1$ or $1 + \beta = \alpha$ in which case $0 < \beta < 1$ (or $\beta = 0.3$). In this case, a series expansion yields (Jumarie 2006),

$$f(x+h) = \sum_{k=0}^{\infty} \frac{h^k}{k!} f^{(k)}(x) + \sum_{k=1}^{\infty} \frac{h^{(\beta k+m)}}{\Gamma(1+m+\beta k)} f^{(\beta k+m)}(x), \beta = \alpha - m$$

And explicitly, for $m = 1$ or $0 < \beta = \alpha - 1 < 1$

$$\begin{aligned} f(x+h) - f^{(0)}(x) &= hf^{(1)}(x) + \frac{h^2}{2} f^{(2)}(x) + \ldots. \\ &+ \frac{h^{(\beta+1)}}{\Gamma(2+\beta)} f^{(\beta+1)} + \frac{h^{(\beta+1)}}{\Gamma(2+2\beta)} f^{(2\beta+1)} + \frac{h^{(3\beta+1)}}{\Gamma(2+3\beta)} f^{(3\beta+1)} + \ldots.. \end{aligned}$$

but if $\beta < 1$ we can retain only these elements $k\beta < 1$ that are smaller than 1. For example, if $\alpha = 1.4, \beta = 0.4 < 1, 2\beta = 0.8 < 1$ and therefore,

$$\begin{aligned} f(x+h) - f^{(0)}(x) &= hf^{(1)}(x) + \frac{h^2}{2} f^{(2)}(x) + \ldots. \\ &+ \frac{h^{(\beta+1)}}{\Gamma(2+\beta)} f^{(\beta+1)} + \frac{h^{(\beta+1)}}{\Gamma(2+2\beta)} f^{(2\beta+1)} \end{aligned}$$

And

$$f(x+h) - f^{(0)}(x) = hf^{(1)}(x) + \frac{h^2}{2} f^{(2)}(x) + \ldots.$$
$$+ \frac{h^{(\beta+1)}}{\Gamma(2+\beta)} f^{(\beta)} f^{(1)} + \frac{h^{(2\beta+1)}}{\Gamma(2+2\beta)} f^{(2\beta)} f^{(1)}$$

The fractional lognormal model $dS(t)/S(t) - R_f dt = -\frac{1}{2}\sigma^2(dt)^{1+\beta}, \alpha = 1+\beta$ may then be written as follows:

$$y(t) = \ln\left(e^{-R_f t} S(t)\right) = -R_f t + \ln S(t) \text{ and}$$
$$dy(t) = -R_f dt + \frac{dS(t)}{S(t)} = -\frac{1}{2}\sigma^2(dt)^{1+\beta}$$

As a result,

$$dy(t) = -\frac{1}{2}\sigma^2(dt)^{1+\beta} \text{ or } y^{1+\beta}(t) = \frac{d^{1+\beta}y(t)}{(dt)^{1+\beta}}$$
$$= \frac{1}{2}\Gamma(2+\beta)\sigma^2 \text{ or } y^{1+\beta}(t) = y^1(t)y^\beta(t)$$

A solution within a Fractional Brownian motion case is considered below.

6 The Fractional Brownian Motion and the Hurst Index H

The Fractional Brownian motion with an index H, is, defined by $\left\{W^H(t), t \geq 0\right\}, \forall t \in \Re_+$ with the following elementary properties: Pr $\left\{W^H(0) = 0\right\} = 1$, $\left\{W^H(t), t \geq 0\right\}$ is a measurable random variable such that it has a null mean and a covariance given by:

$$E\left\{W^H(t)W^H(\tau)\right\} = \frac{\sigma^2}{2}\left(t^{2H} + \tau^{2H} - |t-\tau|^{2H}\right).$$

The variance equation is a solution of the functional equation: $xy = x^2 + y^2 - (x-y)^2$ an therefore, assuming that $x = t^{2H}$, the fractional variance equation above is obtained. Thus setting H = 1/2, $E\{W^{\frac{1}{2}}(t)W^{\frac{1}{2}}(\tau)\} = \sigma^2 t$ which corresponds to standard Brownian motion variance $E\{[W^H(t)]^2\} = \sigma^2(dt)^{2H}$. Its self-similarity is implied also by: $W^H(\rho t) \overset{id}{=} \rho^H W^H(t)$. And an Ito-Like Lemma, for a twice continuous function: $f\left(t, W^H(t)\right)$:

$$df\left(t, W^H(t)\right) = \frac{\partial}{\partial t} f\left(t, W^H(t)\right) dt + \frac{\partial f\left(t, W^H(t)\right)}{\partial W^H(t)} dW^H(t)$$
$$+ \frac{1}{2}\frac{\partial^2 f\left(t, W^H(t)\right)}{\partial\left(W^H(t)\right)^2}(d(t))^{2H}$$

Interestingly, we see that in such a case, the expected differential $E(df)$ may have multiple scale as indicated earlier, since:

$$E(df) = \left(\frac{\partial f}{\partial t}\right) dt + \frac{1}{2} E\left(\frac{\partial^2 f}{\partial (W^H(t))^2}\right) (dt)^{2H}$$

Of course when H = 1/2, we obtain the Taylor series expansion of a Brownian motion model.

The relationship between a Brownian motion and a fractional Brownian motion limited to a positive time interval is given by a kernel (although there may be others) that relates the fractional Brownian motion with a Brownian motion:

$$K_H(t,\tau) = \frac{(t-\tau)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} \quad \text{with} \quad W^H(t) = \int_0^t K_H(t,\tau) dW(t)$$

where $\Gamma\left(H+\frac{1}{2}\right)$ is the Euler Gamma function. The manipulation of these integrals and their implications to financial modeling will be outlined below. First we shall consider a random walk approximation to such fractional stochastic integral. Subsequently, application of a fractional calculus and examples are considered.

From a financial viewpoint, fractional Brownian motion models have important implications for many of the paradigms used in modern financial economics. For example, optimal consumption-savings and portfolio decisions may become extremely sensitive to the investment horizon if stock prices returns were long range dependent. Problems also arise in the pricing of options and futures since the class of models used are incompatible with long term memory. Traditional tests of the capital asset pricing model and the APT (Arbitrage Pricing Theory) are no longer valid since the usual forms of statistical inference do not apply to time series exhibiting such dependence. Tests of an "efficient" markets hypothesis depend therefore precariously on the presence or absence of long term memory. From a fractional viewpoint, note as well that the time scale changes and therefore the information and the calculations implied in the model we use change.

Memory based models such as ARCH-GARCH stochastic volatility models (Engle 1987, Bollerslev 1986), Fractal Brownian motion and Multi Fractals stochastic models (Bianchi et al., this book) have been constructed potentially to better explain the leptokurtic character of rates of returns distributions. Other studies have shown that distributions have tails fatter than the normal distribution. Further, stochastic volatility and fractional Brownian motion models have shown that non-linear volatility growth and non-linear dependence may be observed in fact. In such cases, the assumptions of a "linear time finance" may in practice be doubtful. References to these models and the problems they deal with are numerous. Below we shall consider such models and their mathematical framework that are of particular usefulness in finance and in other areas where long run memory are observed in fact and are not accounted for in conventional stochastic models.

6.1 A Random Walk Approximation to Fractional Brownian Motion

Consider the fractional Brownian motion integral, which as noted below is a weighted function of standard normal events which we approximate by random walks, or:

$$y_H(t) = \sigma \int_0^t W^H(\tau)d\tau = \frac{\sigma}{\Gamma\left(H+\frac{1}{2}\right)} \int_0^t (t-\tau)^{H-\frac{1}{2}-1} W(\tau)d\tau$$

This relationship is proved again by an application of Ito's differential:

$$\int_0^t h(t,\tau)dW(\tau) = h(t,t)W(t) - \int_0^t \frac{\partial h(t,\tau)}{\partial \tau} W(\tau)d\tau$$

The leads to: $\int_0^t (t-\tau)^{H-\frac{1}{2}} dW(\tau) = \left(H-\frac{1}{2}\right) \int_0^t (t-\tau)^{H-\frac{1}{2}-1} W(\tau)d\tau$ and therefore to the equation above. For convenience, set $\alpha = (H-\frac{1}{2})$, then:

$$y_\alpha(t) = \frac{\sigma\alpha}{\Gamma(1+\alpha)} \int_0^t (t-\tau)^{\alpha-1} W(\tau)d\tau, \alpha = \left(H-\frac{1}{2}\right)$$

And let ε_{k+1} be a series of standardized random walks with mean zero and variance 1. In this case, we construct the following approximation:

$$\int_0^t (t-\tau)^\alpha dW(\tau) = \underset{n\to\infty}{Lim}\sqrt{n} \left\{ \sum_{k=0}^{[nt]} \left(\int_{k/n}^{(k+1)/n} (t-\tau)^{\alpha-1} 1_{\{\tau<t\}} d\tau \right) \varepsilon_{k+1} \right\}$$

With its estimate in the time interval $\{(k+1)/n - (k)/n\}$ given by:

$$\int_{k/n}^{(k+1)/n} (t-\tau)^{\alpha-1} d\tau = \frac{1}{\alpha}\left[-(t-\tau)^\alpha\right]\Big|_{k/n}^{(k+1)/n} = \frac{1}{\alpha}\left[\left[t-\frac{k}{n}\right]^\alpha - \left[t-\frac{k+1}{n}\right]^\alpha\right]$$

And therefore,

$$\begin{aligned} y_\alpha(t) &= \frac{\sigma}{\Gamma(1+\alpha)}\left[\left[t-\frac{k}{n}\right]^\alpha - \left[t-\frac{k+1}{n}\right]^\alpha\right] \\ &= \underset{n\to\infty}{Lim}\frac{\sqrt{n}}{\Gamma(1+\alpha)} \left\{ \sum_{k=0}^{[nt]} \left(\left[t-\frac{k}{n}\right]^\alpha - \left[t-\frac{k+1}{n}\right]^\alpha \right) \varepsilon_{k+1} \right\} \end{aligned}$$

which converges in probability to the stochastic integral as stated above and provides a random walks approximation to the long run memory process. Of course, the time derivative is :

$$dy_\alpha(t) = \frac{\sigma}{\Gamma(\alpha)} \underset{n\to\infty}{Lim} \sqrt{n} \left\{ \sum_{k=0}^{[nt]} \left(\left[t - \frac{k}{n} \right]^{\alpha-1} - \left[t - \frac{k+1}{n} \right]^{\alpha-1} \right) \varepsilon_{k+1} \right\} dt$$

Thus, a random walk approximation to a fractional mean reversion model is

$$dx_\alpha(t) = -ax_\alpha(t)dt + dy_\alpha(t)$$

And explicitly:

$$\begin{aligned} dx_\alpha(t) =& - ax_\alpha(t)dt \\ &+ \frac{\sigma}{\Gamma(\alpha)} \underset{n\to\infty}{Lim} \sqrt{n} \left\{ \sum_{k=0}^{[nt]} \left(\left[t - \frac{k}{n} \right]^{\alpha-1} - \left[t - \frac{k+1}{n} \right]^{\alpha-1} \right) \varepsilon_{k+1} \right\} dt \end{aligned}$$

Again, setting $z_\alpha(t) = e^{at}x_\alpha(t)$ we have $e^{-at}dz_\alpha(t) - ax_\alpha(t) = dx_\alpha(t)$ and therefore,

$$dz_\alpha(t) = \frac{\sigma}{\Gamma(\alpha)} \underset{n\to\infty}{Lim} \sqrt{n} \left\{ \sum_{k=0}^{[nt]} \left(e^{at} \left[t - \frac{k}{n} \right]^{\alpha-1} - e^{at} \left[t - \frac{k+1}{n} \right]^{\alpha-1} \right) \varepsilon_{k+1} \right\} dt$$

And finally:

$$\begin{aligned} z_\alpha(t) = z_\alpha(0) + \frac{\sigma}{\Gamma(\alpha)} \underset{n\to\infty}{Lim} \sqrt{n} \\ \left\{ \sum_{k=0}^{[nt]} \varepsilon_{k+1} \left(\int_0^t e^{a\tau} \left[\tau - \frac{k}{n} \right]^{\alpha-1} d\tau - \int_0^t e^{a\tau} \left[\tau - \frac{k+1}{n} \right]^{\alpha-1} d\tau \right) \right\} \end{aligned}$$

6.2 A Note: On the Random Walk Approximation:

*C*onsider a sequence of independent random walks with $\varepsilon_i, i \geq 1$ denoting random variables identically and independently distributed with $P(\varepsilon_i = \pm 1) = 1/2$, for all $i \geq 1$. We construct the random walk : $x_k = \sum_{i=1}^{k} \varepsilon_i, k \geq 1, x_0 = 0$ and consider the function defined by : $x_t^{(n)} = \frac{1}{\sqrt{n}}(x_{[tn]} + (nt - [nt])\varepsilon_{[nt+1]})$ where the brackets [..] are used to denote the integer number of its argument. This is equivalent again to a piecewise linear approximation of the stochastic process where the time interval is divided into n equal intervals of length $1/n$ as seen in the figure below. In the interval $\left[\frac{k}{n}, \frac{k+1}{n}\right]$, $x^{(n)}$is an affine function and in particular, we havethe linear approximation

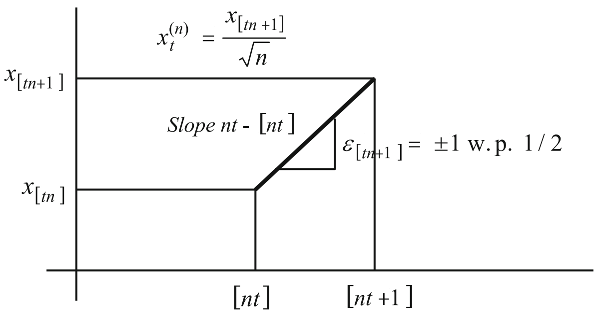


Fig. 1 Piecewise linear approximation

defined by: $x_{k/n}^{(n)} = \frac{1}{\sqrt{n}} x_k$ at the kth point and therefore, in terms of standardized values:

$$x_t^{(n)} = x_{k/n}^{(n)} + n\left(t - \frac{k}{n}\right)\left(x_{(k+1)/n}^{(n)} - x_{k/n}^{(n)}\right) \quad \text{for} \quad t \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$$

between the kth and the (k + 1)st points. At the kth point, it can be verified that Fig. 1,

$$x_{k/n}^{(n)} = \frac{1}{\sqrt{n}}\left(x_k + (1-1)\varepsilon_{[nt+1]}\right) = \frac{1}{\sqrt{n}}(x_k)$$

Within each segment, the time derivative is obviously:

$$\dot{x}_t^{(n)} = \frac{dx_t^{(n)}}{dt} = n\left(x_{(k+1)/n}^{(n)} - x_{k/n}^{(n)}\right) = \sqrt{n}\varepsilon_{k+1}, t \in \left]\frac{k}{n}, \frac{k+1}{n}\right[$$

With these definitions on hand, we consider the stochastic integral defined below and to which we apply Ito's differential rule,

$$\int_0^t h(s)dw(s) = h(t)w(t) - \int_0^t h'(s)w(s)ds$$

where $h(s)$assumes at least a first derivative denoted by $h'(s)$. By Donsker's theorem, we know that the process $x^{(n)}$ converges in probability law to the Brownian motion when $n \to \infty$. Thus, the application of Ito's differential rule can be written equivalently by :

$$\int_0^t h(s)dw(s) = \underset{n\to\infty}{Lim}\left\{h(t)x_t^{(n)} - \int_0^t h'(s)x_s^{(n)}ds\right\},$$

Integration by parts within each interval yields therefore,

$$h(t)x_t^{(n)} - \int_0^t h'(s)x_s^{(n)}ds = \int_0^t h(s)\dot{x}_s^{(n)}ds$$

and as a result, we have the integral expressed by a sum of random walks :

$$\int_0^t h(s)dw(s) \approx \sum_{k=0}^{[nt]} \left(\int_{k/n}^{(k+1)/n} h(s)ds \right) \left(x^{(n)}_{\frac{k+1}{n}} - x^{(n)}_{\frac{k}{n}} \right)$$
$$= \sqrt{n} \left\{ \sum_{k=0}^{[nt]} \left(\int_{k/n}^{(k+1)/n} h(s) 1_{\{s<t\}} ds \right) \varepsilon_{k+1} \right\}$$

where $1_{\{s<t\}}$ is a function which is equal to 1 as long as $s < t$ and it equals zero otherwise. Note that the random walk approximation does not require that the first derivative of the function $h(.)$ exists. Equivalently, we can consider the limit distribution of the random walks and show that it converges in probability law to the Brownian motion. The implication of this result were used above to the approximation of the fractal stochastic integral.

$$y_\alpha(t) = \int_0^t (t-\tau)^\alpha dW(\tau), \alpha \neq 0$$

6.3 *Fractional Calculus and the Fractional Lognormal Model*

We consider next the fractional lognormal Brownian motion model with a Hurst index:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dW^H(t), S(0) = S_0 > 0 \text{ with } W^H(t) = \int_0^t \frac{(t-\tau)^{H-\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)} dW(\tau)$$

As indicated previously, a fractional Brownian motion can be approximated by a converging sum of random walks. In this section, we apply a stochastic fractional calculus. Define for convenience the functional transformation $y(t) = \ln S(t)$. Applying Ito calculus rules, we have:

$$dy(t) = \frac{dS}{S} - \frac{1}{2}\frac{(dS)^2}{S^2} \quad \text{or}$$
$$dy(t) = \frac{dS}{S} - \frac{1}{2}\frac{(dS)^2}{S^2} = \mu dt - \frac{1}{2}\sigma^2 \left(dW^H(t)\right)^2 + \sigma dW^H(t)$$

And therefore, the fractional differential equation:

$$dy(t) = \mu dt - \frac{1}{2}\sigma^2 (dt)^{2H} + \sigma dW^H(t)$$

And explicitly in terms of the stock price:

$$S(t) = S(0)\exp\left\{\mu t - \frac{1}{2}\sigma^2 \int_0^t (dt)^{2H} + \sigma \int_0^t dW^H(t)\right\}$$

We consider first the deterministic (albeit fractional) part given by (when $0 < 2H < 1$:

$$Z(t) = \exp\left\{\mu t - \frac{1}{2}\sigma^2 \int_0^t (dt)^{2H}\right\} \quad \text{or} \quad dZ(t) = \mu Z(t)dt - \frac{1}{2}\sigma^2 Z(t)(dt)^{2H}$$

Whose solution (treated earlier) is defined by $Z(t) = Z_1(t)Z_2(t)$, with $Z_1(t) = Z_1(0)e^{\int_0^t \mu dt}$

However, $\frac{dZ_2(t)}{Z_2} = -\frac{1}{2}\sigma^2(dt)^{2H}$ can be written by:

$$\frac{d^{2H}Z_2(t)}{Z_2} = -\frac{1}{2}\Gamma(1+2H)\sigma^2(dt)^{2H}, 0 < 2H < 1$$

Whose solution is given by the Mittag-Leffler equation:

$$Z_2(t) = E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2 t^{2H}\right) \quad \text{with}$$

$$E_{2H}(h) = \sum_{k=0}^{\infty} \frac{h^k}{\Gamma(1+2Hk)} = \sum_{k=0}^{\infty} \frac{h^k}{(2Hk)\Gamma(2Hk)}$$

Note that if H = 1/2, then $Z_2(t) = E_1(-\sigma^2 t)$ and

$$E_1(h) = \sum_{k=0}^{\infty} \frac{h^k}{\Gamma(1+k)} = 1 + \frac{h^2}{\Gamma(2)} + \frac{h^3}{\Gamma(3)} + ...$$

and thus $Z_2(t) = E_1(-\sigma^2 t) = e^{-\sigma^2 t}$

And therefore

$$Z(t) = Z_1(0)e^{\int_0^t \mu dt} E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2 t^{2H}\right)$$

The solution of the stochastic fractional differential equation is:

$$S(t) = S(0)e^{\int_0^t \mu dt + \sigma \int_0^t dW^H(t)} E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2 t^{2H}\right)$$

Where,

$$E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2 t^{2H}\right) = \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2\right)^k}{(2Hk)\Gamma(2Hk)} t^{2Hk}$$

And an explicit development of the first terms yields:

$$E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\,\sigma^2 t^{2H}\right)$$
$$=1-\frac{1}{2}\sigma^2 t^{2H}+\frac{1}{4}\frac{(\Gamma(1+2H))^2}{\Gamma(1+4H)}\sigma^4 t^{4H}-\frac{1}{8}\frac{(\Gamma(1+2H))^3}{\Gamma(1+6H)}\sigma^6 t^{6H}+....$$

Replacing the fractional Brownian motion by its time equivalent, we have then:

$$S_H(t)=S(0)E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2 t^{2H}\right)\exp$$
$$\left\{\mu+\frac{\sigma}{\Gamma\left(H+\frac{1}{2}\right)}\int_0^t (t-\tau)^{H-\frac{1}{2}}W(\tau)d\tau\right\}$$

Which is not of course a Martingale and therefore it is not a financial pricing equation. When H = 1/2, this pricing model is reduced to the standard model, since the price is reduced to:

$$S(t)=S(0)\exp\left\{\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma W(t)\right\}$$

The Hurst index thus provides a measure for departure from the standard model. Explicitly, under a probability measure $W^Q(t)=W(t)+(\frac{\mu-R_f}{\sigma})t$, the following pricing (Martingale) measure is obtained:

$$S(t)=S(0)\exp\left\{\left(R_f-\frac{\sigma^2}{2}\right)t+\sigma W^Q(t)\right\}$$

And therefore,

$$S(0)=E^Q\left\{e^{-\left(R_f-\frac{\sigma^2}{2}\right)t+\sigma W^Q(t)}S(t)\right\}=E^Q\left\{e^{-R_f t}S(t)\right\}$$

Replacing $W(t)=W^Q(t)-(\frac{\mu-R_f}{\sigma})t$ we have a fractional model measuring departure from market completeness. When, H = 1/2 it is by definition a pricing Martingale measure while when 0 < H < 1/2, it is not a pricing measure as it is incomplete. In this sense, the Hurst index measures a departure from the complete markets hypothesis with a current price given by:

$$S(0)=E^Q$$
$$\left(\frac{e^{-\left\{\frac{\sigma}{\Gamma(H+\frac{1}{2})}\left(\int_0^t (t-\tau)^{H-\frac{1}{2}}W^Q(\tau)d\tau+\int_0^t (t-\tau)^{H-\frac{1}{2}}\left(\frac{\mu-R_f}{\sigma}\right)\tau d\tau\right)\right\}}}{E_{2H}\left(-\frac{1}{2}\Gamma(1+2H)\sigma^2 t^{2H}\right)}S_H(t)\right)$$

Since H < 1/2, this corresponds to time series that tend to revert to their trend. The variance of the underlying process will be smaller than if H > 1/2, in which case, the evolution of the variance tends to increase at a nonlinear rate. The financial implications of a fractional lognormal process as defined above are important as they provide a preliminary approach to pricing the cost of incompleteness. Of course, if $H = \frac{1}{2}$ then, $E^Q\left(S_{\frac{1}{2}}(t)\right) - E^Q(S(t)) = 0$. While we may expect that since for H < 1/2, the process variance may grow at a sub-linear rate that $E^Q(S_{H<1/2}(t)) - E^Q(S(t)) < 0$, the difference accounting for a premium that one pays assuming that the market is complete. Inversely, when the Hurst index is greater than ½, time series have a tendency to have an increasing variance over time since $t^{2H} > t$ (see also Vallois and Tapiero and Tapiero and Vallois numerous references at the end of the paper for the estimation of the Hurst index using range processes). Such situations imply financial models that have unpredictable states with a variance growing much more than presumed by the normal Brownian motion model. In this case, we expect that $E^Q(S_{H>1/2}(t)) - E^Q(S(t)) > 0$ which implies a financial risk premium far greater than presumed by a complete market model. The calculation of such a price is however more complex.

Let $1 < 2H < 2$ which corresponds to fractional time $(dt)^{2H} < dt$ and define $\beta = 2H - 1$ and therefore, $1 + \beta = 2H$ as well as $0 < \beta < 1$. The fractional lognormal model is thus: $dS(t)/S(t) - R_f dt = -\frac{1}{2}\sigma^2 (dt)^{1+\beta}$ and as indicated earlier or $y^{1+\beta}(t) = \frac{d^{1+\beta}y(t)}{(dt)^{1+\beta}} = -\frac{1}{2}\Gamma(2+\beta)\sigma^2$ Or $y^{1+\beta}(t) = y^1(t)y^\beta(t)$. Consider the Laplace Transform of the equation above where we take a first a derivative with respect to 1 and then to β, then

$$L^*(D^{(1+\beta)}y(t)) = p^{1+\beta}L^*(y(t)) - p^{1+\beta-1}y^{(1)}(0) - p^{1-1}y^{(\beta)}(0),$$

A first integration leads to

$$L^*(D^{(1+\beta)}y(t)) = p^{1+\beta}L^*(y(t)) - p^\beta y(0) - y^{(\beta)}(0)$$

And therefore, $y^{(1+\beta)}(t) = (dy(t)/dt)(\sigma/\lambda)\Gamma(1+\alpha)\sigma$. A first integration yields:

$$\frac{dy(t)}{dt} = \frac{dy(0)}{dt}E_\beta\left(-\frac{1}{2\lambda}\sigma^2\Gamma(2+\beta)t^\beta\right),$$

While a second integration yields:

$$y(t) = \frac{dy(0)}{dt}\int_0^t E_\beta\left(-\frac{1}{2\lambda}\sigma^2\Gamma(2+\beta)\tau^\beta d\tau\right), 1 < 2H < 2 \text{ and } \beta = 2H - 1$$

$$y(t) = \frac{dy(0)}{dt}\int_0^t E_\beta\left(-\frac{1}{2\lambda}\sigma^2\Gamma(2+\beta)\tau^\beta d\tau\right), 1 < 2H < 2$$

7 Long Run Memory and Fractional Models: Theoretical Properties

A stationary process X_t for which there exists a real number $\alpha \in (0,1)$ and a constant $C_\rho > 0$ is said to have lon run memory if $\underset{t\to\infty}{Lim}\,\rho(t)/(C_\rho t^{-\alpha}) = 1$. X_t has other names such as long range dependence, strong dependence, or a stationary process with slowly decaying or long range correlation. Such fractional models are also defined by the Hurst exponent, H (often used in Bloomberg, as a "chaos coefficient" and more formally, a self-similarity index, see Vallois 1993, 1995, 1996; Vallois and Tapiero 1995, 1996, 1997; Tapiero and Vallois 2007 for numerous studies on the Hurst index). In this case,

$$\underset{t\to+\infty}{Lim}\frac{\operatorname{var}\left(\sum_{i=1}^{t} X_i\right)}{C_\gamma t^{2H}} = \frac{1}{H(2H-1)}$$

It implies that a time series innovation $\varepsilon_t = y_t - E_{t-1}(y_t)$, has a fractional property or a long run memory. For $H = 1/2$, as shown above, we have a linear time variance while for $1/2 < H < 1$ we have a fractional series with self-similarity parameter H. The implication of this (Hurst) self-similarity index is that although increments are independent, the following holds:

$$Y_t \overset{(d)}{=} t^H Y_1, t > 0 \text{and } Y_1 \neq 0 \text{ with positive probability.}$$

This means that:

$$H < 0 \Rightarrow Y_t \overset{(d)}{\to} 0$$

$$H = 0 \Rightarrow Y_t \overset{(d)}{\to} Y_1$$

$$H > 0 \text{ and } Y_t \neq 0 \Rightarrow |Y_t| \overset{(d)}{\to} \infty$$

Thus, excluding an initial condition, $Y_t = 0$, Y_t is not stationary, unless $H = 0$. If $H > 0$, variance growth is nonlinear, or: $E(Y_t - Y_s)^2 = \sigma^2(t-s)^{2H}$ while the covariance is:

$$E(Y_t Y_s) - E(Y_t)E(Y_s) = \gamma_Y(t,s) = \frac{1}{2}\sigma^2\Big[t^{2H} - (t-s)^{2H} + s^{2H}\Big].$$

By the same token, if we consider increments $X_i = Y_i - Y_{i-1}, i = 1,2,3,\cdots$ the covariance of these increments is: $\gamma_X(k) = \operatorname{cov}(X_{i,}, X_{i+k}) = \operatorname{cov}(X_{1,}, X_{k+1})$. Using the self-similarity property, we have then:

$$\gamma_X(k) = \frac{1}{2}\sigma^2\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \text{ and}$$

$$\rho_X(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right].$$

It is then simple to show (as stated above) that:

$$\left[\frac{\rho_X(k)}{H(2H-1)k^{2H-2}}\right]_{k\to+\infty}^{\to}1$$

This means that if the correlation decays slowly such that for $1/2 < H < 1$, $\sum_{k=-\infty}^{+\infty}\rho_X(k) = +\infty$ leads to a series being unpredictable (i.e. in this case, an infinite variance which differs from Knight's uncertainty that presumes that future states are unknown). While for a Hurst exponent of $0 < H < 1/2$, $\sum_{k=-\infty}^{+\infty}\rho_X(k) = 0$. The case, although very rare has been observed while studying trades of Florentine proveditori degli cambiatori, a special flower traded over long periods of time in the Middle Ages. Similar results are obtained when we consider the Brownian motion. In this case, transformation of variables yields: $EB(ct) = E[B(ct) - B(0)] = \sqrt{c}EB(t) = 0$.

The fractional Brownian motion expressed in terms of a self-similarity index H is thus (where $B(u)$ is a standard Brownian motion):

$$B_H(t) = C_H\int w_H(t,u)dB(u) \quad \text{with} \quad \begin{cases} w_H(t,u) = 0 \text{ for } t \le u \\ w_H(t,u) = (t-u)^{H-\frac{1}{2}} \text{ for } 0 \le u \le t \\ w_H(t,u) = (t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} \text{ for u} < 0 \end{cases}$$

Elementary manipulations will then show that: $w_H(ct,u) = c^{H-\frac{1}{2}}w_H(t,uc^{-1})$ and therefore,

$$B_H(ct) = C_H\int w_H(ct,u)dB(u) = C_Hc^{H-\frac{1}{2}}\int w_H(t,uc^{-1})dB(u)$$

Substituting $v = uc^{-1}$, we have: $C_Hc^{H-\frac{1}{2}}\int w_H(t,v)dB(cv)$ and by self-similarity, we obtain at last:

$$C_Hc^{H-\frac{1}{2}}\int w_H(t,v)dB(cv) = C_Hc^{\frac{1}{2}}c^{H-\frac{1}{2}}\int w_H(t,v)dB(v) = c^HB_H(t),$$

For additional references see Bunde and Havlin 1991; Feller 1951; Hulliet 2002; Hurst 1951; Imhoff 1985, 1992, Jain and Orey 1968; Jain and Pruitt 1972; Peter 1995; Scalas 2006; Viswanathamn et al. 1999; Barkai 2001; Daley 1999.

8 Stratonovich Calculus

We introduce at this time the Stratonovich calculus as it can be useful in applying fractional calculus. A formal stochastic differential equation may be written as we saw above by:

$$dx(t) = f(x,t)dt + \sigma(x,t)dW, x(0) = x_0$$

The stochastic variable x is defined, however, only if the above equation is meaningful or if its stochastic integral:

$$\int_{t_0}^{t} \sigma(x,t)dW$$

is meaningful and computable. In which case, the following solution follows:

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x,t)dt + \int_{t_0}^{t} \sigma(x,t)dW$$

The first integral is well defined in Riemann's calculus. The second integral involves however, a random variable with an unbounded variation due to the Wiener process, $dW = W(t+dt)-W(t)$. This unbounded variation introduces some difficulty requiring that we specify precisely what we mean by random (or stochastic) integrals. This is equivalent to specifying the computational method which allows the calculation of the stochastic integral. The methods we apply to computing these stochastic integrals, differentiates the stochastic calculus we apply. To define this integral we partition the time interval $\Gamma : [t_0, t]$ into N steps of length $t^{\ell}_{j+1} - t^{\ell}_{j}$, $j = 0,1,\cdots N-1$ where $\ell = \max(t^{\ell}_{j+1} - t^{\ell}_{j})$ and let $t = t_N$. Then we proceed by letting ℓ be the maximum time difference $t^{\ell}_{j+1} - t^{\ell}_{j}$, tend to zero, and obtain the Ito stochastic integral:

$$\int_{\Gamma} \sigma(x,t)dW = \underset{\ell \to 0}{Lim} \sum_{j=1}^{N-1} \sigma(x(t^{\ell}_{j}), t^{\ell}_{j}) \left[W(t^{\ell}_{j+1}) - W(t^{\ell}_{j})\right]$$

However, using an alternative calculation of the stochastic integral, say as that given below, then its value may not be the same as that of the Ito-integral. Rather, we have an integral suggested by Stratonovich whose basic difference is that it follows the Riemanian (deterministic) calculus. In this case, we have the following procedure to calculate a stochastic integral:

$$\int_{\Gamma} \sigma(x,t)dW = \underset{\ell \to 0}{Lim} \sum_{j=1}^{N-1} \sigma\left(\frac{x(t^{\ell}_{j}) + x(t^{\ell}_{j+1})}{2}, t^{\ell}_{j}\right) \left[W(t^{\ell}_{j+1}) - W(t^{\ell}_{j})\right]$$

For these integrals to be computable, two essential terms must be understood. First, the concept of non-anticipating function and second the notion of mean square convergence. The function $\sigma(x,t)$is non anticipating if for all times s and t, the function is independent of the random future events $W(s)-W(t)$. This means that the function $\sigma(x,t)$is independent of future values that the Wiener process will assume (which is a reasonable assumption for most practical and tractable problems). This independence allows us to multiply the function and future values of the Wiener process and compute its moments (since there will be no correlation between the two). Mean square, or m.s. stands however, for existence of the limit sum which computes the stochastic integral in the sense that the squared error is minimal.

Although the Ito and the Stratonovich definition of integrals are not the same they are related to one another by a relationship which holds with probability one. In other words, discretizing a random process following the Ito or Stratonovich stochastic procedures will not lead to the same results and will thus involve stochastic calculus rules that tend to differ. They are, however, measuring the "same thing" and therefore, the relationship between their integration rules must remain the same. This is summarized by the following:

$$\int_{Straton.} \sigma(x(t),t)\, dW(t) = \int_{Ito} \sigma(x(t),t) dW(t) + \frac{1}{2}\int_{Ito} \frac{\partial\sigma(x(t),t)}{\partial x}\sigma(x(t),t) dW(t)$$

Further, if we were to use a Langevin formalism (with $w(t) = dW/dt$) then, we could write as well:

$$\int_{Straton.} \sigma(x(t),t) w(t)(dt)^{\frac{1}{2}} \text{ which can be generalized to: } \int_{Straton.} \sigma(x(t),t) w(t)(dt)^{2\alpha}$$

to denote an alternative calculus we shall be concerned with and based on the Riemann-Liouville function.

Consider at present a random function $y = h(t, x)$. Taylor series developments (differential rules) under the Ito and the Stratonovich calculus, as indicated above, will not be the same. Under a Stratonovich calculus, the rules of Riemanian calculus are applied, in which case:

$$dy_S = \frac{\partial h}{\partial t} dt + \frac{\partial h}{\partial x} dx$$

While under an Ito calculus, we have:

$$dy_I = \frac{\partial h}{\partial t} dt + \frac{\partial h}{\partial x} dx + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(dx)^2$$

And therefore,

$$dy_S = dy_I + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(dx)^2 \text{ and } dy_I = dy_S - \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(dx)^2$$

An equivalence od stochastic differential equations may then be proved. Let the Ito stochastic differential equation:

$$dx_I = f(x,t)dt + \sigma(x,t) dW_I(t)$$

Then, its equivalent Stratonovich stochastic differential equation (or $dx_I \equiv dx_S$) is:

$$dx_S = \left(f(x,t) - \frac{1}{2}\sigma(x,t)\frac{\partial\sigma(x,t)}{\partial x}\right) dt + \sigma(x,t) dW_S(t)$$

In this sense, a drift alters the mean evolution of the process. Inversely, if a Stratonovich differential equal is given by:

$$dx_S = f(x,t)dt + \sigma(x,t)dW_S(t)$$

Then, its equivalent Ito stochastic differential equation is specified by:

$$dx_I = \left(f(x,t) + \frac{1}{2}\sigma(x,t)\frac{\partial \sigma(x,t)}{\partial x} \right) dt + \sigma(x,t)dW_I(t)$$

For example, if $\sigma(x,t) = \sigma(t)$, then both equations provide the same results. However for a lognormal Ito stochastic differential equation, $\sigma(x,t) = \sigma x$ we have the Stratonovich Ito equivalent equation given by:

$$dx_S = \left(\mu x - \frac{1}{2}\sigma^2 x \right) dt + \sigma x dW_S(t) \text{ or } \frac{dx_S}{x} = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_S(t)$$

Which is under the Ito calculus the transformation of $y_I = \ln(x_I)$. Similarly, consider the Taylor series expansion of the Ito lognormal process:

$$\frac{dx_I}{x} = \mu dt + \sigma dW_I(t) \text{ and setting } y(x) = \ln x, \text{ thus: } \frac{\partial y}{\partial x} = \frac{1}{x}, \frac{\partial^2 y}{\partial x^2} = -\frac{1}{x^2} \text{ and:}$$

While under the Stratonovitch calculus, we have:

$$dy_S = dy_I - \frac{1}{2}\frac{1}{x^2}(dx)^2 \quad \text{and} \quad dy_I = dy_S + \frac{1}{2}\frac{1}{x^2}(dx)^2$$

The financial implications is that computing numerically a financial stochastic integral model may either point out to a an increase in rates of returns or in its decrease—relative to the computational (Ito or Stratonovich) approach used.

Using an integration form to Ito stochastic differential equation,

$$dx = \sigma(x(t),t)dW_I$$

Then since

$$\int_{Straton.} \sigma(x(t),t)dW_S(t) = \int_{Ito} \sigma(x(t),t)dW_I(t) + \frac{1}{2}\int_{Ito} \frac{\partial \sigma(x(t),t)}{\partial x}\sigma(x(t),t)dW_I(t)$$

Or

$$\int_{Straton.} \sigma(x(t),t)dW_S(t) = \int_{Ito} \left(\sigma(x(t),t) + \frac{1}{2}\frac{\partial \sigma(x(t),t)}{\partial x}\sigma(x(t),t) \right) dW_I(t)$$

Or

$$\int_{Straton.} \sigma(x(t),t)dW_S(t) = \int_{Ito} \left(\frac{1}{2}\frac{\partial \sigma(x(t),t)}{\partial x}\sigma(x(t),t)dt \right) + \int_{Ito} \sigma(x(t),t)dW_I(t)$$

And therefore

$$dx = \sigma(x(t), t)dW_S(t) = \left(\frac{1}{2}\frac{\partial\sigma(x(t), t)}{\partial x}\sigma(x(t), t)dt\right) + \sigma(x(t), t)dW_I(t)$$

And inversely,

$$dx = \sigma(x(t), t)dW_I(t) = \sigma(x(t), t)dW_S(t) - \left(\frac{1}{2}\frac{\partial\sigma(x(t), t)}{\partial x}\sigma(x(t), t)dt\right)$$

Where numerically, the integrals are calculated by:

$$\sigma(x, t)dW = \sigma(x(t_j), t_j)\left[W_I(t_{j+1}) - W_I(t_j)\right] \quad \text{(Ito)}$$

$$\sigma(x, t)dW = \sigma\left(\frac{x(t_{j+1}) + x(t_j)}{2}, t_j\right)\left[W_S(t_{j+1}) - W_S(t_j)\right] \quad \text{(Stratonovich)}$$

For example, consider a Stratonovich stochastic differential equation is given by a generalized lognormal process:

$$dx(t) = \mu x^q dt + \sigma x^q dW_S(t) \quad \text{with} \quad \partial(\sigma x^q)/\partial x = q\sigma x^{q-1}$$

In which case,

$$\begin{aligned}\frac{dx_S}{x^q} &= \mu dt + \sigma dW_S \\ \frac{dx_I}{x^q} &= \left(\mu + \frac{1}{2}q\sigma^2 x^{q-1}\right)dt + \sigma dW_I(t)\end{aligned}$$

Which results with q = 1 to their equivalent values for a lognormal process. Inversely, we have:

$$\begin{aligned}\frac{dx_I}{x^q} &= \mu dt + \sigma dW_I \\ \frac{dx_S}{x^q} &= \left(\mu - \frac{1}{2}q\sigma^2 x^{q-1}\right)dt + \sigma dW_S(t)\end{aligned}$$

The choice of a mathematical system and the computation of stochastic integrals in financial analysis thus matters. Although in the cases treated above, we obtained an equivalence, allowing to define one approach with respect to the other, is clearly specifies that some theoretical financial results may be valid under a mathematical system, it might not be valid under another. The same rational applies to the definitions of stochastic integrals in fractional calculus.

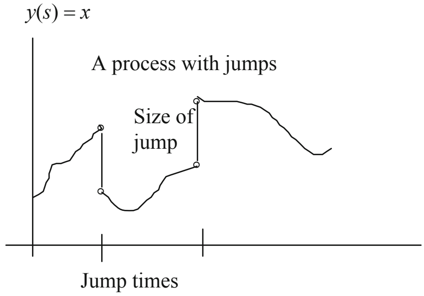


Fig. 2 A jump process

9 The Poisson Process and Long Run Memory

Assume at first that a process $y(t)$ is described by a differential equation, given by:

$$dy = f(y,t)dt, y(s) = x$$

At random times τ_i, $s < \tau_1 < \tau_2 < \cdots < \tau_i < \cdots$. the process above jumps by a known (or random) quantity z_i. This means that at time τ_i, the value of the variable $y(\tau_i)$ instantly increases to $y(\tau_i^+) = y(\tau_i^-) + z_i$. Mathematically, this can be written as follows (Fig. 2):

$$y(t+dt) = \begin{cases} y(t) + f(y,t)dt \ \text{ at } \ t \in (\tau_i, \tau_{i+1}) \\ y(t) + z_i \ \text{ at } \ t = \tau_i, i = 1,2,3, \ldots\ldots.. \\ y(s) = x \end{cases}$$

where it is convenient to write $\tau_0 = s$. In such a process, the equation behaves between jumps as if it were an ordinary differential equation, while at the jump times it creates a discontinuous change in the process variable $y(t)$ which may be deterministic or stochastic. For example, if the jumping process is Poisson then the probability of a jump occurring in dt is λdt where λ is a known parameter and the inter-jump event times are exponential. That is, if $n(dt)$ denotes the number of jumps occurring in dt, then

$$\begin{cases} P(n(dt) = 1) = \lambda dt + o(dt) \\ P(n(dt) = 0) = 1 - \lambda dt + o(dt) \\ P(n(dt) \geq 2) = o(dt) \end{cases}$$

and the size of the jump equals one. As a result, the time between jumps has an exponential distribution and is memory-less since at any instant of time, the probability of a jumps are independent and independent of their previous history. The ordinary differential equation with jumps stated above would then be written as follows :

$$y(t+dt) = \begin{cases} y(t) + f(y,t)\, dy \ \text{ at } \ t \in (\tau_i, \tau_{i+1}) \\ y(t) + 1 \ \text{ at } \ t = \tau_i, i = 1,2,3, \ldots\ldots.. \\ y(s) = x \end{cases}$$

where τ_i is defined by a Poisson probability distribution with parameter λ. Next, let the size of the jump z be arbitrary, possibly defined by a probability distribution $f(z)$. For simplification purposes, we assume that the jump times τ_i and the jump magnitudes z are statistically independent leading therefore to a compound Poisson (jump) process. We then write for convenience.

$$dy = f(y,t)\,dt + \mu(z, n(dt))$$

$$\mu(z, n(dt)) = \begin{cases} 0 \text{ if } n(dt) = 0 \\ z \text{ if } n(dt) = 1 \end{cases}$$

Since $n(dt)$ is a Poisson process, we have,

$$P(dy) = \sum P\left[dz\,|n(dt)\right] P\left[n(dt)\right]$$

where $n(dt) = 1{,}0$. Thus, $P(dy) = P[f(y,t)dt][1 - \lambda dt] + P(z)\lambda dt$ where $P(.)$ denotes the probability distribution of its argument. Generally, it is convenient to write formally a stochastic differential equation with jumps as follows:

$$dy = f(y,t)dt + \sum z_i \delta(t - \tau_i);\, y(s) = x$$

where $\delta(.)$ is the Dirac Delta function, z_i is the i^{th} jump at time τ_i, both possibly random. Further, if the underlying process (without jumps) is also stochastic, or given by a Wiener process, we can proceed as before and write more generally:

$$dy = f(y,t)dt + \sigma(y,t)dw + \sum z_i \delta(t - \tau_i),\, y(s) = x$$

Applications of such processes in finance and in insurance are numerous. These include insurance problems where claims arrive following a Poisson process and claims sizes have a known distribution or events occurring contributing to a jump of the underlying continuous time process. Model where jumps occur due to a past history of past jumps are memory based jump processes. Below we consider some simple jump stochastic processes.

9.1 A forward Rate Process:

Consider at time t, the forward rates for all future maturity $T > t$ and say that it follows a difference equation where the time interval is denoted by Δt:

$$f(t + \Delta t, T) = f(t,T) + a(t,T)\Delta t + \sigma(T)\xi_1(t + \Delta t)$$
$$\sqrt{\Delta t} + N_{\lambda,\Delta t}\xi_2(t + \Delta t), \forall T > t$$

where $f(t,T)$ is the one period forward rate at time T, as observed at time t. The drift coefficient of the normal (Wiener) process is $a(t,T)$ while the diffusion coefficient is

$\sigma(T)$ which is independent of time t. Further, $\xi_1(.)$ is a normal random variable with zero mean and unit variance while $\xi_2(.)$ is a discrete random variable independent of $\xi_1(.)$ and defined as follows:

$$\xi_2 = \begin{cases} \mu + \gamma & \text{with probablity 1/2} \\ \mu - \gamma & \text{with probability 1/2} \end{cases}$$

Finally, the $N_{\lambda,\Delta t}$ represents the point process, assuming a value of 1 with probability $\lambda\Delta t$ and zero with probability $(1 - \lambda\Delta t)$. In other words, the parameter λ denotes the jump rate. As a result, the forward rate is a Wiener-Jump process where the size of jumps have an expected mean μ which occur at the exponential rate $1/\lambda$.

The Poisson process as introduced at the beginning of this paper is a counting process with no memory as the time between events is exponentially distributed, and therefore memory-less. It is defined by the probability distribution:

$$f(n) = \frac{\lambda^n e^{-\lambda n}}{n!}$$

with a mean λ equal its variance, which may be restrictive. Let $P(\tau)$ be the probability that a given interval of time between events is equal or greater than τ. Let $\theta(\tau)$ be the probability distribution of this time, then:

$$P(\tau) = 1 - \int_0^\tau \theta(u)du \quad \text{and} \quad \text{therefore} \quad \theta(\tau) = -\frac{dP(\tau)}{d\tau} \quad \text{and} \quad \theta(\tau) = \lambda e^{-\lambda\tau}$$

We shall compare this result to Laskin's (2003) result for a fractional Poisson process.

Crow and Bardwell have suggested instead that a super Poisson distribution can be constructed to account for a "Poisson-Like" distribution with a greater variance than its mean (a super-Poisson). In which case, the inter-event distribution would no longer be exponential. It is given by:

$$P(n) = f(n : \xi, \lambda) = \frac{\Gamma(\xi)\lambda^n}{{}_1F_1(1;\xi;\lambda)\Gamma(\xi + n)}, n = 0,1,2,3...$$

Where $\lambda, \xi > 0$ and $F_1(1;\xi;\lambda)$ is a confluent hyper-geometric series given by:

$${}_1F_1(1;\xi;\lambda) = 1 + \frac{\lambda}{\xi} + \frac{\lambda^2}{\xi(1+\xi)} + \frac{\lambda^3}{\xi(1+\xi)(2+\xi)} + ...$$

Whose mean and variance are:

$$E(n) = \lambda + (1 - \xi)\left(1 - \frac{1}{{}_1F_1(1;\xi;\lambda)}\right) > \lambda$$

$$\mathrm{var}\,(n) = \lambda(1 + E(n)) + E(n)(1 - E(n) - \xi) > \lambda$$

When $\xi = 1$ then $E(n) = \lambda,\ \mathrm{var}(n) = \lambda + \lambda E(n) - (E(n))^2 = \lambda$ which are the parameters of a standard Poisson distribution. If $\lambda > \xi$ then ${}_1F_1(1;\xi;\lambda) \approx 1 + \frac{\lambda}{\xi}$ and therefore,

$$E(n) = \lambda\left(1 + \frac{(1-\xi)^2}{\lambda+\xi}\right) \quad \text{while,}$$

$$\mathrm{var}(n) = \lambda + (1+\lambda-\xi)\left(1 + \frac{(1-\xi)^2}{\lambda+\xi}\right) + \left(\lambda\left(1 + \frac{(1-\xi)^2}{\lambda+\xi}\right)\right)^2$$

These provide a system of two moment equations in the two parameters of the Poisson distribution. For example, consider a period of time t and calculate the number of events a stock price has increased. Let the data indicate a mean and variance estimates, (μ, s^2), then using the two equations above, a mean-variance approximation for the super Poisson model can be guessed. Maximum Likelihood Estimated (MLE) may provide specific results.

9.2 The valuation of an option in a jump process (Merton 1976):

The valuation of an option when the price process is modeled by both a diffusion and a jump process is difficult because the application of the arbitrage argument it requires (since it is necessary to find a certainty equivalent for both the diffusion process and the jump). Explicitly, let the price process be lognormal with jumps:

$$\frac{dp}{p} = \alpha dt + \sigma dW + K dQ$$

where dQ is an adapted Poisson process with parameter $q\Delta t$. In other words, $Q(t + \Delta t) - Q(t)$ has a Poisson distribution function with mean $q\Delta t$ or for infinitesimal time intervals:

$$dQ = \begin{cases} 1 & w.p. \quad qdt \\ 0 & w.p. \quad (1-q)dt \end{cases}$$

Let $F = F(p,t)$ be the option price. When a jump occurs, the new option price is $F(p(1+K))$. As a result,

$$dF = [F(p(1+K)) - F]\, dQ$$

When no jump occurs, we have:

$$dF = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial p}dp + \frac{1}{2}\frac{\partial^2 F}{\partial p^2}(dp)^2$$

and explicitly, letting $\tau = T - t$ be the remaining time to the exercise date, we have:

$$dF = \left[-\frac{\partial F}{\partial \tau} + \alpha p\frac{\partial F}{\partial p} + \frac{1}{2}p^2\sigma^2\frac{\partial^2 F}{\partial p^2}\right]dt + p\sigma\frac{\partial F}{\partial p}dW$$

Combining these two equations, we obtain:

$$dF = adt + bdw + cdQ$$
$$a = \left[-\frac{\partial F}{\partial \tau} + \alpha p \frac{\partial F}{\partial p} + \frac{1}{2} p^2 \sigma^2 \frac{\partial^2 F}{\partial p^2} \right];\ b = p\sigma \frac{\partial F}{\partial p};\ c = F(p(1+K)) - F$$

with

$$E(dF) = [a + qc]dt \quad \text{since} \quad E(dQ) = qdt$$

To eliminate the stochastic elements (and thereby the risks implied) in this equation, we shall construct a portfolio consisting of the option and a stock. To eliminate the "Wiener risk", i.e. the effect of " dw", we let the portfolio Z consist of a future contract whose price is p for which a proportion v of stock options is sold (which will be calculated such that this risk disappears). In this case, the value of the portfolio is:

$$dZ = p\alpha dt + p\sigma dW + pKdQ - [vadt + vbdW + vcdQ]$$

If we set $v = p\sigma/b$ and insert in the equation above (as done by Black-Scholes), then we will eliminate the "Wiener risk" since:

$$dZ = p(\alpha - \sigma a/b)dt + (p\sigma - vb)dW + p(K - \sigma c/b)dQ$$

or

$$dZ = p(\alpha - \sigma a/b)dt + p(K - \sigma c/b)dQ$$

In this case, if there is no jump the evolution of the portfolio follows the differential equation:

$$dZ = p(\alpha - \sigma a/b)dt \quad \text{if there is no jump}$$

However, if there is a jump, then the portfolio evolution is:

$$dZ = p(\alpha - \sigma a/b)dt + p(K - \sigma c/b)$$

Since the jump probability equals $qdt,$ we obviously have:

$$\frac{E(dZ)}{dt} = p(\alpha - \sigma a/b) + pq(K - \sigma c/b)$$

There remain a risk in the portfolio due to the jump. To eliminate it we can construct another portfolio using an option F' (with exercise price E') and a future contract such that the terms in dQ are eliminated as well. Then, constructing a combination of the first (Z) portfolio and the second portfolio (Z'), both sources of uncertainty will be reduced. Applying an arbitrage argument (stating that there cannot be a return to a riskless portfolio which is greater than the riskless rate of return r) we obtain the proper proportions of the riskless portfolio.

Alternatively, finance theory (and in particular, application of the CAPM-Capital Asset Pricing Model) state that any risky portfolio has a rate of return in a small time interval dt which is equal the riskless rate rplus a return premium for the risk assumed which is proportional to its effect. Thus, using the CAPM we can write:

$$E\frac{dZ}{Zdt} = r + \lambda\frac{p(K - \sigma c/b)}{Z}$$

where λ is assumed to be a constant and expresses the "market price" for the risk associated to a jump. This equation can be analysed further leading to the following partial differential equation which remains to be solved (once the boundary conditions are specified):

$$-\frac{\partial F}{\partial \tau} + \left[(\lambda - q)\left\{pK\frac{\partial F}{\partial p} - F(p(1+K) - F\right\}\right] + \frac{1}{2}\frac{\partial^2 F}{\partial p^2}p^2\sigma^2 - rF = 0$$

with boundary condition:

$$F(T) = Max(0, p(T) - E)$$

Of course, for an American option, it is necessary to specify the right to exercise the option prior to its final exercise date, or

$$F(t) = Max(F^*(t), p(t) - E)$$

where $F^*(t)$ is the value of the option which is not exercised at time t and given by the solution of the equation above. The solution of this equation is of course much more difficult than the Black-Scholes partial differential equation we have seen in the previous chapter. For this reason, any complication in the terms of the option or the underlying theoretical assumptions which are made, requires that we apply numerical and simulation techniques to value the option.

9.3 The Fractional Poisson Process

An extension to a fractional (long run memory) Poisson process is a non-Markovian counting process which is considered next. Such a distribution captures the long run memory effect which results from a non-exponential interval time between events observed in a broad set of problems in finance, in insurance as well as in many "counting" problems where the time between events follows a power law. The fractional Poisson distribution providing a distribution of the number of counted events in a time interval (0,t) is was initially developed by (see Laskin 2003, 2009). The underlying process parallel to the Poisson process which solves the following system of differential-difference equations:

$$\frac{dP(n,t)}{dt} = \nu(P(n-1,t) - P(n,t)), \quad \frac{dP(0,t)}{dt} = P(0,t)$$

Laskin suggests instead, the following model:

$$\frac{d^{\alpha}P_{\alpha}(n,t)}{dt}=\nu(P_{\alpha}(n-1,t)-P_{\alpha}(n,t)),\quad \frac{dP_{\alpha}(0,t)}{dt}=-P_{\alpha}(0,t)+\frac{t^{-\alpha}}{\Gamma(1-\alpha)}$$

Where $\frac{d^{\alpha}P_{\alpha}(n,t)}{dt}$ is the operator of the time derivative of fractional order α defined as the Riemann-Liouville function, or:

$$\frac{d^{\alpha}P_{\alpha}(n,t)}{dt}=\frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}\int_0^t(t-\tau)^{\alpha}P_{\alpha}(n,\tau)d\tau, 0<\alpha<1$$

Explicitly,

$$\frac{d^{\alpha}P_{\alpha}(0,t)}{dt}=\frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}\int_0^t(t-\tau)^{\alpha}P_{\alpha}(0,\tau)d\tau=-P_{\alpha}(0,t)+\frac{t^{-\alpha}}{\Gamma(1-\alpha)}$$

The probability generating function of these equations are given by:

$$G_{\alpha}(z,t)=\sum_{n=0}^{\infty}z^nP_{\alpha}(n,t)\quad\text{and}\quad\text{its}\quad\text{inverse}\quad P_{\alpha}(n,t)=\frac{1}{n!}\frac{\partial^nG_{\alpha}(z,t)}{\partial z^n}|_{z=0}$$

They allow a definition of the PGF of $\frac{d^{\alpha}P_{\alpha}(n,t)}{dt}$ since:

$$\begin{aligned}\frac{d^{\alpha}G_{\alpha}(z,t)}{dt}&=\nu\left(\sum_{n=0}^{\infty}z^nP_{\alpha}(n-1,t)-\sum_{n=0}^{\infty}z^nP_{\alpha}(n,t)\right)\\&=\nu(z-1)G_{\alpha}(z,t)+\frac{t^{-\alpha}}{\Gamma(1-\alpha)}\end{aligned}$$

Whose solution is a Mittag-Leffler function

$$G_{\alpha}(z,t)=E_{\alpha}(\nu t^{\alpha}(z-1))\quad\text{with } E_{\alpha}(h)=\sum_{m=0}^{\infty}\frac{h^m}{\Gamma(1+\alpha m)}$$

And thus $E_{\alpha}(.)$ is a function given by:

$$E_{\alpha}(z,t)=\nu+\frac{\nu t^{\alpha}(s-1)}{\Gamma(1+\alpha)}+\frac{\nu t^{2\alpha j}(s-1)^2}{\Gamma(1+2\alpha)}+\sum_{j=3}^{\infty}\frac{\nu t^{\alpha j}(s-1)^j}{\Gamma(1+\alpha j)}$$

In this case, we have:

$$\begin{aligned}&E_{\alpha}(..)|_{s=1}=E_{\alpha}(0)=\nu\\&\frac{\partial E_{\alpha}(..)}{\partial s}|_{s=1}=\frac{\partial E_{\alpha}(..)}{\partial z}\frac{\partial z}{\partial s}\text{and } E(n,t)=\frac{\nu t^{\alpha}}{\Gamma(1+\alpha)}\\&\frac{\partial E_{\alpha}(..)}{\partial s^2}|_{s=1}=E(n^2,t)=\frac{2\nu t^{2\alpha}}{\Gamma(1+2\alpha)}\end{aligned}$$

Further, the probabilities are:

$$P_\alpha(n,t) = \frac{(\nu t^\alpha)^n}{n!} \sum_{i=0}^{\infty} \frac{(i+n)!}{i!} \frac{(-\nu t^\alpha)^i}{\Gamma(1+\alpha(i+n))}, \quad 0 < \alpha < 1$$

These probabilities as pointed out by Laskin can be represented by the Mitag-Leffler equation:

$$P_\alpha(n,t) = \frac{(-z)^n}{n!} \frac{d^n}{dz^n} E_\alpha(z)\,|_{z=-\nu t^\alpha} \text{ and } P_\alpha(0,t) = E_\alpha(-\nu t^\alpha)$$

When $\alpha = 1$, it is reduced to a Poisson process with

$$P_\alpha(n,t) = \frac{(\nu t)^n}{n!} \sum_{i=0}^{\infty} \frac{(i+n)!}{i!} \frac{(-\nu t)^i}{\Gamma(1+(i+n))} = \frac{(\nu t)^n}{n!} e^{-\nu t}$$

The first two moments are also

$$E(n_\alpha) = \frac{\nu t^\alpha}{\Gamma(1+\alpha)} \quad \text{and} \quad E(n_\alpha^2) = n_\alpha + (n_\alpha)^2 \frac{\sqrt{\pi}\,\Gamma(1+\alpha)}{2^{2\alpha-1}\Gamma(1/2+\alpha)}$$

Therefore a variance (see Wikipedia on Fractional Poisson processes):

$$\sigma_\alpha^2 = E(n_\alpha^2) - (n_\alpha)^2 = n_\alpha + (n_\alpha)^2 \left(\frac{\sqrt{\pi}\,\Gamma(1+\alpha)}{2^{2\alpha-1}\Gamma(1/2+\alpha)} - 1 \right)$$

Note that $\sigma_1^2 = n_1 + (n_1)^2 \left(\frac{\sqrt{\pi}}{2\Gamma(1/2)} - 1 \right) = \nu t + (\nu t)^2 \left(\frac{\sqrt{\pi}}{2\Gamma(1/2)} - 1 \right)$ with $\frac{\sqrt{\pi}}{2\Gamma(1/2)} - 1 = 0$

As a result a comparison with the Poisson process yields:

$$\text{Mean} : E(n_\alpha) = \frac{\nu t^\alpha}{\Gamma(1+\alpha)} \text{ versus } \frac{1}{2}\nu t \text{ at } \alpha = 1$$

$$\text{Variance} : \sigma_\alpha^2 = E(n_\alpha^2) - (n_\alpha)^2 = n_\alpha + (n_\alpha)^2 \left(\frac{\sqrt{\pi}\,\Gamma(1+\alpha)}{2^{2\alpha-1}\Gamma(1/2+\alpha)} - 1 \right)$$

$$\text{Versus } \sigma_1^2 = n_1 + (n_1)^2 \left(\frac{2\sqrt{\pi}}{\Gamma(1/2)} - 1 \right) = n_1 = \frac{1}{2}\nu t$$

Explicit expressions for the initial probabilities are thus:

$$P_\alpha(0,t) = \sum_{i=0}^{\infty} \frac{(-\nu t^\alpha)^i}{\Gamma(1+i\alpha)}, P_\alpha(1,t) = (\nu t^\alpha) \sum_{i=0}^{\infty} \frac{i(-\nu t^\alpha)^i}{\Gamma(1+\alpha(i+1))} \text{etc.}$$

Finally, the inter-event time distribution can be calculated by noting that:

$$\theta_\alpha(\tau) = -\frac{dP_\alpha(\tau)}{d\tau} \quad \text{where } P_\alpha(\tau) = 1 - \sum_{n=1}^{\infty} P_\alpha(n,\tau) = E_\alpha(-\nu t^\alpha)$$

And therefore,

$$\theta_\alpha(\tau) = \nu\tau^{\alpha-1}E_{\alpha,\alpha}(-\nu t^\alpha), \quad E_{a,b}(z) = \sum_{m=0}^{\infty}\frac{z^m}{\Gamma(am+b)}, 0 < \alpha < 1$$

where $E_{a,b}(z)$ is the two-parameter Mittag-Leffler equation.

The Fractional Compound Poisson Process, defined by $X(t) = \sum_{i=1}^{N(t)} Y_i$ where $N(t)$ is a fractional Poisson process. Laskin 2003, p. 211, in particular proves that the moment generating function of the Compound Poisson process is:

$$J_\alpha(s,t) = E_\alpha(\nu t^\alpha(g(s)-1)) = \sum_{m=0}^{\infty}\frac{(\nu t^\alpha(g(s)-1))^m}{\Gamma(1+\alpha m)}$$

$$\frac{d^\alpha}{ds}\frac{J_\alpha(s,t)}{\partial s} = \frac{d^\alpha}{ds}\sum_{m=0}^{\infty}\frac{(\nu t^\alpha)^m}{\Gamma(1+\alpha m)}(g(s)-1)^m$$

At s = 0, we have

$$\frac{\partial J_\alpha(s,t)}{\partial s}\mid_{s=0} = E(Y_i)\frac{\nu t^\alpha}{\Gamma(1+\alpha)}$$

where $g(s)$ is the moment generating function of the random variables Y_i, which indicates the independence of the events times of the fractional Poisson process and the random events Y_i.

9.4 The Inter-Event Distribution

Inter-event time distributions for the fractional Poisson process are due to Laskin as well as subsequent development by several authors (for example, Cahoy and Polito 2013). Again consider the counting fractional Poisson process over time $N^\nu(t)$, $0 < \nu < 1$. Let T^ν be the tail distribution of the time between events, or $P^\nu(T^\nu > t) = E_\nu(-\lambda t^\nu)$ with $E_\nu(z) = \sum_{n=0}^{\infty}\frac{z^n}{\Gamma(\nu n+1)}$ the Mittag-Leffler equation. The financial importance of this Poisson process is that it is a counting process whose time between events are not independent and therefore provides a far greater modeling flexibility for insurance, credit risks and other events than the Poisson process where inter-event times are exponentially (and thus independent and memory-less) distributed. The inter-event time distribution fractional Poisson Process is then:

$$f^\nu(t) = \lambda t^\nu E_{\nu,\nu}(-\lambda t^\nu),\ t > 0$$

Where

$$E_{\nu,\kappa}(-\lambda t^{\nu}) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\nu n + \kappa)}$$

is the two-parameters Mittag-Leffler function. Further, the kth fractional moment of the random inter-event time is:

$$E^{\nu}(T^{\nu})^k = \frac{\Gamma(1+k)}{\Gamma(k/\nu)\Gamma(1-k))} \frac{\pi}{\lambda^k \sin(\pi k/\nu)},\ 0 < k < \nu$$

Thus, the probability density function of mth event is:

$$f_m^{\nu}(t) = \lambda^m \frac{t^{\nu m} - 1}{(m-1)!} E_{\nu,\nu}^{(m-1)}(-\lambda t^{\nu}),\ t > 0$$

Where $E_{\nu,\nu}^{(m-1)}$ denotes an $\mathrm{m}-1$ derivative of the two parameter Mittag-Leffler equation. Its Laplace Transform is:

$$L^{*}(f_m^{\nu}(t)) = \int_0^{\infty} e^{-st} f_m^{\nu}(t)dt = \frac{\lambda^m}{(\lambda + s^{\nu})^m}$$

Thus, when $\nu = 1$, this is reduced to the Laplace Transform of a Gamma Probability distribution and of course when $\nu = 1$ and $m = 1$ we obtain the Laplace Transform of the exponential probability distribution. A comparison of the essential properties of the Poisson and the Fractional Poisson are given below (taken from Cahoy and Polito 2013).

	Poisson $\nu = 1$	FP $\nu < 1$
$P^{\nu}(T^{\nu} > t)$	$e^{-\lambda t}$	$E_{\nu,\nu}(-\lambda t^{\nu})$
$f^{\nu}(t)$	$\lambda e^{-\lambda t}$	$\lambda t^{\nu} E_{\nu,\nu}(-\lambda t^{\nu})$
$p_k^{\nu}(t)$	$\frac{(\lambda t)^k}{k!} e^{-\lambda t}$	$\frac{(\lambda t^{\nu})^k}{k!} \sum_{i=0}^{\infty} \frac{(i+k)!(-\lambda t^{\nu})^i}{\Gamma(\nu(i+k)+1)}$
Mean	λt	$\lambda t^{\nu}/\Gamma(\nu+1)$
variance	λt	$\frac{\lambda t^{\nu}}{\Gamma(\nu+1)} + (\lambda t^{\nu})^2 \left[\frac{1}{\nu\Gamma(2\nu)} - \frac{1}{\Gamma^2(\nu+1)}\right]$
Laplace Transform	$\left(\frac{\lambda}{\lambda+s}\right)^m$	$\left(\frac{\lambda}{\lambda+s^{\nu}}\right)^m$

Such a counting distribution may then be used to calculate the number of credit defaults over time, the number of events and so on where the time between default events are not necessarily exponential. For further study, see Leonenko and Merzbach 2011; Beghin and Orsingher 2009, 2010; Herbin and Merzbach 2006; Ivanov and Merzbach 2000; Leonenko et al. 2013a, b, c; Wang et al. 2003, 2006, 2007 and Wen 2003, 2006, 2007; Cahoy et al. 2010.

9.5 *The Exponential Model: The Hawkes Process*

Consider next the following (Hawkes) mean reverting process:

$$dx(t) = -\lambda(x(t) - \mu(t)) + \xi dN_t,\ x(0) > 0$$

Where N_t is a counting process. A solution for this equation can be reached as follows. First set $y(t) = e^{\lambda t}x(t)$ and therefore, $dy(t) = \lambda e^{\lambda t}x(t) + e^{\lambda t}dx(t)$ or $e^{-\lambda t}dy(t) - \lambda e^{-\lambda t}y(t)dt = dx(t)$. As a result,

$$e^{-\lambda t}dy(t) = dz(t) = (\lambda\mu(t))dt + \xi dN_t,\ y(0) = x(0)$$

And therefore for μ constants, we have

$$z(t) = \lambda\mu t + \int_0^t \xi dN_t \text{ or } y(t) = \lambda\mu t e^{\lambda t} + e^{\lambda t}\xi \int_0^t dN_t \text{ and finally,}$$

$$x(t) = \lambda\mu t + \xi N_t$$

The interesting question however is to let the counting process be a fractional Poisson process.

10 Short Memory

Short term memory, unlike autoregressive and model based memory is based on a "model selection" that depends on the process data that has occurred in the immediate or in the current past. For example, is the last observed price, contributing to the propensity for the price to increase or decrease next. Is a series of specific outcomes altering the underlying process. Such a series may be defined by a series of events or by a statistical realization of a series of events. It is therefore data based and defined by identifiable and occurring (or recurring) events. In actuarial science for example, much use is made Poisson jump processes with specific events (such as a claim by an insured) used to increase the insured premium). Similarly in finance Brownian motion is used to capture "the noise" in an otherwise well predicted process while specific events may alter future expectations and thereby the propensities of financial agents to buy or sell assets. These models assume for mathematical convenience that events are independent. Short memory is thus a departure from these models. Weiss 1994, 2002 has pointed out that certain random phenomena have in some manner, a "biased" randomness (see also Balinth 1986; Masoliver et al. 1993, 2003, 2007; Ben Avraham and Havlin 2000; Pottier 1996; Claes and van den Broeck 1987; Cresson et al. 2007; Weiss and Rubin 1983; Wu et al. 2000; Viswanathan et al. 1999, Patlak 1953 and Taylor 1921).

Such a bias can induce a persistent "noise" and therefore alters the underlying current stochastic process. For example, is a bank failure increasing the probability of another failure; Is an accident by an insured alters the probability of a subsequent accident probability? In some ways, processes do change by learning, by adapting etc. and thereby their underlying process may change accordingly. These are however different manifestations of short memory. The short memory process differs in several manners from the memory presumed in several papers and other applications areas where memory is deemed important. For example, Telesca and Lovallo (2006) while studying the advent of terrorist attacks suggest that “memory” is associated to correlated events—in the sense that inter-events times are not statistically independent. Such an observation implies explicitly that terror events do not have a Poisson distribution. Ferguson and Bazant (2005) while providing a number of avenues to model polymers in solution characterized by long flexible chains as Random walk models of various complexities, suggest also the potential use of a memory imbedded in a correlation between subsequent steps in a random walk. In this case, “persistence” (in their sense) still exhibits a “normal diffusion”. In the longer run, fractal analysis of random walks, defined by a Hurst exponent, seeks as indicated earlier, to capture a long run memory, based as well on some “correlation” within the walk that lead to “sub or superdiffusion” processes. Such approaches assume usually that short term memory is well accounted for by Markov models (Cresson et al. 2007) while long-range memory typically gives rise to Non-Markovian walks (Hurst 1951; Mandelbrot 1982; Bunde and Havlin 1991; Huillet 2002; Tapiero and Vallois 1996; Peter 1995). This is in contradiction to our analysis of a short term memory model, based on a concept of memory of the previous random event and not only the memory of the previous random state. These “memory processes” remember only the past state (or several ones) and not their random properties. For example, “average persistence of random walks” studied by Rieger and Igloi (1999) is essentially based on a birth-death (B-D) random walk (their equation 1) which assumes adsorbing boundaries applied to the B-D random walk and which leads to a “persistence” defined by the probability that an expected walker’s position has not returned to its initial position. Persistence of past direction in a random walk or internal bias, was indicated in an early paper by Clifford Patlak (1953) who meant that the process will travel in a given direction need not be the same for all directions but depends solely on the particle’s previous direction of motion. An external bias then arises from an anisotropy of the external force on the particle. Application of such a model is used to study diffusion and long chain polymers.

In financial time series, (such as the S&P, the Euro-Dollar exchange rate), we found evidence of short term memory. For example, over a period of 759 consecutive days, we found that in a standardized process, the probability of a price increase on the S&P will follow a price increase was 0.4774 while the probability of such an increase following an actual price decline in the previous day was 0.50964. By the same token, the probability of a price decline following a price increase was 0.5254 while a decline following a decline was found to have a probability of 0.4820. Analysis of other time series (such as on intraday data) may lead to more pronounced results. Over shorter periods of time (51 days), these probabilities were found to be even more pronounced (0.5925 and 0.4583) and (0.4074 and 0.54166). These observations lead to some financial traders to devise trading strategies based on persistence and profit from financial markets' incompleteness. Increasingly, individuation of insurance contracts and individual health histories and their time record are both useful and essential to measure the propensity of an insured to claim or an individual to recede. Further, initially healthy individuals may eventually become ill, altering the probabilities of their subsequent health events. In predicting the weather, counting natural events (such as the number of Hurricanes over a period of time), their occurrence might not be, necessarily, statistically independent as presumed by a Poisson counting processes (a fractional Poisson process may however be used). For example, to what extent do years of intense Hurricane activity follow each other, or vice versa, what would be the propensity of an active Hurricane period to follow a period of non-activity? In many theoretical models, we presume that these events are statistically independent, meaning that there is no statistical memory, justifying thereby counting processes with linear time growth. Similarly in financial time series, we use a random walk or a Brownian motion driven stochastic process (although empirical data regarding financial time series of stocks, exchange rates, interest rates etc. indicate otherwise).

Finally, "learning" with a short memory model differs from other approaches. For example, Actuarial approaches mostly assume that prior events do not alter the basic probabilities laws that determine the occurrence of subsequent events. To update probability estimates, a statistical credibility theory is used to evaluate the objectivity and the subjectivity of a risk source and devise a statistical "learning" mechanism that allows the updating of the underlying future event probability. Using Bayesian statistics, credibility theory divides risk events into a number classes each with a propensity for the event to occur which are updated using subjective prior estimates of risk classes and an accrued experience—the process history observed. The goal of credibility theory is then to set up an experience rating system. Unlike credibility theory, this "persistent" approach presumes that there may be an inherent persistence in an underlying process that will dictate the probability laws (and therefore the randomness) with which subsequent events occur. The probability of a subsequent event will then be determined by the past memory (in our case, specific past events) rather than be determined by a statistical estimator based on the accrued evidence of past events. Explicitly, while credibility theory seeks to integrate "experience" in estimating the propensity of an event to occur, a persistence approach, an inherent property of the underlying process determines, conditionally on the "observed memory", the actual probabilities with which events may occur or not.

In the following sections, we shall consider some essential and mathematical results. To simplify our presentation, we shall consider first a one period memory model in a discrete time random walk and summarize for brevity basic results (where proofs are given in previous papers published by the authors as well as other authors). The "simple", one period memory, will demonstrate analytically the short term and the long term effects of memory. In particular, we will show that the variance grows initially at a nonlinear rate and subsequently, becoming linear. The linear trend however would have endogenized the short term memory effects. To confirm our analysis, we also obtain as special cases well known ergodic results as well as the standard random walk processes. The effects of memory have many implications, and applications, some of which will be discussed in this paper In particular, we point out fundamental differences between short term stochastic models based current "events" and long term stochastic models based on theoretical constructs justified by extensive and large data sets. These differences are naturally important for financial trading when the short term is of greater concern.

10.1 The Random Counting Process with a 1-Short memory

Assume that an event in any given period of time can be in one of two states: (0,1) (or $+1$ or -1). The first state "0" states that the event has not occurred within the period while "1" states that the event has occurred in the period. For example, a loan in a bank portfolio loan has defaulted or not, an insured has claimed following an accident or not, etc. Data can be gathered to determine the probability of an event occurring conditional on its past realization which we denote by its probability $\alpha > 0$. By the same token, if the event has not occurred in the period, then the probability of the event occurring in the following year would be $\beta > 0$. These observations define a simple two-states Markov chain, given by:

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

Thus, if we let y_t be an event occurring at time t, where $0 < \alpha, \beta < 1$ we have four potential events:

$$\begin{cases} (y_t = 1 \,|y_{t-1} = 0\,) & \text{and } P(y_t = 1\,|y_{t-1} = 0\,) = P(\varepsilon_t(\alpha) = 1) = \alpha \\ (y_t = 0 \,|y_{t-1} = 0\,) & \text{and } P(y_t = 0\,|y_{t-1} = 0\,) = P(\varepsilon_t(\alpha) = 1) = 1-\alpha \\ (y_t = 1 \,|y_{t-1} = 1\,) & \text{and } P(y_t = 1\,|y_{t-1} = 1\,) = P(\varepsilon_t^{'}(\beta) = 1) = 1-\beta \\ (y_t = 0 \,|y_{t-1} = 1\,) & \text{and } P(y_t = 0\,|y_{t-1} = 1\,) = P(\varepsilon_t^{'}(\beta) = 0) = \beta \end{cases}$$

In other words,

$$\text{If } y_{t-1} = 0 \text{ then the probability law of } y_t \text{ is identical to that of } \varepsilon_t(\alpha)$$

$$If\, y_{t-1} = 1 \text{ then the probability law of } y_t \text{ is identical to that of } \varepsilon_t^{'}(1-\beta)$$

For example, $(y_t = 1\,|y_{t-1} = 0\,)$ may define the event that a stock price has increased at time t given that it did not increase previously. We also define the parameter $\rho = 1-\alpha-\beta, \quad \rho \in [0,1]$. For example, say that $\alpha = 0.3, \quad \beta = 0.6$, this means that is a stock price has not increased, its probability of increasing next is 0.3, while it has a probability 0.4 if it has increased previously. In other words, there is in this case a momentum and $\rho = 0.1$.

Over a period of time t, the total number of events (claims, or an increasing or decreasing stock price etc.), is then:

$$x_t = \sum_{j=0}^{t} y_j$$

These movements are conditional on the initial event (the current memory of the event), denoted by: $y_0 = 0$ or $y_0 = 1$. A recursive single memory model is then:

$$y_t = (1 - y_{t-1})\tilde{\varepsilon}_t(\alpha) + y_{t-1}\tilde{\varepsilon}_t^{'}(1-\beta) \text{ or } y_t = \tilde{\varepsilon}_t(\alpha) - y_{t-1}(\tilde{\varepsilon}_t(\alpha) + \tilde{\varepsilon}_t^{'}(1-\beta))$$

Note that y_{t-1} is an observed value, know at time t while $y_t = 1{,}0$ is observed only at time $t+1$. Consider a series of observed values:

$$y_{t-i} = (1 - y_{t-(i+1)})\tilde{\varepsilon}_{t-i}(\alpha) + y_{t-(i+1)}\tilde{\varepsilon}_{t-i}^{'}(1-\beta), \quad i = 1, \ldots .t-1$$

And assume that $\tilde{\varepsilon}_{t-i}(\alpha)$ and $\tilde{\varepsilon}_{t-i}^{'}(1-\beta)$ are time independent. Then system of n equations above provide an estimate for $\tilde{\varepsilon}(\alpha)$ and $\tilde{\varepsilon}(1-\beta)$. First not that for two periods,

$$y_{t-i} = \tilde{\varepsilon}(\alpha) - y_{t-(i-1)}(\tilde{\varepsilon}(\alpha) - \tilde{\varepsilon}^{'}(1-\beta))$$

And therefore a least square estimate yields

$$Min\ E \sum_{i=1}^{t} \left\{ y_{t-i} - \tilde{\varepsilon}(\alpha) + y_{t-(i-1)}(\tilde{\varepsilon}(\alpha) - \tilde{\varepsilon}^{'}(1-\beta)) \right\}^2$$

Which we rewrite as follows, with a simplified notation $\tilde{\varepsilon}(\alpha) \equiv \tilde{\alpha}$ and $\tilde{\varepsilon}^{'}(1-\beta) = 1-\tilde{\beta}$:

$$\begin{aligned} Min \sum_{i=1}^{t} & \left\{ (y_{t-i})^2 - 2y_{t-i}E(\tilde{\alpha}) + E(\tilde{\alpha})^2 \right\} \\ & + (y_{t-(i+1)})^2 \left\{ E(\tilde{\alpha})^2 + E(1-\tilde{\beta})^2 - 2E(\tilde{\alpha}(1-\tilde{\beta})) \right\} \\ & - 2y_{t-(i+1)} \left\{ E(\tilde{\alpha})^2 - E(\tilde{\alpha}(1-\tilde{\beta})) \right\} + 2(y_{t-i})y_{t-(i+1)}(\tilde{\alpha} - (1-\tilde{\beta})) \end{aligned}$$

Elementary manipulation yields the following:

$$\underset{\alpha,\beta}{Min} E \sum_{i=1}^{t} \left\{ \begin{array}{l} (y_{t-i})^2 - 2(y_{t-i})\left\{1 - y_{t-(i+1)}\right\}\tilde{\alpha} + \left\{(1-y_{t-(i+1)})^2\right\}\tilde{\alpha}^2 \\ - \left\{2(y_{t-i})y_{t-(i+1)}\right\}(1-\tilde{\beta}) + (y_{t-(i+1)})^2\left\{(1-\tilde{\beta})^2\right\} \\ +2y_{t-(i+1)}\left\{1-y_{t-(i+1)}\right\}\tilde{\alpha}(1-\tilde{\beta}) \end{array} \right\}$$

And in expectation (assuming that $\tilde{\alpha}, \tilde{\beta}$ are not correlated):

$$\underset{\alpha,\beta}{Min} E \sum_{i=1}^{t} \left\{ \begin{array}{l} (y_{t-i})^2 - 2(y_{t-i})\left\{1 - y_{t-(i+1)}\right\}\alpha + \left\{(1-y_{t-(i+1)})^2\right\}\alpha^2 \\ - \left\{2(y_{t-i})y_{t-(i+1)}\right\}(1-\beta) + (y_{t-(i+1)})^2\left\{(1-\beta)^2\right\} \\ +2y_{t-(i+1)}\left\{1-y_{t-(i+1)}\right\}\alpha(1-\beta) \end{array} \right\}$$

And therefore, least square estimates are defined by the solution of the following system of linear equations:

$$\begin{bmatrix} \sum_{i=1}^{t} \left(1 - y_{t-(i+1)}\right)^2 & -\sum_{i=1}^{t} (y_{t-(i+1)}(1 - y_{t-(i+1)})) \\ \sum_{i=1}^{t} (y_{t-(i+1)}(1 - y_{t-(i+1)})) & \sum_{i=1} \left(y_{t-(i+1)}\right)^2 \end{bmatrix}$$

$$\begin{bmatrix} \alpha \\ 1-\beta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{t} (y_{t-i}(1-y_{t-(i+1)})) \\ \sum_{i=1} ((y_{t-i})(y_{t-(i+1)})) \end{bmatrix}$$

or,

$$\begin{bmatrix} \hat{\alpha} \\ 1-\hat{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{t} (1-y_{t-(i+1)})^2 & -\sum_{i=1}^{t} (y_{t-(i+1)}(1-y_{t-(i+1)})) \\ \sum_{i=1}^{t} (y_{t-(i+1)}(1-y_{t-(i+1)})) & \sum_{i=1} (y_{t-(i+1)})^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{t} (y_{t-i}(1-y_{t-(i+1)})) \\ \sum_{i=1} ((y_{t-i})(y_{t-(i+1)})) \end{bmatrix}$$

For example,

$$\hat{\alpha} = \frac{\sum y_{t-(i+1)} \sum_{i=1} \left(y_{t-i}(y_{t-(i+1)})\right)}{\left\{ \sum_{i=1} (y_{t-(i+1)})^2 \sum_{i=1} (1-y_{t-(i+1)})^2 + \left(\sum_{i=1} \left(y_{t-(i+1)}(1-y_{t-(i+1)})\right) \right)^2 \right\}}$$

And therefore,

$$E(y_{t-i}) = \hat{\alpha} - y_{t-(i-1)}(\hat{\alpha} - (1-\hat{\beta})) = \hat{\alpha} + \hat{\rho}_t y_{t-(i-1)} \text{ with } \hat{\rho}_t = 1 - \hat{\alpha} - \hat{\beta}$$

where $\hat{\rho}_t$ is an estimate at time t (since it is based on the estimation of past observations).

Recursive estimates provide also a relationship based on the expectation of past outcomes. Thus, at time t,

$$E(y_t) = \alpha(1 - E(y_{t-1})) + (1-\beta)E(y_{t-1}) = \alpha + (1-\alpha-\beta)E(y_{t-1})$$

By recursion, we have future events recurring as a function of (α, β) and the longer memory of event y_0:

$$E(y_t) = \alpha \left[\frac{1-(1-\alpha-\beta)^t}{\alpha+\beta} \right] + (1-\alpha-\beta)^{t-1} E(y_0)$$

In other words, given only y_0, the initial condition, the expected value of the event at time t is a nonlinear function of time and the probabilities (α, β). If the underlying process is a pure random walk, $\alpha + \beta = 1$, $E(y_t) = \alpha + E(y_0)$ (since each occurrence is independent). By the same token, the expected autocorrelation $E(y_t y_{t-1})$ (and therefore events are memory dependent) is:

$$y_t = \tilde{\varepsilon}_t(\alpha) - y_{t-1}(\tilde{\varepsilon}_t(\alpha) + \tilde{\varepsilon}_t^{'}(1-\beta))$$

$$
\begin{aligned}
E(y_t y_{t-1}) = & (\tilde{\varepsilon}_t(\alpha) - y_{t-1}(\tilde{\varepsilon}_t(\alpha) + \tilde{\varepsilon}_t^{'}(1-\beta))) \\
& (\tilde{\varepsilon}_{t-1}(\alpha) - y_{t-2}(\tilde{\varepsilon}_{t-1}(\alpha) + \tilde{\varepsilon}_{t-1}^{'}(1-\beta))) \\
= & (\tilde{\varepsilon}_t(\alpha)\tilde{\varepsilon}_{t-1}(\alpha) - y_{t-1}(\tilde{\varepsilon}_t(\alpha)\tilde{\varepsilon}_{t-1}(\alpha) + \tilde{\varepsilon}_t^{'}(1-\beta)\tilde{\varepsilon}_{t-1}(\alpha))) \\
& - (y_{t-2}\tilde{\varepsilon}_t(\alpha)(\tilde{\varepsilon}_{t-1}(\alpha) + \tilde{\varepsilon}_{t-1}^{'}(1-\beta))) - y_{t-2}y_{t-1}(\tilde{\varepsilon}_t(\alpha)(\tilde{\varepsilon}_{t-1}(\alpha)) \\
& + 2\tilde{\varepsilon}_t^{'}(1-\beta)(\tilde{\varepsilon}_{t-1}(\alpha)) + \tilde{\varepsilon}_t^{'}(1-\beta)\tilde{\varepsilon}_{t-1}^{'}(1-\beta))
\end{aligned}
$$

Assuming independence of α, β and their stationary estimates $\hat{\alpha}, \hat{\beta}$, we have:

$$
E(y_t y_{t-1}) = (\hat{\alpha}^2 - (y_{t-1} + y_{t-2})\hat{\alpha}(1 + \hat{\alpha} - \hat{\beta}) + y_{t-2}y_{t-1}(1 + \hat{\alpha} - \hat{\beta})^2
$$

Thus,

$$
\begin{aligned}
& E(y_t y_{t-1} \,|y_{t-1} = 0, y_{t-2} = 0\,) = \hat{\alpha}^2, E(y_t y_{t-1} \,|y_{t-1} = 0, y_{t-2} = 1\,) = -\hat{\alpha}(1 - \hat{\beta}) \\
& E(y_t y_{t-1} \,|y_{t-1} = 1, y_{t-2} = 0\,) = -\hat{\alpha}(1 - \hat{\beta}), E(y_t y_{t-1} \,|y_{t-1} = 1, y_{t-2} = 1\,) \\
& \qquad = (1 - \beta)^2
\end{aligned}
$$

As a result, at any one time t, the co-variation of an event depends on the previous two observed values.

For t periods, we define the process $\{X_t, t \in T\}$ by $X_t = \sum_{i=1}^{t} y_i$. In this case, the following results are obtained with their proof in Vallois and Tapiero 2007.

The properties of such a short memory process can be found easily as follows. Let $\rho = 1 - \alpha - \beta$, then in $(0, t)$:

$$
\begin{cases}
E(x_t \middle| y_0 = 0) = \dfrac{\alpha}{\alpha + \beta}\left(t - \dfrac{(1 - \rho^{t+1})}{\alpha + \beta}\right) \\
E(x_t \middle| y_1 = 1) = \dfrac{1}{\alpha + \beta}\left(\alpha(t + 1) + \beta\dfrac{(1 - \rho^{t+1})}{\alpha + \beta}\right)
\end{cases}
$$

With their estimate calculated based on the observations of past events up to time t−1. The second moment (and therefore a variance calculated based on the second and the expectations above) is:

$$
E\left\{x^2(t)\middle|x(0) = 0\right\} = \alpha \begin{Bmatrix} \frac{t(t+1)}{6}[-2\beta t^2 + (3\beta - \alpha + 2)t + \alpha - \beta + 1] \\ +\sigma_3[-2\beta(\alpha + \beta)t + \beta^2 - \alpha^2 - 4\beta + 2\alpha] \end{Bmatrix} \quad \text{and}
$$

$$
\begin{aligned}
E\left\{x^2(t)\middle|x(0) = 1\right\} = & \frac{(t + 1)}{6}\begin{Bmatrix} 6 + t[-\beta^2 - 3\beta\alpha - 5\beta + 6] + \\ t^2[-\beta^2 + 3\beta\alpha - 4\beta] + 2\beta^2 t^3 \end{Bmatrix} \\
& + \sigma_3\beta[2\beta(\alpha + \beta)t + \beta^2 + 3\alpha^2 + 4\alpha\beta + 2\beta - 4\alpha]
\end{aligned}
$$

where $\sigma_3 = \sigma_3(\alpha+\beta, t)$ with $\sigma_3 = \sigma_3(s, t)$ a real valued term defined by the following relation: $1-(1-s)^{t+1} = s(t+1) - t(t+1)\frac{s^2}{2} + \frac{t(t^2-1)}{6}s^3 + s^4\sigma_3(s,t)$.

These results indicate non-linear mean and variance growth rates. This nonlinear effect is indeed very small if ρ is small and disappears gradually over time to be time linear with $\underset{t\to l\arg e}{E(x_t \,|y_0 = 0\,)} = \frac{\alpha}{\alpha+\beta}[t]$. When $\alpha + \beta = 1$, such a process corresponds to a standard random walk, at $\rho = 0$. And $E(x_t \,|y_0 = 0\,) = \alpha t/(\alpha+\beta)$ as it is expected. Note that if both the probabilities of an event occurring whether or not the event has occurred previously are very small, then $\rho = (1-\beta) - \alpha$ should be very small as well, and therefore $\rho \approx 0$ and as a result again we obtain the familiar random walk. Similar results are obtained when we consider second order moments. To this effect we calculated the variance (see Vallois and Tapiero 2007).

These results indicate a number of potential approximations to the underlying probability distributions of short memory processes. Explicitly, if the parameter α is not too small but $\alpha + \beta = 1$, then the Binomial random walk (and its normal approximation) provide a “good” approximation to the short memory random walk process. When α is sufficiently small such that $\alpha >> \alpha^2$ holds, then a Poisson moments approximations to the short memory counting process is also appropriate. Of course, for other parameters, both the binomial and the Poisson distributions may turn out to be poor approximations (albeit over the short rather than in the long run). In this sense, an observation of the US/Euro dollar exchange time series indicated (as seen above), that there is a volatility effect due to a short range memory effect. The implications of such an observation are numerous. For example, an option price (due to memory effects) may be relatively larger (or smaller) the smaller the option remaining time to its exercise (since the memory effect would “kick in”. By the same token, long range options (such as perpetual American options) are more justified in their use of an underlying price process whose stochastic driving element is the Brownian motion (or a linear growth rate for the underlying process variance). Such a short time memory may also explain the increased variance on short time (intra-day data time series). Over the longer run, these effects are dissipated. Again, for simplicity, we summarize long run (ergodic) results. The asymptotic mean and variance of the persistent-one period memory random walk is given by:

$$\begin{cases} E\left(\frac{x_t}{t}\middle| x_0 = 0\right) = \frac{\alpha}{\alpha+\beta} + \frac{1}{t}\left(\frac{1-\alpha-\beta}{(\alpha+\beta)^2}\right) + \frac{\alpha}{(\alpha+\beta)^2}\frac{\rho^{t+1}}{t} \\ E\left(\frac{x_t}{t}\middle| x_0 = 1\right) = \frac{\alpha}{\alpha+\beta} + \frac{1}{t}\left(\frac{\alpha^2+\alpha\beta+\beta}{(\alpha+\beta)^2}\right) - \frac{\beta}{(\alpha+\beta)^2}\frac{\rho^{t+1}}{t} \end{cases}$$

And by

$$\begin{cases} \frac{1}{t}Var(x_t \big| x_0 = 0) = \frac{\alpha}{(\alpha+\beta)^3}\{4\alpha^2 - \beta^2 + 3\alpha\beta - 4\alpha + 2\beta\} \\ \qquad + \frac{1}{t}\frac{\alpha(1-\alpha-\beta)(\beta^2+\alpha\beta+\alpha-4\beta)}{(\alpha+\beta)^4} + 0(1) \\ \frac{1}{t}Var(x_t \big| x_1 = 1) = \frac{\alpha\beta(2-\alpha-\beta)}{(\alpha+\beta)^3}t - \\ \qquad - \frac{1}{t}\frac{\beta(\alpha^3+\alpha\beta^2+2\alpha^2\beta-5\alpha^2-4\alpha\beta+\beta^2+4\alpha)}{(\alpha+\beta)^4} + 0(1) \end{cases}$$

These results are of course consistent with the random walk, while at the same time they indicate that memory effects remain and alter the structure of random walks. Of course, over the long run, the underlying process variance has a linear variance growth. Further, when $\alpha + \beta = 1$, we have also:

$$E\left(\frac{x_t}{t}\,|x_0 = 0\right) = \alpha, E\left(\frac{x_t}{t}\,|x_0 = 1\right) = \alpha + \frac{1}{t}$$

While the variance is proportional to time t as it is the case for both random walks and Brownian motion. If this is not the case, we obtain a constant variance correction due only to the short term memory effect. While over time, such an effect may be negligible, over shorter periods of time, such an effect can be important and to be reckoned with.

10.2 Moments and Distributions of the Short Memory Process

Other moments as well as explicit distributions can be obtained by calculating the underlying probability generating function (PGF) defined by $G(\lambda,t) = E\{\lambda^{x_t}\}$. Again, it is given by Vallois and Tapiero (2007) as follows. Let $\rho = 1-\alpha-\beta, \rho \in]-1,1[$ then the PGF of the short memory process is:

$$G(\lambda,t) = \frac{G(\lambda,1)}{\sqrt{\delta}}\Psi_t - \lambda\rho\frac{G(\lambda,0)}{\sqrt{\delta}}\Psi_{t-1}, t \geq 1$$

With boundary conditions

$$G(\lambda,0) = P(x_0 = 0) + \lambda P(x_0 = 1)$$

$$G(\lambda,1) = (1-\alpha+\lambda\alpha)P(x_0 = 0) + \lambda(1+(\lambda-1)(\alpha+\rho))P(x_0 = 1)$$

and parameters defined by $a = 1-\alpha+\lambda\alpha, \delta = (a+\lambda\rho)^2 - 4\lambda\rho, \Psi_t = \mu_0^t - \mu_1^t$ where:

$$\mu_0 = \frac{1}{2}(a+\lambda\rho+\sqrt{\delta}), \mu_1 = \frac{1}{2}(a+\lambda\rho-\sqrt{\delta})$$

where $P(x_0 = 0)$ is the probability that the initial event (memory) is null (for example, no default occurred, no accident occurred, etc.) while $P(x_0 = 1)$ denotes the probability that an event has occurred (e.g. a loan has defaulted).

Such a generating function provides the means to calculate the moments if the number of defaulting loans in a given time interval for an underlying persistent default process as well as their number of defaults probabilities in the given time interval. Further, it clearly points out to the effects of the index $\rho = 0$ on the default process. When $\rho = 0$ then $\beta = 1 - \alpha$ and whatever the previous outcome (whether a default has occurred or not), the subsequent probability of default is α while that of no default is $1 - \alpha$. When the persistence index is positive, $\rho > 0$ then $\beta = 1 - \alpha - \rho$. That is, if a loan defaults at some time, then the probability that a loan default subsequently is has a smaller probability. And vice versa, when the persistence index is negative $\rho < 0$ and the underlying default stochastic process would point out to a "contagious" default process (for example, with a portfolio's loans defaulting at an increasing rate). Inversely, for $\rho > 0$ it will indicate that the portfolio of loans has a built-in "incentive effect", reducing a default probability in a given period following a default made in the previous one.

Our result allows an explicit calculation of the moments of the persistent counting process, defined recursively as shown below. Note that:

$$\frac{\partial G}{\partial \lambda}(\lambda, t) = E(x_t \lambda^{x_t - 1}), \quad \frac{\partial^2 G}{\partial \lambda^2}(\lambda, t) = E(x_t(x_t - 1)\lambda^{x_t - 2}),$$
$$\frac{\partial^3 G}{\partial \lambda^3}(\lambda, t) = E(x_t(x_t - 1)(x_t - 2)\lambda^{x_t - 3})$$

and so on for higher order terms. Using these terms and setting $\lambda = 1$, we obtain the necessary equations which allow the calculation of the mean, the variance, the kurtosis and other moments of the short memory process distribution. It is also useful to derive a number of special and well known cases to confirm the validity of our results. First note that when $\lambda = 1$ then $G(1, t) = 1$ as expected. Further, when there is no persistence (i.e. $\rho = 0$), we have $G(\lambda, t) = \frac{G(\lambda,1)}{a}\Psi_t$, $\mu_0 = a$, $\mu_1 = 0$ and $\Psi_t = a^t$ and therefore:

$$G(\lambda, t) = \{P(x_0 = 0) + \lambda P(x_0 = 1)\}\, a^t$$

In particular, if initially, $P(x_0 = 0) = 1$, then $G(\lambda, t) = a^t$ which corresponds as expected to the Probability Generating Function of a binomial distribution. However, if $P(x_0 = 1) = 1$, then $G(\lambda, t) = \lambda a^t$. A convenient recursive expression for the generating function can be found by noting that $\Psi_t = {\mu_0}^t - {\mu_1}^t$ where μ_0 and μ_1 solve $\mu^2 - \mu(\lambda(\rho + \alpha) + 1 - \alpha) + \lambda\rho = 0$ and verify the second order equation:

$$\Psi_{t+2} - (a + \lambda\rho)\Psi_{t+1} + \lambda\rho\Psi_t = 0$$

As a result, the probability generating function $G(\lambda, t)$ satisfies as well the second order recursive equation given by:

$$G(\lambda, t + 2) = (1 - \alpha + \lambda(\alpha + \rho))G(\lambda, t + 1) - \lambda\rho G(\lambda, t)$$

Deriving this equation with respect to λwith $\lambda = 1$ we obtain a recursive expression for the moments of the process. Concentrating our attention on the first moments only, derivatives yield the following recursive equation:

$$G(\lambda, t+2) = (1-\alpha+\lambda(\alpha+\rho))G(\lambda, t+1) - \lambda\rho G(\lambda, t)$$

$$\frac{\partial^k G(\lambda, t+2)}{\partial\lambda^k} = (1-\alpha+\lambda(\alpha+\rho))\frac{\partial^k G(\lambda, t+1)}{\partial\lambda^k} - \lambda\rho\frac{\partial^k G(\lambda, t)}{\partial\lambda^k} + k(\alpha+\rho)\frac{\partial^{k-1} G(\lambda, t+1)}{\partial\lambda^{k-1}} - k\rho\frac{\partial^{k-1} G(\lambda, t)}{\partial\lambda^{k-1}}, k = 1,2,3,....$$

With initial conditions:

$$G(\lambda, 0) = P(x_0 = 0) + \lambda P(x_0 = 1)$$

$$G(\lambda, 1) = (1-\alpha+\lambda\alpha)P(x_0 = 0) + \lambda(1+(\lambda-1)(\alpha+\rho))P(x_0 = 1)$$

$$\frac{\partial G(\lambda, 0)}{\partial\lambda} = P(x_0 = 1),$$

$$\frac{\partial G(\lambda, 1)}{\partial\lambda} = \alpha P(x_0 = 0) + (1+(2\lambda-1)(\alpha+\rho))P(x_0 = 1)$$

$$\frac{\partial^2 G(\lambda, 0)}{\partial\lambda^2} = 0, \quad \frac{\partial^2 G(\lambda, 1)}{\partial\lambda^2} = 2(\alpha+\rho)P(x_0 = 1)$$

$$\frac{\partial^j G(\lambda, 0)}{\partial\lambda^j} = 0, \quad \frac{\partial^j G(\lambda, 1)}{\partial\lambda^j} = 0, \text{for } j \geq 3$$

At $\lambda = 1$, we can write these expressions in the following manner which simplifies their numerical solution:

$$G(1, t+2) = (1+\rho)G(1, t+1) - \rho G(1, t)$$

$$\frac{\partial^k G(1, t+2)}{\partial\lambda^k} = (1+\rho)\frac{\partial^k G(1, t+1)}{\partial\lambda^k} - \rho\frac{\partial^k G(1, t)}{\partial\lambda^k} + k(\alpha+\rho)\frac{\partial^{k-1} G(1, t+1)}{\partial\lambda^{k-1}} - k\rho\frac{\partial^{k-1} G(1, t)}{\partial\lambda^{k-1}},$$

While the initial conditions stated above in equation (11) and leading to:

$$G(1,0) = 1, \quad G(1,1) = 1, \quad \frac{\partial G(1,0)}{\partial\lambda} = P(x_0 = 1),$$

$$\frac{\partial G(1,1)}{\partial\lambda} = \alpha + (1+\rho)P(x_0 = 1)$$

$$\frac{\partial^2 G(1,0)}{\partial\lambda^2} = 0, \quad \frac{\partial^2 G(1,1)}{\partial\lambda^2} = 2(\alpha+\rho)P(x_0 = 1),$$

$$\frac{\partial^j G(1,0)}{\partial\lambda^j} = 0, \quad \frac{\partial^j G(1,1)}{\partial\lambda^j} = 0, \quad \text{for } j \geq 3.$$

Similarly, we can calculate the probabilities of the persistent default process by setting $\lambda = 0$ in the derivatives of the generating functions. In this case, the probabilities are given by:

$$p_i(t) = \frac{1}{i!}\frac{\partial^i G(\lambda, t)}{\partial \lambda^i}, \ i = 0, 1, 2, 3, \cdots, t \text{ with:}$$

$$G(0, t+2) = (1-\alpha)G(0, t+1)$$

$$\frac{\partial^k G(0, t+2)}{\partial \lambda^k} = (1-\alpha)\frac{\partial^k G(0, t+1)}{\partial \lambda^k} + k(\alpha+\rho)\frac{\partial^{k-1} G(0, t+1)}{\partial \lambda^{k-1}} - k\rho\frac{\partial^{k-1} G(0, t)}{\partial \lambda^{k-1}},$$

With the initial conditions:

$$G(0{,}0) = P(x_0 = 0), \quad G(0{,}1) = (1-\alpha)P(x_0 = 0)$$

$$\frac{\partial G(0, 0)}{\partial \lambda} = P(x_0 = 1), \quad \frac{\partial G(0{,}1)}{\partial \lambda} = \alpha P(x_0 = 0) + (1-\alpha-\rho)P(x_0 = 1)$$

$$\frac{\partial^2 G(0{,}0)}{\partial \lambda^2} = 0, \quad \frac{\partial^2 G(0{,}1)}{\partial \lambda^2} = 2(\alpha+\rho)P(x_0 = 1),$$

$$\frac{\partial^j G(\lambda, 0)}{\partial \lambda^j} = 0, \quad \frac{\partial^j G(\lambda, 1)}{\partial \lambda^j} = 0, \text{for } j \geq 3.$$

These equations define a numerical approach to calculating both the moments and the probabilities of a persistent process. A more direct approach will be outlined subsequently however.

Explicit results for the first two moments are provided below with proofs found directly from the moments equations. Let $x_0 = 0$, then the expected number of default loans and its second moment are:

$$E(x(t)) = (1+\rho)E(x(t-1)) - \rho E(x(t-2)) + \alpha, \quad t \geq 2$$

$$E(x(t)^2) = (1+\rho)E(x(t-1)^2) - \rho E(x(t-2)^2) + \frac{\alpha}{1-\rho}\left[2\alpha t + \rho + 1 - 2\left(\frac{\alpha + (1-\alpha)\rho^t - \rho^{t+1}}{1-\rho}\right)\right], \quad t \geq 2$$

In these equations, note that we have as expected (Vallois and Tapiero 2007):

$$E(x(t)) = \frac{\alpha}{1-\rho}\left[t + 1 - \frac{1-\rho^{t+1}}{1-\rho}\right]$$

This clearly indicates the nonlinear time effects of persistence in such default processes. A verification can be reached by setting $\rho = 0$ and $E(x(t)^2) = E(x(t-1)^2) + \alpha[1 + 2\alpha(t-1)]$. Summing for $1 \leq t \leq n$, we obtain $E(x(n)^2) = n\alpha(1 + \alpha(n-1))$. Since $x(n)$ has a binomial distribution $B(n, \alpha)$, we have:

$$E(x(n)) = n\alpha, \quad Var(x(n)) = n\alpha(1-\alpha),$$

and thereby:

$$E(x(n)^2) = n\alpha(1-\alpha) + n^2\alpha^2 = n\alpha(1+(n-1)\alpha) \quad \text{as} \quad \text{expected.}$$

Explicit expressions for the persistent counting probabilities can be determined as well using the recursive probability generating functions. In this case, we calculate $(p_k(t); 0 \le k \le t)$ by recurrence. Initially these are specified by:

$$p_0(0) = 1,\ p_0(1) = 1-\alpha,\ p_1(1) = \alpha$$

Further, $(p_k(t+2); 0 \le k \le t+2)$ is defined as a function of $(p_k(t+1); 0 \le k \le t+1)$ and $(p_k(t); 0 \le k \le t)$ by using the recursive equation

$$p_k(t+2) = (1-\alpha)p_k(t+1) + (\alpha+\rho)p_{k-1}(t+1) - \rho p_{k-1}(t),$$

for all $0 \le k \le t+2$, and by convention, we set $p_k(t) = 0$ if $k < 0$ or $k > t$.

Subsequent calculations will indicate the underlying process probabilities. In particular, we have for the first 3 probabilities:

$$p_0(t) = (1-\alpha)^t, \quad p_1(t) = \alpha(1-\alpha)^{t-2}((1-\alpha-\rho)t+\rho), t \ge 1$$

$$p_2(t) = \alpha \left\{ \begin{matrix} (\alpha+\rho)(1-\alpha)^{t-2} + \frac{1}{2}(1-\alpha)^{t-4}(1-\alpha-\rho)(t-2) \\ (\alpha(1-\alpha-\rho)t + \alpha + 2\rho - \alpha^2 + \alpha\rho) \end{matrix} \right\},\ t \ge 2$$

Of course, when there is no short memory, this is reduced as expected to:

$$p_0(t) = (1-\alpha)^t,\ p_1(t) = \alpha t(1-\alpha)^{t-1}, \quad t \ge 1,$$

$$p_2(t) = \frac{1}{2}\alpha^2(1-\alpha)^{t-2}t(t-1), \quad t \ge 2.$$

Additional and general results are given in Vallois and Tapiero (2007) and in Herrmann and Vallois 2010. These are summarized by:

$$E(y_t y_{t-k}) = \left[\frac{\alpha}{\alpha+\beta}(1-\rho)^k + \rho^k\right]\left[\frac{\alpha}{\alpha+\beta} + \rho^{t-k-1}\left(E(y_1) - \frac{\alpha}{\alpha+\beta}\right)\right]$$

$$E(X_t) = \frac{\alpha}{\alpha+\beta}t + \frac{1-(1-\alpha-\beta)^t}{\alpha+\beta}\left(E(y_1) - \frac{\alpha}{\alpha+\beta}\right)$$

$$\text{var}\,(X_t) = E(X_t)[1-E(X_t)] + 2\sum_{i=1}^{t}\sum_{k=1}^{t-1} E(y_i y_{i-k}) \quad \text{where}$$

$$\sum_{i=1}^{t}\sum_{k=1}^{t-1} E(y_i y_{i-k}) = \left(\frac{\alpha}{\alpha+\beta}\right)^2 \sum_{i=1}^{t}\sum_{k=1}^{t-1}\left[(1-\rho)^k + \frac{\alpha+\beta}{\alpha}\rho^k\right] \left[1 - \rho^{i-k-1}\left(1 - \frac{\alpha+\beta}{\alpha}E(y_1)\right)\right]$$

Which provides an analytical (rather than an empirical defined earlier) co-variation between events.

10.3 Compound Processes with Short Memory

Finally, if an event has risk consequences (for example, a loss, a gain, etc.) then their aggregate has a generating function that differs from the standard approach we use in calculated compound processes. Explicitly, assume that every loan defaulting has a loss probability distribution $\tilde{Z}_j$. The total loss over a time interval (0,t) is thus a Compound Loss process given by $\xi_t = \sum_{j=1}^{x_t} \tilde{Z}_j$ with mean $\mu = E(\tilde{Z}_i)$ and known variance var $(\tilde{Z}_i)$. Then, for independent loss processes, the mean loss up to time t and its variance are:

$$E\left(\sum_{i=1}^{x_t} \tilde{Z}_i\right) = E(x_t)E(\tilde{Z}_i), \quad \text{var}\left(\sum_{i=1}^{x_t} \tilde{Z}_i\right) = [E(\tilde{Z}_i)]^2 Var(x_t) + Var(\tilde{Z}_i)E(x_t).$$

However, since default is persistence-dependent a more general expression can be found. Explicitly, consider the following random default:

$$S_t = \sum_{i=0}^{t} \tilde{Z}_i^0 1_{\{y_i=0\}} + \sum_{i=0}^{t} \tilde{Z}_i^1 1_{\{y_i=1\}} \text{ where } x_t = \sum_{i=0}^{t} y_i$$

where $\tilde{Z}_i^0$ is a "normal expense" occurring in any regular period (defined by the fact that no specific loan has defaulted) while $\tilde{Z}_i^1$ is a "large loss due to a loan defaulting" (of course, if $\tilde{Z}_i^0 = \tilde{Z}_i^1$ then $S_t = \xi_t$ as stated above). We assume that $\left\{\tilde{Z}_i^0, \tilde{Z}_i^1\right\}$ are random variables independent of each other and independent of the Markov (persistent) claims. In this case, the compound claim mean and variance and the claim probability generating function are given by the following (with proofs provided in the appendix):

$$E(S_t) = (t+1)E\left(\tilde{Z}_1^0\right) + \left(E\left(\tilde{Z}_1^1\right) - E\left(\tilde{Z}_1^0\right)\right)E(x_t)$$

$$\begin{aligned} E(S_t^2) &= (t+1)\left[\text{var}\left(\tilde{Z}_1^0\right) + (t+1)\left(E\left(\tilde{Z}_1^0\right)\right)^2\right] + E\left(x_t^2\right)\left[E\left(\tilde{Z}_1^1\right) - E\left(\tilde{Z}_1^0\right)\right]^2 \\ &\quad + E(x_t)\left[\text{var}\left(\tilde{Z}_1^1\right) - \text{var}\left(\tilde{Z}_1^0\right) + 2(t+1)E\left(\tilde{Z}_1^0\right)\left(E\left(\tilde{Z}_1^1\right) - E\left(\tilde{Z}_1^0\right)\right)\right] \end{aligned}$$

While their Laplace Transform is:

$$E\left(e^{-\lambda S_t}\right) = \left[E\left(e^{-\lambda \tilde{Z}_i^0}\right)\right]^{t+1} G(z,t), z = \frac{E\left(e^{-\lambda \tilde{Z}_1^1}\right)}{E\left(e^{-\lambda \tilde{Z}_1^0}\right)}, \ \lambda \geq 0$$

where $G(z,t)$ is a probability generating function:

$$G(z,t) = E(z^{x_t}) = \sum_{i=0}^{t+1} z^i P(x_t = i).$$

Note that when $\tilde{Z}_1^0$ is null, then, we find that:

$$E(S_t) = E(\tilde{Z}_1^1)E(x_t) \text{ and } Var(S_t) = E(x_t)Var(\tilde{Z}_1^1) + Var(x_t)(E(\tilde{Z}_1^1))^2$$

These results can then be used to obtain a pricing of the portfolio as seen earlier as well as be used to assess a prospective default of loans over time (providing a default referential on the basis of which calculations can be made). Note however that the variance of this process has increased over time due to the process persistence. Further, based on these moments, the Value at Risk (VaR) risk exposure can be determined which would use these two moments as a first approximation (although higher order moments can be calculated as well, the generating function of the persistent Compound Process).

A numerical analysis of our equations will reveal some of the characteristics of default persistent process. As expected, the mean evolution of the persistent process has an almost linear growth as indicated in our equation. In the long run, the variance turns out to be also almost linear, as it is the case for random walks. However, persistence ($\rho > 0$) has the effect of increasing the variance as shown in figure below. In the short term however, the variance evolution is nonlinear as our equations have indicated. Interestingly, the rate of change in variance is not constant and growing over time which indicates a "persistent volatility". Such a phenomenon is observed in fractal stochastic differential equations. Of particular interest is the evolution of the third moment of the persistent claim distribution. Empirical analyses (see also Vallois and Tapiero…) have also shown that initially, it was increasing (over 4 periods) and subsequently declining (although remaining positive for ρ positive). When ρ is negative we noted that for the first few periods the evolution of the mean and the variance are indeed nonlinear. This is particularly the case for the variance as shown in figure below. In this figure, the variance initially declines, then increases and again decreases. Finally, it converged to a linear growth. This behavior is indicative of the short term effects of memory on the stochastic process as indicated earlier. This particular observation will allow us to make mean variance approximations validated over the long run(or at least subsequent to a short period when the process variance may be volatile).

Further, we noted in some cases we note the divergence in the growth of volatility when the parameter rho is negative. Finally, the third moment is positive which demonstrate. This latter observation is particularly important for it may be used to explain partly the skew of certain time series, presuming that this skew is due to the short term memory effects prevalent in such series (for example, in financial time series). Some graphs are drawn below (Figs 3 and 4).

These graphs point out that short memory processes are both of theoretical and practical importance.

10.4 Short Memory, Lognormal Process and a Binomial Approximation

The lognormal price process is a normally distributed rates of returns process. Explicitly, it is defined by:

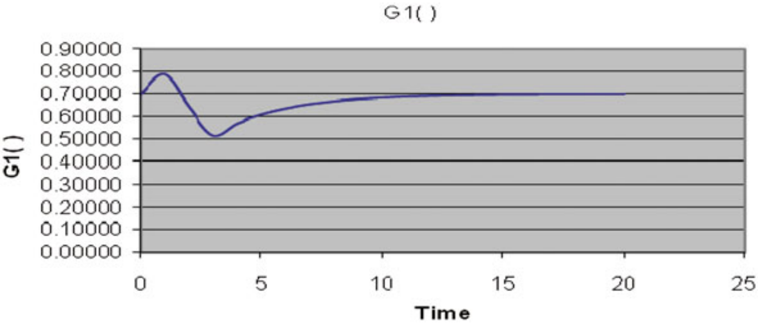


Fig. 3 $\rho = -0.3$

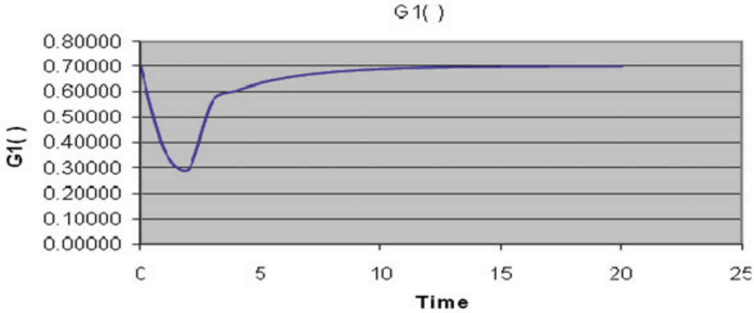


Fig. 4 The Probability of $\rho = 0.3$

$$\frac{dS(t)}{S(t)} = \alpha dt + \sigma dW(t), S(0) > 0$$

Letting $y(t) = \ln S(t)$, then an Ito's calculus transformation yields a rates of returns $\{y, t \geq 0\}$ process:

$$dy(t) = \left(\alpha - \frac{1}{2}\sigma^2\right) \Delta t + \sigma dW(t), S(0) > 0$$

A short memory process in such a model may be constructed in different manners. For discussion purposes, we consider a binomial approximation to such a process given by:

$$\Delta y^+(t) = \{\Delta y(t) \,|\, y(t-1) = 1\} = \begin{cases} Y^{++}(y,t) - y & wpP^{++}(y,t) \\ Y^{-+}(y,t) - y & wp1 - P^{++}(y,t) \end{cases}$$

$$\Delta y^-(t) = \{\Delta y(t) \,|\, y(t-1) = 0\} = \begin{cases} Y^{+-}(y,t) - y & wpP^{-+}(y,t) \\ Y^{--}(y,t) - y & wp1 - P^{-+}(y,t) \end{cases}$$

In other words, given an outcome (1, − 1, 0, etc.) the probability of a rates of returns increasing or decreasing to say by $Y^{++}(y,t) - y$ or $Y^{-+}(y,t) - y$ depends on the previous outcome (in case the rates of returns previously increased). Its probability of increasing or decreasing is similarly dependent on that previous outcome. As a result, a financial pricing model is in fact a 6 parameters model (4 state prices and 2 probabilities, defined under an appropriate probability measure) rather than two state prices and one probability. Such a pricing model requires then far more information to be defined. Further, it implies that a model that negates the effects of memory in fact embed short memory in the pricing process. For demonstration purposes, say that short memory has an effect only on the probability measure defining the subsequent increase or decrease in a rates of returns while state prices remain identical. In this case, the 2 binomial processes defined above are reduced to the following:

$$\Delta y^{+}(t) = \{\Delta y(t)\,|y(t-1) = 1\} = \begin{cases} Y^{+}(y,t) - y & wpP^{+}(y,t) \\ Y^{-}(y,t) - y & wp1 - P^{+}(y,t) \end{cases}$$

$$\Delta y^{-}(t) = \{\Delta y(t)\,|y(t-1) = 0\} = \begin{cases} Y^{+}(y,t) - y & wpP^{-}(y,t) \\ Y^{-}(y,t) - y & wp1 - P^{-}(y,t) \end{cases}$$

In a one period memory, $\Delta y^{+}(t)$ is the increment in $\Delta y(t)$ when at time t, the stock price has increased previously. Similarly, $\Delta y^{-}(t)$ is the increment of a same size as $\Delta y^{+}(t)$ (but differs in its expectation) when in the previous period, the price was down. The probabilities $P^{+}(y,t)$ and $P^{-}(y,t)$ differ however, expressing a propensity for a market to increase or decrease following a price increase or decrease. If $P(y,t) = P^{+}(y,t) = P^{-}(y,t)$, this corresponds to a random walk with no short memory. In this sense, $\left\{Y^{+}(y,t), Y^{-}(y,t)\right\}$ define two states, one corresponding to a stock price increase and the other corresponding to a stock price decrease. Memory of the last period is then expressed by the probability of the stock increasing or decreasing based on the previous movement—being up or down. The following probabilities result:

$$\begin{cases} P\left(y_{t-1}^{+}|y_{t-1}^{-}\right) = P_{t-1}^{-}, P\left(y_{t-1}^{-}|y_{t-1}^{-}\right) = 1 - P_{t-1}^{-} \qquad [+ \qquad -] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{or} \\ P\left(y_{t-1}^{+}|y_{t-1}^{+}\right) = 1 - P_{t-1}^{+}, P\left(y_{t-1}^{-}|y_{t-1}^{+}\right) = P_{t-1}^{+} \\ \mathbf{P} = \begin{bmatrix} - \\ + \end{bmatrix} \begin{bmatrix} 1 - P_{t-1}^{-} & P_{t-1}^{-} \\ P_{t-1}^{+} & 1 - P_{t-1}^{+} \end{bmatrix} \end{cases}$$

And therefore, a rate of returns increase has 4 possibilities:

$$\Delta y(t) = \begin{cases} \Delta y_t^{+-} = Y^{+}(y_{t-\Delta t}^{-}) - y_{t-\Delta t}^{-} & wp & P_{t-\Delta t}^{-} \\ \Delta y_t^{--} = Y^{-}(y_{t-\Delta t}^{-}) - y_{t-\Delta t}^{-} & wp & 1 - P_{t-\Delta t}^{-} \\ \Delta y_t^{++} = Y^{+}(y_{t-\Delta t}^{+}) - y_{t-\Delta t}^{+} & wp & 1 - P_{t-\Delta t}^{+} \\ \Delta y_t^{-+} = Y^{-}(y_{t-\Delta t}^{+}) - y_{t-\Delta t}^{+} & wp & P_{t-\Delta t}^{+} \end{cases}$$

Corresponding to two Binomial models with $(P_{t-\Delta t}^{+}, P_{t-\Delta t}^{-})$, the unknown probabilities. Next assume a stock price whose rate of returns and volatility are respectively

(μ, σ) (and therefore, a lognormal model). If we were to equate the discretized and binomial approximated such model (note that it is not a pricing model) then in a time interval Δt, a mean-variance approximation equating the normal process with mean $\left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right)$ and its variance $\sigma^2\Delta t$ yields the following equalities to the short memory process defined above:

$$\begin{aligned}\left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right) &= E(\Delta y(t)) \\ &= \left(\Delta y_t^{+-}\right) P_{t-\Delta t}^{-} + \left(\Delta y_t^{--}\right)\left(1 - P_{t-\Delta t}^{-}\right) + \left(\Delta y_t^{+-}\right)\left(1 - P_{t-\Delta t}^{+}\right) \\ &\quad + \left(\Delta y_t^{-+}\right) P_{t-\Delta t}^{+}\sigma^2\Delta t = E(\Delta y(t))^2 - [E\,(\Delta y(t))]^2 \\ &= \left(\Delta y_t^{+-}\right)^2 P_{t-\Delta t}^{-} + \left(\Delta y_t^{--}\right)^2\left(1 - P_{t-\Delta t}^{-}\right) \\ &\quad + \left(\Delta y_t^{++}\right)^2\left(1 - P_{t-\Delta t}^{+}\right) + \left(\Delta y_t^{-+}\right)^2 P_{t-\Delta t}^{+} - [E\,(\Delta y(t))]^2\end{aligned}$$

This is a system of two equations in the two probabilities $(P_{t-\Delta t}^{-}, P_{t-\Delta t}^{+})$ when the four states $\Delta y_t^{+-}, \Delta y_t^{++}, \Delta y_t^{--}, \Delta y_t^{-+}$ are given by definition. Thus, two equations in $P_{t-\Delta t}^{+}$ and $P_{t-\Delta t}^{-}$:

$$\begin{aligned}&\left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right) - \left(\Delta y_t^{--} + \Delta y_t^{+-}\right) = \left(\Delta y_t^{+-} - (\Delta y_t^{--})\right) P_{t-\Delta t}^{-} \\ &+ \left(\Delta y_t^{-+} - \Delta y_t^{+-}\right) P_{t-\Delta t}^{+}\left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right)^2 \\ &+ \sigma^2\Delta t - \left(\left(\Delta y_t^{--}\right)^2 + \left(\Delta y_t^{++}\right)^2\right) = \left(\left(\Delta y_t^{+-}\right)^2 - \left(\Delta y_t^{--}\right)^2\right) \\ &P_{t-\Delta t}^{-} + \left(\left(\Delta y_t^{-+}\right)^2 - \left(\Delta y_t^{++}\right)^2\right) P_{t-\Delta t}^{+}\end{aligned}$$

And therefore

$$P_{t-\Delta t}^{+} = \frac{\begin{Bmatrix}\left((\Delta y_t^{+-})^2 - (\Delta y_t^{--})^2\right)\left((\mu\Delta t - \frac{1}{2}\sigma^2\Delta t) - (\Delta y_t^{--} + \Delta y_t^{+-})\right) \\ -\left(\Delta y_t^{+-} - (\Delta y_t^{--})\right)\end{Bmatrix} \left\{\left[\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right]^2 + \sigma^2\Delta t - \left((\Delta y_t^{--})^2 + (\Delta y_t^{++})^2\right)\right\}}{\begin{array}{c}\left\{\left((\Delta y_t^{+-})^2 - (\Delta y_t^{--})^2\right)(\Delta y_t^{-+} - \Delta y_t^{+-}) \right.\\ \left. - \left(\Delta y_t^{+-} - (\Delta y_t^{--})\right)\left((\Delta y_t^{-+})^2 - (\Delta y_t^{++})^2\right)\right\}\end{array}}$$

Similarly, the probability $P_{t-\Delta t}^{-}$ can be defined. Note that this probability is defined in terms of all previous potential results. By the same token solving for μ and σ^2 will provide a solution in terms of both probabilities. To highlight their dependence we consider a simple numerical example. Say that:

$$\Delta y(t) = \begin{cases}\Delta y_t^{+-} = 0.12 & wp \quad P_{t-\Delta t}^{-} \\ \Delta y_t^{--} = -0.08 & wp \quad 1 - P_{t-\Delta t}^{-} \\ \Delta y_t^{++} = 0.10 & wp \quad 1 - P_{t-\Delta t}^{+} \\ \Delta y_t^{-+} = -0.04 & wp \quad P_{t-\Delta t}^{+}\end{cases}$$

Then

$$\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t = 0.12P^-_{t-\Delta t} - 0.08(1 - P^-_{t-\Delta t})$$
$$+ 0.10(1 - P^+_{t-\Delta t}) - 0.04P^+_{t-\Delta t}$$
$$\sigma^2\Delta t + \left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right)^2 = (0.12)^2 P^-_{t-\Delta t} + (0.08)^2(1 - P^-_{t-\Delta t})$$
$$+ (0.10)^2(1 - P^+_{t-\Delta t}) + (0.04)^2 P^+_{t-\Delta t}$$

And

$$\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t - 0.02 = 0.04P^-_{t-\Delta t} - 0.14P^+_{t-\Delta t}\sigma^2\Delta t + \left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right)^2$$
$$-(0.0164) = (0.008)P^-_{t-\Delta t} - 0.0084P^+_{t-\Delta t}$$

And therefore,

$$P^+_{t-\Delta t} = \frac{\sigma^2\Delta t + \left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right)^2 - 0.2\left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t - 0.02\right)}{0.0196}$$
$$P^-_{t-\Delta t} = 178.56\left(\sigma^2\Delta t + \left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t\right)^2 - 0.0164\right)$$
$$+ 10.71\left(\mu\Delta t - \frac{1}{2}\sigma^2\Delta t - 0.02\right)$$

And inversely, the lognormal model mean and variance imply the probabilities $(P^-_{t-\Delta t}, P^+_{t-\Delta t})$:

$$\mu\Delta t = 0.0282 + 0.044P^-_{t-\Delta t} - 0.1442P^+_{t-\Delta t} - \frac{1}{2}(0.02 + 0.04P^-_{t-\Delta t} - 0.14P^+_{t-\Delta t})^2$$
$$\sigma^2\Delta t = 0.0164 + (0.008)P^-_{t-\Delta t} - 0.0084P^+_{t-\Delta t} - (0.02 + 0.04P^-_{t-\Delta t} - 0.14P^+_{t-\Delta t})^2$$

For small time intervals, we have then,

$$P^+_{t-\Delta t} = 52.04(\sigma^2\Delta t) - 2.0833(\mu\Delta t) - 0.04166$$
$$P^-_{t-\Delta t} = 173.205(\sigma^2\Delta t) + 10.71(\mu\Delta t) - 3.1425$$

And inversely,

$$\mu\Delta t = \frac{52.04\left\{P^-_{t-\Delta t} + 3.1425\right\} - 173.205\left\{P^+_{t-\Delta t} + 0.04166\right\}}{52.04 * 10.71},$$
$$\sigma^2\Delta t = 0.02P^+_{t-\Delta t} - 0.0523P^-_{t-\Delta t}$$

Note that in this case, the variance is very small compared to the mean.

Next assume instead a risk neutral pricing model under a probability measure Q,

$$\frac{dS(t)}{S(t)} = R_f dt + \sigma dW(t), S(0) > 0 \text{ and } S(0) = e^{-R_f t} E^Q(S(t))$$

In this case, we may have a short memory pricing process with probabilities $(P_{t-\Delta t}^{-Q}, P_{t-\Delta t}^{+Q})$ given by:

$$R_f \Delta t = \frac{52.04 \left\{ P_{t-\Delta t}^{-Q} + 3.1425 \right\} - 173.205 \left\{ P_{t-\Delta t}^{+Q} + 0.04166 \right\}}{52.04 * 10.71}$$
$$\sigma^2 \Delta t = 0.02 P_{t-\Delta t}^{+Q} - 0.0523 P_{t-\Delta t}^{-Q}$$

Or probabilities defined as a function of a stock price volatility:

$$P_{t-\Delta t}^{+Q} = 52.04(\sigma^2 \Delta t) - 2.0833(R_f \Delta t) - 0.04166$$
$$P_{t-\Delta t}^{-Q} = 173.205(\sigma^2 \Delta t) + 10.71(R_f \Delta t) - 3.1425$$

If However, a call option, $C(0) = e^{-R_f t} E^Q(Max(S(t) = K, 0))$ its implied volatility can be calculated and therefore assuming $\sigma^2 \Delta t = 0.015$ as well as a risk free rate of 0.08, we have the following implied risk neutral probabilities for each of these two binomial processes:

$$P_{t-\Delta t}^{+Q} = 0.7802 - 2.0833(R_f \Delta t) = 0.6135$$
$$P_{t-\Delta t}^{-Q} = 10.71(R_f \Delta t) - 0.5444 = 0.2556$$

We shall see below that such a process may converge to three limits: a deterministic, a normal distribution and finally to a Telegraphic equation. Each of these limits depends on the memory conditional Markov probabilities. While the proof of these results requires an extensive analysis, we shall summarize basic results and refer to Herrmann and Vallois (2010). A discussion of these results follows.

10.5 Continuous Time Limits and Convergence

Continuous time short memory models, based on limits of their underlying transitions may seem a contradiction, with limits pointing out to "long run" results based on infinitesimal Markov transitions and their short memory effects on the limit stochastic models. In this section we consider a number of results and theirs implications based mostly on the Herrmann and Vallois 2010 paper. To do so, we assume that the Markov transition probabilities consist of the discrete (binomial) probabilities considered earlier and denoted by (α_0, β_0) as well as small perturbations $(\Delta_\alpha, \Delta_\beta)$ defining the Markov chain transition probabilities:

$$\alpha = \alpha_0 + \Delta_\alpha(\Delta_\alpha \to 0) \text{ and } \beta = \beta_0 + \Delta_\beta(\Delta_\beta \to 0)$$

Table 1 A summary of limit results

	$\Delta_x \to \Delta_t$	$(\Delta_x)^2 \to \sigma\Delta_t$	$(\Delta_x)^3 \to \xi\Delta_t$
$\rho_0 = 0$	Deterministic walk	Random walk	–
$\rho_0 \neq 0$	–	BM with drift	–
$\rho_0 = 1$	Planar Poisson process	Telegraphic equation	If $\rho_0 = -1$, BM with no drift

where $\rho_0 = 1 - (\alpha_0 + \beta_0)$ is used as an index of asymmetry. These result in the Markov chain where for convenience we write: $\Delta_\alpha = c_0\Delta_x$ and $\Delta_\beta = c_1\Delta_x$ with Δ_x denoting a transition probability perturbation which is converges at a time linear rate $\Delta_x \to r\Delta_t$ or converging in its quadratic perturbation $(\Delta_x)^2 \to \sigma\Delta_t$. As a result, we obtain the following 2 states Markov Chain, with transitions in $(1, -1)$ (rather than 1,0 as considered earlier):

$$\mathbf{P} = \begin{array}{c} \\ +1 \\ -1 \end{array}\begin{array}{c} \begin{array}{cc} \quad +1 \quad\quad\quad\quad & \quad\quad -1 \end{array} \\ \begin{bmatrix} 1-(\alpha_0 + c_0\Delta_x) & \alpha_0 + c_0\Delta_x \\ \beta_0 + c_1\Delta_x & 1-(\beta_0 + c_1\Delta_x) \end{bmatrix} \end{array}, \quad 0 < \alpha, \beta < 1$$

A number of cases will be considered based on both the index of asymmetry of the underlying short memory process and the perturbations $(\Delta_\alpha, \Delta_\beta)$. These limits determine the convergence of the continuous time short memory process to either a constant process, a Brownian motion process with drift (albeit with the memory parameters embedded in the process drift and its volatility) reminiscent of a central limit theorem and finally a convergence to Telegraphic equation (a wave partial differential equation). The implications of these results, are that the underlying assumption of most financial Brownian models are either "hiding" memory parameters or are just limit models in an environment and profession which is based essentially on short run transactions and trades. A number of cases are summarized in the following 1.

Continuous time limits are reached by considering a partition of time into n discrete events Y_i with, $X_t = Y_0 + Y_1 + Y_2 + \cdots + Y_t$, $\quad t \geq 0$ while the time between events is assumed exponential. Note that Y_n defines a random walk if and only if $\rho_0 = 0$ or $1 = \alpha_0 + \beta_0$ where ρ_0 is an index of asymmetry with $\rho_0 = 1 - \alpha_0 - \beta_0$. For convenience we also set $\eta_0 = \alpha_0 - \beta_0$.

As a result, the process Y_n defines a Markov chain with two states, $+1$ to denote a price increase and -1 to denote a price decrease. Note that this process, when one is a given state and the probability of switching to another process is extremely small, the process will converge at the limit to a telegraphic partial differential equation. We normalize the process $\{X_t, t > 0\}$ associated to event Y_t by considering a small interval of time Δ_t and define a limit counting process $Z^\Delta(s\Delta t) = \Delta_x X_s, \quad k \in \mathbb{N}$ or $Z_s^\Delta = \Delta_x X_{s/\Delta t}$. A continuous time approximation is given by $\tilde{Z}_s^\Delta$ obtained through an interpolation of Z_t^Δ. Intuitively, we have then two processes resulting in a single events counting process, each corresponding to another past event (the current memory) which we count by $N_t^{c_0,c_1} = \sum_{k\geq 1} \mathbf{1}_{\left\{\sum_{i=1} \lambda_i e_i \leq t\right\}}$. Note here that the e_i are exponential times taken to have a mean 1 while $1/\lambda_i$ are the mean events defined by the transition probabilities of both processes. Explicitly, if $\rho_0 = 1$, then $\alpha_0 = \beta_0 = 0$ and evolution of the short memory process depends only on the probabilities c_0, c_1 with:

$$1/\lambda_i = \begin{cases} c_0 & if\ i \text{ is odd} \\ c_1 & otherwise \end{cases}$$

In such a process, when the process is in state $+1$, the probability of remaining in that state is $1 - c_0\Delta_x$ which is almost equal to 1 since Δ_x is very small. While the probability of switching to -1 in a small interval of time is very small. Similarly if the process is in state -1, the probability of switching to a $+1$ state is extremely small resulting in a process alternating "infrequently" from one state to another. For example, is pricing model is defined by two states, one of price expansion and the other of price depreciation, then it will be defined by an underlying process that remains an appreciable amount of time in the $+1$ state followed by an appreciable amount of time in its -1 state. Of course, the time between a switch from one state to another will depend on the exponential time of transition (a Gamma Probability distribution, defined by the number of exponential events occurring for a process switch to occur) and the means of these exponentials when in one of the two processes. In the special case $c_0 = c_1$, the number of events process $\left\{N_t^{c_0}, t \geq 0\right\}$ is then a usual Poisson process whose Poisson rate is $c_0 t$ and $N_0^{c_0} = 0$. Generally, Herrmann and Vallois prove that the probability distribution of the counting process is:

$$P\left\{N_t^{c_0}\right\} = \begin{cases} P\left\{N_t^{c_0} = 2k\right\} = \dfrac{(c_0 c_1)^k \alpha_k(t)}{2^{2k} k!(k-1)!} e^{\frac{c_0+c_1}{2}} \\ \text{with } \alpha_k(t) = \int\limits_{-t}^{t} (t-z)^{k-1}(t+z)^k e^{\frac{c_1-c_0}{2}} \\ P\left\{N_t^{c_0} = 2k+1\right\} = \dfrac{{c_0}^{k+1} {c_1}^k \bar{\alpha}_k(t)}{2^{2k}(k!)^2} e^{\frac{c_0+c_1}{2}} \\ \text{with } \bar{\alpha}_k(t) = \int\limits_{-t}^{t} (t-z)^k (t+z)^k e^{\frac{c_1-c_0}{2}} \end{cases}$$

As a result, the process Z_s^{Δ} corresponding to the counting process $N_t^{c_0,c_1}$ (for both processes, $+1$ and -1) is defined by:

$$Z_t^{c_0,c_1} = \int\limits_0^t (-1)^{N_u^{c_0,c_1}} du$$

In other words, when the number of events was odd, then the memory is -1 and otherwise it is necessarily $+1$. This latter process is in fact a Poisson planar process which can be shown under certain condition to converge to a Brownian motion model. Herrmann and Vallois (2010) point out that the process $Z_t^{c_0,c_1}$ is non-Markovian and that its conditional probability distribution (on the counting process $N_t^{c_0}$) provides the following distribution:

$$P\left\{Z_t^{c_0,c_1}\right\} = \begin{cases} P\left\{Z_t^{c_0,c_1} \in dz \,\middle|\, N_t^{c_0} = 2k\right\} \\ = \dfrac{1}{\alpha_k(t)}(t-z)^{k-1}(t+z)^k e^{\frac{c_1-c_0}{2}} 1_{[-1,1]}(z) \quad k \geq 1 \\ P\left\{N_t^{c_0} \in dz = 2k+1\right\} \\ = \dfrac{1}{\bar{\alpha}_k(t)}(t-z)^k(t+z)^k e^{\frac{c_1-c_0}{2}} 1_{[-1,1]}(z) \qquad k \geq 0 \end{cases}$$

Additional results, such as special cases for $c_0 = c_1$ are also derived. In addition, the probability distribution of $Z_t^{c_0,c_1}$ is calculated explicitly and while $Z_t^{c_0,c_1}$ is non Markovian, $\left\{Z_t^{c_0,c_1}, N_t^{c_0,c_1}\right\}$ is Markovian while by definition:

$$\frac{dZ_t^{c_0,c_1}}{dt} = (-1)^{N_t^{c_0,c_1}}$$

Further for $\rho_0 = 1$, we shall see that at the limit, the short memory Markov process converges to a Telegraphic (wave) stochastic process (see case 4).

Given these definitions, the following special cases result, each of which is discussed and with essential proofs to be found in (see Herrmann and Vallois 2010):

Case 1: $\rho_0 \neq 1$ and $\Delta_x \to r\Delta_t$

This case corresponds to a "slow convergence" of the Markov probability to its linear time increment, $\Delta_x = r\Delta_t, r > 0$. The counting process $\{\tilde{Z}_s^{\Delta}, s > 0\}$ converges then to a linear trend deterministic Φrt with $\Phi = \frac{(\alpha_0-\beta_0)}{1-\rho_0}$ as $\Delta_x \to 0$. For example, if $\alpha_0 = 0.4, \quad \beta_0 = 0.3$, then the limit is: $-0.333rt$.

Case 2: $\rho_0 \neq 1$ and **and** $(\Delta_x)^2 \to r\Delta_t$

This case corresponds to a quadratic convergence of the Markov probability to linear time increment, $(\Delta_x)^2 \to r\Delta_t$. Consider the random increment $Z_s^{\Delta} = \Delta_x X_{s/\Delta t}$ and defined the following variable:

$$\tilde{\xi}_t^{\Delta} = \tilde{Z}_t^{\Delta} + \Phi\sqrt{r}\frac{t}{\sqrt{\Delta_t}}$$

Where $t/\sqrt{\Delta_t}$ ids the number of time increments. Then, Herrmann and Vallois (2010) prove that $\xi^{\Delta}(t)$ converge in distribution to a Brownian motion with drift and volatility defined by $\mu t + \sigma B_t$ where B_t a standard Brownian:

$$\mu = r\left(-\frac{\bar{c}}{1-\rho_0} - \frac{(\alpha_0-\beta_0)c}{(1-\rho_0)^2}\right) \quad \text{and} \quad \sigma = \sqrt{\frac{r(1+\rho_0)}{1-\rho_0}\left(1 - \frac{(\alpha_0-\beta_0)^2}{(1-\rho_0)^2}\right)}$$

Or defining an underlying Ito linear stochastic differential equation which we denote by dB_t^{ρ} to denote a memory prone Brownian motion, since the above process reduces to a standard Brownian motion when there is no memory. Thus,

$$dB_t^{\rho} = -r\left(\frac{1}{1-\rho_0}(\bar{c}-\Phi c)\right)dt + \sqrt{\frac{r(1+\rho_0)}{1-\rho_0}(1-\Phi^2)}dB_t,\ B_0^{\rho} = Y_0$$

where $\bar{c} = c_1 - c_0$ and $c = c_1 + c_0$ and therefore $\bar{c} - \Phi c = c_1(1 - \Phi) - c_0(1 + \Phi)$. In other words, short memory perturbs the Brownian motion. A pure random walk is then defined if the transition probabilities are symmetric, i.e. $\Phi = 0$ as well as $\bar{c} = c_1 - c_0 = 0$ and further $\rho_0 = 0$. As a result,

$$dB_t^\rho = \sqrt{r} dB_t, \quad B_0^\rho = Y_0$$

And r is an arbitrary parameter, which can be set to 1, thus $B_t^\rho \equiv B_t$. In this sense, short memory perturb predictably (in our case) the Brownian motion by altering both a process drift and its volatility. Explicitly, for a lognormal price process, with mean and volatility (μ, σ) and a short memory perturbed Brownian motion, we have:

$$\frac{dS(t)}{S(t)} = \mu dt + \sigma dB_t^\rho, S(0) > 0$$

We have instead:

$$\frac{dS(t)}{S(t)} = \left\{ \mu - r \left(\frac{1}{1 - \rho_0} (\bar{c} - \Phi c) \right) \right\} dt + \sigma \sqrt{\frac{r(1 + \rho_0)}{1 - \rho_0} (1 - \Phi^2)} dB_t, S(0) > 0$$

In this case, the rate of returns on the price is:

$$\begin{aligned} dR(t) = & \left\{ \mu - r \left(\frac{1}{1 - \rho_0} (\bar{c} - \Phi c) \right) - \frac{1}{2} \sigma^2 \left(\frac{r(1 + \rho_0)}{1 - \rho_0} (1 - \Phi^2) \right) \right\} dt \\ & + \sigma \sqrt{\frac{r(1 + \rho_0)}{1 - \rho_0} (1 - \Phi^2)} dB_t, S(0) > 0 \end{aligned}$$

Further, under an assumption of complete markets, and relative to a risk neutral probability measure, we have:

$$\begin{aligned} & \frac{dS(t)}{S(t)} = R_f dt + \sigma \sqrt{\frac{r(1 + \rho_0)}{1 - \rho_0} (1 - \Phi^2)} dB_t, S(0) > 0 \quad \text{And} \\ & S(0) = e^{R_f t} E^Q(S(t)) dt \end{aligned}$$

The risk premium in this case is:

$$\pi = \frac{\mu - r \left(\frac{1}{1 - \rho_0} (\bar{c} - \Phi c) \right) - R_f}{\sigma \sqrt{\frac{r(1 + \rho_0)}{1 - \rho_0} (1 - \Phi^2)}}$$

Of course, if $\rho_0 \neq 0$ while $\alpha_0 = \beta_0$ ($\Phi = 0$) then the risk premium is a function of the short memory asymmetry in its basic Marko probabilities and in their perturbations:

$$\pi = \frac{(\mu - R_f)(1 - \rho_0) - r(c_1 - c_0)}{\sigma \sqrt{r(1 - \rho_0{}^2)}}$$

And if $c_1 = c_0$ and r = 1, then,

$$\pi = \left(\frac{\mu - R_f}{\sigma}\right)\sqrt{\frac{1-\rho_0}{1+\rho_0}}$$

And therefore for $\rho_0 > 0$ or $\alpha_0 + \beta_0 < 1$, the risk premium is smaller while for $\rho_0 < 0$, the risk premium is greater.

The implications of these results imply that although a short memory process may converge in distribution to a Brownian motion, the process drift and its volatility are also a function of the underlying process short memory as well as the asymmetry in the process probabilities when they move from one state (say -1, or $+1$) to the other state (in this case, $+1$ and -1 respectively). Further, both the mean rate of return and the volatility depend on a common set of parameters that define the process short memory. For a short term pricing model and based on momentum trading using a lognormal pricing model, we may over or underestimate the effects of short memory. Of course, for no short term memory, $\rho_0 = 0$ (and therefore the probability on an increase or a decrease in rates of returns are statistically independent) and hereby, reduced to the standard risk neutral pricing model.

Case 3: $\rho_0 = -1$ and $\alpha_0 = \beta_0 = 1$ and $(\Delta_x)^2 \to r\Delta_t,\ r > 0$

This case corresponds to $\rho_0 = -1 = 1 - \alpha_0 - \beta_0$ and therefore $\alpha_0 = \beta_0 = 1$ which implies that once in $+1$ or in a -1 process, we remain in this process (i.e. it is an absorbing state). However, if these probabilities are perturbed by a quadratic variation, Herrmann and Vallois (2010) show that $\tilde{Z}_t^{\Delta}$ converges to the Gaussian distribution with mean zero and variance equal to $r\,|c_0|\,t$ where $c_1 = c_0 < 0$. In other words, since $\Phi = 0$, we have $\tilde{\xi}_t^{\Delta} = \tilde{Z}_t^{\Delta}$ and therefore at the limit, the following stochastic process results:

$$dB_t^{\rho} = r\,|c_0|\,dB_t$$

Case 4: $\rho_0 = 1$ **and** $\Delta_x \to \Delta_t$

This case corresponds to $\alpha_0 = \beta_0 = 0$ and therefore to the Markov switching model defined by the perturbations:

$$\mathbf{P} = \begin{array}{c} \\ +1 \\ -1 \end{array}\!\!\begin{array}{c} \begin{array}{cc} \quad +1 \quad\quad & \quad\quad -1 \end{array} \\ \begin{bmatrix} 1 - c_0\Delta_x & c_0\Delta_x \\ c_1\Delta_x & 1 - c_1\Delta_x \end{bmatrix} \end{array},$$

As note earlier, $\Delta_x = \Delta_t$ and $c_0 = c_1$ then the counting process $N_t^{c_0}$ is a Poisson process. Further, if $\Delta_x = \Delta_t$ and the initial condition is $Y_0 = X_0 = +1$ then the continuous time process $\{\tilde{Z}_s^{\Delta}, s > 0\}$ converges in distribution to the counting process $\left\{Z_s^{c_0,c_1}, s \geq 0\right\}$ defined earlier. Similarly, if $\Delta_x = \Delta_t$ and $Y_0 = X_0 = -1$ then $\left\{\tilde{Z}_s^{\Delta}, s > 0\right\}$ converges in distribution to the counting process $\left\{-Z_s^{c_0,c_1}, s \geq 0\right\}$. Of particular interest is the convergence of such a process to a telegraphic (wave distribution) equation. Consider the following equation:

$$u(x,t) = \frac{1}{2}\left\{f(x+at) + f(x-at)\right\}, x \in \mathbb{R}, t \geq 0$$

Such a distribution is defined as the sole solution of the wave partial differential equation:

$$\frac{\partial^2 u(x,t)}{\partial t^2} = a^2 \frac{\partial^2 u(x,t)}{\partial x^2} \quad \text{with} \quad \mathrm{u}(x,0) = f(x) \quad \text{and} \quad \frac{\partial u(x,0)}{\partial t} = 0$$

If we set $c_0 = c_1 = c$ then $Z_t^{c,c} = \int_0^t (-1)^{N_u^c} du$ with N_u^c a Poisson process and $Z_t^{c,c}$ defining a sum of $+1$ and -1, occurring with the same inter-event exponential probability distribution. In this particular case, since $\left\{N_s^{c_0,c_1}, Z_s^{c_0,c_1}\right\}$ is a Markov process, the function $w(t,x) = \mathbf{E}\left\{u(x, Z_s^{c,c})\right\}$, satisfies the partial differential equation:

$$\frac{\partial^2 w(x,t)}{\partial t^2} + 2c\frac{\partial w(x,t)}{\partial t} = a^2 \frac{\partial^2 w(x,t)}{\partial x^2} \quad \text{with} \quad w(x,0) = f(x) \quad \text{and} \quad \frac{\partial u(x,0)}{\partial t} = 0$$

which is a Telegraphic equation.

When $c_0 \neq c_1$ a general result called the Integrated Telegraphic Noise (ITN) by Herrmann and Vallois 2010, provides the probability distribution of $Z_t^{c_0,c_1}$, given in terms of modified Bessel functions:

$$I_\lambda(\xi) = \sum_{i \geq 0} \frac{(\xi/2)^{\lambda+2i}}{i!\Gamma(\lambda+i+1)}$$

Letting,

$$f(t,x) = \frac{1}{2}\left[\sqrt{\frac{c_0 c_1 (t+x)}{t-x}} I_1\left(\sqrt{c_0 c_1 (t^2 - x^2)}\right) + c_0 I_o\left(\sqrt{c_0 c_1 (t^2 - x^2)}\right)\right] e^{(c_1 - c_0)x}$$

We obtain the probability distribution:

$$\mathbf{P}(Z_t^{c_0,c_1} \in dx) = e^{-c_0 t}\delta_t(dx) + e^{-(c_0+c_1)} f(t,x)\mathbf{1}_{[-t,t]}(x)$$

where $\delta_t(dx)$ is a Dirac-Delta function.

This particular Short memory case important as it provides a rational framework for random disturbances that are "cyclical" in the sense that a system may be operating under a certain model (say $+1$) for a random time, switch then following some Jump probability model) to a another model (say -1) and then return at a later random time to the model $+1$. The current application to this model to financial problems is in its infancy however.

Acknowledgement The authors are grateful to G. Jumarie (University of Quebec in Montreal), to O. Tapiero (University of Paris I-Pantheon Sorbonne) and to S. Herrmann for the right they gave us to reproduce some of their results in this papers.

References

Atkinson R. C., and R. M. Shiffrin. 1968. Human memory: A proposed system and its control process. In *The psychology of learning and motivation,* ed. K. W. Spence and J. T. Spence, 2 vols, 89–195. New York: Academic Press.

Bachelier, L. 1900. Théorie de la speculation. *Annales Scientifiques de L'École Normale Supérieure* 17:21–86 (English translation by A. J. Boness in Cootner, P. H. (ed.). 1964. *The random character of stock market prices*. 17–75. Cambridge: MIT Press).

Bachelier, L. 1901. Théorie mathématique du jeu. *Annales de L'École Normale Supèrieure* 18: 143–210.

Baddeley, A. 1994. The magical number 7after all these years. *Psychological Review* 101 (2): 353–356.

Baddeley, A. 2003. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* 4 (10): 829–839.

Balinth, Toth. 1986. Persistent random walks in random environments. *Probability Theory and Related Fields* 71:615–625.

Barkai, E.. 2001. Fractional Fokker-Planck equation, solution, and application. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics* 63:046118.

Beghin, L., and E. Orsingher. 2009. Fractional Poisson processes and related planar random motions. *Electronic Journal of Probability* 14 (61): 1790–1827.

Beghin L., and E. Orsingher. 2010. Poisson-type processes governed by fractional and higher-order recursive differential equations. *Electronic Journal of Probability* 15 (22): 684–709.

Belair, J. 1987. Sur le calcul de la dimension fractale. *Ann. Sci. Math* 11 (I): 7–13.

Ben Avraham, D., and S. Havlin. 2000. *Diffusion and reactions in fractals and disordered systems*. Cambridge: Cambridge University Press.

Beran, J. 1992. Statistical methods for data with long-range dependence. *Statistical Science* 7:404427.

Bollerslev, T. 1986. Generalized autoregresive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.

Boltzmann, L. (1878). Zur theorie der elastischen nachwirkung. *Annalen der Physik* 241 (11): 430–432.

Borges E, P., and I. Roditi. September 1998. A family of non-extensive entropies. *Physics Letters, A* 246:399–402.

Bunde A., and S. Havlin. 1991. *Fractals and disordered systems*. Berlin: Springer.

Cahoy, D. O., and F. Polito. 2013. Renewal processes based on generalized Mittag-Leffler waiting times. *Communications in Nonlinear Science and Numerical Simulation* 18 (3): 639–650.

Cahoy, D. O., V. V. Uchaikin, and W. A. Woyczynski. 2010. Parameter estimation for fractional Poisson processes. *Journal of Statistical Planning and Inference* 140 (11): 3106–3120.

Claes I., and C. Van den Broeck. October 1987. Random walk with persistence. *Journal of Statistical Physics* 49 (1–2): 383–392.

Cont, R., and P. Tankov. 2004. *Financial modelling with jump processes*. London: Chapman & Hall/CRC.

Cresson, J. C., M. A. Alves da Silva, and G. M. Viswanathan. 2007. Amnestically induced persistence in random walks. *Physical review letters* 98:070603.

Daley, D. J. 1999. The Hurst index for a long-range dependent renewal processes. *Annals of Probablity* 27 (4): 2035–2041.

Damien, Campbell. www.cetcapital.com.

Duncan, T. E., Y. Hu, and B. Pasik-Duncan. 2000. Stochastic calculus for fractional Brownian motion, I. Theory. *SIAM Journal on Control and Optimization* 38:582–612.

Einstein, A.,1905. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Annalen der* Physik 17:549–560.

Engle, R. 1987. Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation. *Econometrica* 50:987–1008.

Fama, Eugene. 1963. Mandelbrot and the stable paretian hypothesis. *Journal of Business* 36 (4): 420–429 (reprinted in Paul Cootner (ed.). 1964. The random character of stock prices. MIT Press).

Fama, Eugene. 1965a. The Behavior of stock market prices. *Journal of Business* 38:34–105.

Fama, Eugene. 1965b. Portfolio analysis in a stable Paretian market. *Management Science* 11: 404–419.

Fama, Eugene. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25:383–473.

Feller, W.. 1951. The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics* 22:427–432.

Feller, W. 1957, 1966. *An introduction to probability theory and its applications*. 1 and 2 Vols. New York: John Wiley and Sons, Inc.

Ferguson, A., and M. Z. Bazant. 2005. Lecture 19: poly models: Persistence and self-avoidance, MIT Lecture Notes.

Fitouhi, A., et al. 2005. Asymptotic approximations in quantum calculus. *Journal of Nonlinear Mathematical Physics* 12:586–606.

Fox, R., and M. S. Taqqu. 1985. Non-central limit theorems for quadratic forms in random variables having long range dependence. *The Annals of Probability* 13:428–446.

Goldstein, S. 1951. On diffusion by discontinuous movements, and on the telegraph equation. *The Quarterly Journal of Mechanics and Applied Mathematics* 4:129–156.

Graf, H. P. 1983. Long-range correlations and estimation of the self-similarity parameter. PhD Diss., ETH, Zurich.

Granger, C. W., and J. Joyeux. 1980. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1 (1): 15–29.

Gray, H. L., N. R. Zhang, and W. A. Woodward. 1989. On generalized fractional processes. *Journal of Time Series Analysis* 10:233–257.

Herbin, E., and E. Merzbach. 2006. A set-indexed fractional Brownian motion. *Journal of Theoretical Probability* 19 (2): 337–364.

Herrmann, S. and P. Vallois. 2010. From persistent random walk to the telegraph noise. *Stochastics and Dynamics* 10 (2): 161–196.

Hu, Y., and B. Øksendal. 2003. Fractional white noise calculus and applications to finance. *Infinite Dimensiona Analysis, Quantum Probability and Related. Topics* 6:1–32.

Huillet, T. 2002. Renewal processes and the Hurst effect. *Journal of Physics A* 35:4395–4413.

Hurst, H. E. 1951. Long terms storage of reservoirs. *Transactions of the American Society of Civil Engineers* 116:770–779.

Imhoff, J. P. 1985. On the range of brownian motion and its inverse process. *The Annals of Probility* 13 (3): 1011–1017.

Imhoff, J. P. 1992. A construction of the brownian motion path from BES (3) pieces. *Stochastic Processes and Applications* 43:345–353

Ivanov, G., and E. Merzbach. 2000. *Set-indexed Martingales*. Chapman & Hall, London.

Jackson, F. H. 1910. On a q-definite integrals. *Quarterly Journal of Pure and Applied Mathematics* 41:193–203.

Jain, N.C., and S. Orey. 1968. On the range of random walk. *Israel Journal of Mathematics* 6:373–380.

Jain, N. C., and W. E. Pruitt. 1972. The range of random walk. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability,*3 Vols, ed. Lucien Le Cam, Jerzy Neyman, and Elizabeth L. Scott, 31–50. Berkeley: University of California Press.

Jumarie, G. 1993. Stochastic differential equations with fractional Brownian motion input. *International Journal of Systems Science* 6:1113–1132.

Jumarie, G. 2004. Fractional Brownian motions via random walk in the complex plane and via fractional derivative. Comparison and further results on their Fokker–Planck equations. *Chaos, Solitons & Fractals* 4:907–925.

Jumarie, G. 2005. On the representation of fractional Brownian motion as an integral with respect to (dt)a. *Applied Mathematics Letters* 18:739–748.

Jumarie, G. 2005. On the solution of the stochastic differential equation of exponential growth driven by fractional Brownian motion. *Applied Mathematics Letters* 18:817–826.

Jumarie, G. 2006. Modified Riemann–Liouville derivative and fractional Taylor's series of non-differentiable functions further results. *Computers and Mathematics with Applications* 51: 1367–1376.

Jumarie, Guy. 2008. Stock exchange fractional dynamics defined as a fractional exponential growth driven by (usual) Gaussian white noise. Applications to fractional Black-Scholes equations. *Insurance Mathematics and. Economics* 42:271–287.

Jumarie, Guy. 2010. Fractional multiple birth-death processes with birth probabilities. *Journal of the Franklin Institute* 347:1797–1813.

Jumarie, Guy. 2012a. Deviation of an amplitude of information in the setting of a new family of fractional entropies. *Information Sciences* 216:113–137.

Jumarie, Guy. 2012b. On the fractional solution of the equation $f(x + y) = f(x)f(y)$ and its application to fractional Laplace's transform. *Applied Mathematics and Computation* 219:1625–1643.

Jumarie, Guy. 2013. *Fractional differential calculus for non-differentiable functions*. Saarbrücken: Lap Lambert Academic Publishing.

Kac, V., and P. Cheung. 2002. *Quantum calculus*. New York: Springer.

Laskin, Nick. 2000. Fractional market dynamics. *Physica A* 287:482–492.

Laskin, N. 2003. Fractional Poisson process. *Communications in nonlinear science and numerical simulations* 8 (3–4): 201–213.

Laskin, N. 2009. Some applications of the fractional Poisson probability distribution. *Journal of Mathematical Physics* 50:113513. (12 Pages).

Laskin, Nick. 2011. Some applications of the fractional poisson probability distribution, arXiv1193v2 [math-ph] 17 No.

Leonenko, N., and E. Merzbach. 2013. Fractional Poisson fields. *Methodology and Computing in Applied Probability*. doi:10.1007/s11009-013-9354-7

Leonenko, N. N., M. M. Meerschaert, and A. Sikorskii. 2013a. Fractional Pearson diffusions. *Journal of Mathematical Analysis and Applications* 403:532–546.

Leonenko, N. N., M. M. Meerschaert, and A. Sikorskii. 2013b. Correlation structure of fractional Pearson diffusions. *Computers and Mathematics and Applications* 66:737–745. doi:10.1016/j.camwa.2013.01.009.

Liouville, J. 1832. Sur le calcul des differentielles a indices quelconques. *Journal of Ecole Polytechnique* 13:71.

Mandelbrot, B. B. 1982. *The Fractal Geometry of Nature*. San Francisco: Freeman.

Mandelbrot, B., and M. Taqqu. 1979. Robust R/S analysis of long run serial correlation. *Bulletin of the International Statistical Institute* 48 (Book 2): 59–104.

Mandelbrot, B., and J. Van Ness. 1968. Fractional Brownian motion, fractional noises and applications. *SIAM Review* 10:422–437.

Masoliver, J., J. M. PorrVa, and G. H. Weiss. 1993. Some two and three-dimensional persistent random walks. *Physica A: Statistical Mechanics and its Applications* 193:469.

Masoliver, J., M. Montero, and G. H. Weiss. 2003. Continuous time random walk model for financial distributions. *Physics Review E* 67:1–9.

Masolivera, Jaume, Miquel Monteroa, Josep Perello, and George H. Weiss. 2007. The CTRW in finance: Direct and inverse problems with some generalizations and extensions. *Physica A: Statistical Mechanics and its Applications* 379:151–167.

Miller, G. A. 1956. The magical number 7. *The Psychological Review* 63:81–97.

Miller, K. S., and B. Ross. 1973. *An Introduction to the fractional calculus and fractional differential equations*. New York: Wiley.

Mittag-Leffler, G. M. 1903, 1904, 1905. CR Acad Sci Paris 1903;137(2):554–558; R Acad dei Lincei, Rendiconti 1904;13 (5): 3–5; Acta Math 1905;29:101–182.

Montroll, E. W., and G. H. Weiss. 1965. Random walks on lattices, II. *Journal of Mathematical Physics* 6:167–181.

Montroll, E. W., and B. J. West. 1979. On an enriched collection of stochastic processes. In *Fluctuation phenomena,* ed. E. W. Montroll and J. Leibowitz, 61–175. Amsterdam: North-Holland.

Oldham, K. B., and J. Spanier. 1974. *The fractional calculus*. New York: Academic Press.

Osler, T. J. 1971. Taylor's series generalized for fractional derivatives and applications. *SIAM Journal on Mathematical Anaysis* 2 (1): 37–47.

Patlak, Clifford.1953. Random walk with persistence and external bias. *Bulletin of Mathematical Biology* 15 (3): 311–338.

Peter Edgar, E. 1995. *Chaos and order in capital markets*. New York: Wiley.

Picinbono, B., and C. Bendjaballah. 2006. Poisson processes with integrable density. *IEEE Transactions on Information Theory* 52:5606–5613.

Pottier, Noelle. 1996. Analytic study of the effect of persistence on a one-dimensional biased random walk. *Physica A: Statistical Mechanics and its Applications* 230:563–576.

Rieger H., and F. Igloi. 1999. Average persistence in random walks. *Europhysics Letters* 45: 673–679.

Rostek, S., and R. Schöbel. 2013. A note on the use of fractional Brownian motion for financial modeling. *Economic Modelling* 30:30–35.

Samko, S. G., A. A. Kilbas, and O. I. Marichev. 1987. *Fractional integrals and derivatives and their applications* (translated by Harwood Academic). Minsk: Nauka I Tekhnika.

Samorodnitsky, G., and M. S. Taqqu. 1994. *Stable non-gaussian, random processes: Stochastic models with infinite variance*, Stochastic Modeling Series. 1 Vols. New York: Chapman & Hall.

Scalas, Enrico. 2006. The application of continuous time random walks in finance and economics. *Physica A: Statistical Mechanics and its Applications* 362:225–239.

Tapiero, C. S. 1975. Random walk models of advertising, their diffusion approximations and hypothesis, testing. *Annals of Economics and Social Measurement* 4:293–309.

Tapiero, C. S. 1978. Optimal advertising and goodwill under uncertainty. *Operations Research* 26 (3): 450–463.

Tapiero, C. S. 1979. A Generalization of the Nerlove-Arrow model to multi firms advertising under uncertainty. *Management Science* 25:907–915.

Tapiero, C. S. 1982. Stochastic model of consumer behavior and optimal advertising. *Management Science* 28:1054–1064.

Tapiero, C. S. 1983. Stochastic diffusion models with advertising and word-of-mouth effects. *European Journal of Operational Research* 12 (4): 348–356.

Tapiero, C. S. 1988. *Applied stochastic models and control in management.* New York: North Holland.

Tapiero, C. S. 2004. *Risk and financial management: Mathematical and computational methods*. London: Wiley.

Tapiero, C. S. 2005. Advertising and advertising claims over time. In *Optimal control and dynamic games, applications finance, management science and economics,* ed. C. Deissenberg and R. F. Hartl. Springer, Berlin, New York.

Tapiero, C. S. 2010. *Risk finance and assets pricing*. New York: Wiley.

Tapiero, C. S. 2012a. *Engineering risk and finance*. New York: Springer.

Tapiero, Oren. 2012b. *Finance and incomplete state preferences uncertainty: Measurement and implications*. Israel: Doctoral Dissertation, Bar Ilan University.

Tapiero, Oren. 2013a. Financial decisions and q-calculus. *Risk and Decision Analysis* 4 (4): 291–301.

Tapiero, Oren. 2013b. The economics of uncertainty. In *Engineering risks and finance,* ed. C. S. Tapiero. Springer, New York.

Tapiero, Oren. 2013c. A maximum (non-extensive) entropy approach to equity options bid-ask spread. *Physica A: Statistical Mechanics and its Applications* 392:3051–3060.

Tapiero, Oren. 2013d. The relationship between risk and incomplete states uncertainty: A Tsallis entropy perspective. *Journal of Algorithmic Finance* 2:141–150.

Tapiero, C. S., and P. Vallois. 1996. Run length statistics and the Hurst exponent in random and birth-death random walks. *Chaos, Solitons and Fractals* 7 (9):1333–1341.

Taqqu, M. S. 1986. A bibliographical guide of self-similar processes and long-range dependence. In *Dependence in probability and statistics: A survey of recent results,* ed. Ernst Eberlein and Murad S. Taqqu. Basel: Birkhauser.

Taqqu, M. S. 2003. Fractional Brownian motion and long-range dependence. In *Theory and applications of long-range dependence*, Birkhäuser, ed. P. Doukhan, G. Oppenheim, and M. S. Taqqu, 5–38. Boston: Springer.

Telesca, L., and M. Lovallo. 2006. Are global terrorist attacks time correlated. *Physica A: Statistical Mechanics and its Applications* 362:480–484.

Tsallis, C. 1988. Possible generalization of Boltzmann—Gibbs statistics. *Journal of Statistical Physics* 52 (1): 479–487.

Tsallis, C. 2009. Nonadditive entropy and nonextensive statisticsal mechanics-an over view after 20 years. *Brazilian Journal of Physics* 39:337–356.

Tsallis, C., C. Anteneodo, L. Borland, and R. Osorio. 2003. Nonextensive statistical mechanics and economics. *Physica A: Statistical Mechanics and its Applications* 324 (1–2): 89–100.

Vallois, P. 1993. Diffusion arrê tée au premier instant où le processus de l'amplitude atteint un niveau donné. *Stochastics and Stochastic Reports* 43:93–115.

Vallois, P. 1995. On the range process of a Bernoulli random walk. In *Proceedings of the sixth international symposium on applied stochastic models and data analysis*. II Vols., ed., J. Janssen and C.H. Skiadas, 1020–1031. World Scientific.

Vallois, P. 1996. The range of a simple random walk on Z. *Advances in Applied Probability* 28: 1014–1033.

Vallois P., and C. S. Tapiero. 1995. Moments of an amplitude process in a random walk. *Recherche Operationnelle/Operation Research (RAIRO)* 29 (1): 1–17.

Vallois P., and C. S. Tapiero. 1996. The range process in random walks: Theoretical results and applications. In *Advances in computational economics*, ed. H. Ammans, B. Rustem, and A. Whinston. Springer, Berlin.

Vallois P., and C. S. Tapiero. 1997. Range reliability in random walks. *Mathematics of Operations Research* 45 (3): 325–345. (Zeitchrifts fur Operations Research).

Vallois P., and C. S. Tapiero. 2007. Memory-based persistence in a counting random walk process. *Physica A: Statistical Mechanics and its Applications* 386:303–317.

Vallois P., and C. S. Tapiero. 2008. A claims persistence process and insurance. *Insurance: Economics and Mathematics* 44:367–373.

Viano, M. C., C. Deniau, and G. Oppenheim. 1994. Continuous time fractional ARMA processes. *Statistics and Probability Letters* 21:323–336.

Viswanathan, G. M., S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. Reposo, and H. E. Stanley. 1999. Optimizing the success of random searches. *Nature* 401:911–914.

Wang, X. T., and Z.-X. Wen. 2003. Poisson fractional processes. *Chaos Solitons and Fractals* 18 (1): 169–177.

Wang, X. T., Z.-X. Wen, and S.-Y. Zhang. 2006. Fractional Poisson process II. *Chaos, Solitons & Fractals* 28 (1): 143–147.

Wang, X. T., S.-Y. Zhang, and S. Fan. 2007. Nonhomogeneous fractional Poisson processes. *Chaos, Solitons & Fractals* 31 (1): 236–241.

Weiss, G. H. 1994. *Aspects and applications of the random walk*. Amsterdam: North-Holland.

Weiss, G. H. 2002. Some applications of persistent random walks and the telegrapher's equation. *Physica A: Statistical Mechanics and its Applications* 311:381–410.

Weiss, G. H., and R. J. Rubin. 1983. Random walks: Theory and selected applications. *Advances in Chemical Physics* 52:363–505.

Willinger, W., and V. Paxson. 1998. Where mathematics meets the internet. *Notices of the AMS* 45:961–970.

Wolpert, R. L., and M. S. Taqqu. 2005. Fractional Ornstein-Uhlenbeck Levy processes and the Telecom process: Upstairs and downstrais. *Signal Processing* 85:1523–1545.

Wu, H., B. L. Li, T. A. Springer, and W. H. Neil. 2000. Modelling animal movement as a persistent random walk in two dimensions: expected magnitude of net displacement. *Ecological Modelling* 132:115.

Asset Price Modeling: From Fractional to Multifractional Processes

Sergio Bianchi and Augusto Pianese

1 Introduction and Motivation

The 2007–2009 crisis has re-awakened the interest in modeling financial assets and their prices among academics and practitioners, and increased the awareness of the limits that the standard financial paradigm shows in describing real world data. A large number of research contributions over the last quarter century combined with overwhelming financial crises have provided ample evidence that financial markets are not always complete. In particular, Brownian based stochastic processes underlying fundamental models of finance were shown to fail by several magnitudes when predicting the shocks that financial markets have been subjected to (Tapiero 2010; Tapiero et al. 2013; Tapiero et al. 1996). By undermining the bricks of the efficient markets model, these failures have motivated the research of models that seek to account for the complexity and the anomalies of financial markets (see Yen 2008 and Tapiero 2007 for a survey). Cont 2001 for example summarizes some of the most relevant stylized facts, including:

- the absence of autocorrelations in the log-price variations and the slow decay of autocorrelation in absolute (or squared) returns;
- the asymmetric behavior of stock prices that produces large and sudden drawdowns, but only slow upward movements;
- the volatility clustering, meaning that high-volatile periods tend to cluster in time and to generate, in this way, the so called intermittency;
- the presence of heavy tails (mostly, with a tail index between two and five) in the unconditional distributions of returns and the presence of conditional heavy tails

S. Bianchi (✉)
Department of Economics and Law, University of Cassino,
Cassino, Italy
e-mail: sbianchi@eco.unicas.it

A. Pianese
Department of Finance and Risk Engineering, New York University,
New York, USA

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,*
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_7,

(smaller with respect to the unconditional distributions) even for the residuals obtained by correcting returns for volatility clustering;

- the aggregate Gaussianity, meaning that distributions tend to the Gaussian law as one increases the time scale used to calculate the returns;
- the correlation between volatility and traded volumes;
- an asymmetry in time scales, meaning that fine-scale measures of volatility predict coarse-grained volatility worse than the other way round.

The considerations above suggest that the Efficient Market Hypothesis (EMH) – that is the most influential idea of the modern investment theory stating that market prices fully and instantaneously reflect all the available information (Fama 1970)—is not always met in practice. In this regard, it is not by chance that the *Behavioral Finance* (BF) is gaining consensus (Sewell 2007). Nonetheless, while it is widely recognized that behavioral biases and psychology of market participants can intervene hither and thither in decision making, it is still unclear to what extent they affect trading and investment management (Lo 2005). To provide a rationale for behavioral case studies and heuristics, (Lo 2004, 2012) proposed an *Adaptive Market Hypothesis* (AMH) based on concepts of evolutionary biology such as competition, mutation, reproduction, and natural selection, as a synthesis between behavioral constructs and the EMH. The combination of these forces determines then the efficiency of markets and is ultimately responsible for the waxing and waning of finance. The EMH is thus the "frictionless ideal" recovered once all market imperfections, both technical and psychological, are accounted for.

Beyond the plausibility of this interpretation, the BF and the AMH are not sufficiently developed and formalized to represent well-established paradigms. Indeed, if it is true that behavioral versions of some basic financial schemes were proposed (see e.g. Kahneman and Tversky 1979 for the utility theory, Shefrin and Statman 2000 for the portfolio theory, Shefrin and Statman 1994 for the C.A.P.M., or Shefrin and Thaler 1988 for the Life Cycle Hypothesis), such models met a limited diffusion among academics and practitioners. Furthermore, most of them succeed in explaining only partially the complexity embedded in real data, and therefore their pros are judged not substantial enough when compared with the cons affecting the current paradigm.

In order to overcome this limit, it is useful to step back and start again from the definition of efficiency; this can be linked very parsimoniously to the new and promising analytical financial models that will be discussed in this chapter.

Thus, the notion of efficiency suggests that the price S_t of an individual stock discounts all the information $\mathcal{F}_t$ accumulated up to time t, as a consequence of the quick and wide spread of news. This assumption is traditionally introduced and tested in terms of the expected value of properly discounted payoffs (Fama 1970). With respect to the filtered probability space $(\Omega, \mathcal{F}_t, (\mathcal{F})_{0\le t\le T}, \mathbb{P})$, the condition requires that for $t < \tau < T$

$$S_t = \mathbb{E}_t\left(Y_{t,\tau} X_\tau\right) \tag{1}$$

or, equivalently, that

$$\mathbb{E}_t\left(Y_{t,\tau}\frac{X_\tau}{S_t}\right) = 1, \tag{2}$$

where as usual $\mathbb{E}_t(\,\cdot\,)$ is the short notation for the conditional expectation $\mathbb{E}(\,\cdot\,|\mathcal{F}_t)$. In words, the current price of a financial asset equals the conditional expectation of its payoff X_τ discounted by the stochastic discount factor $Y_{t,\tau}$[1] that accounts for the dynamics of risk premia. Equation (1) can be easily written in terms of excess returns with respect to the risk-free rate $r_{t,\tau}$ as

$$\mathbb{E}_t\left(Y_{t,\tau}\left(\frac{X_\tau}{S_t} - (1+r_{t,\tau})^{\tau-t}\right)\right) = 0.$$

For $\tau = t+1$, setting $R^*_{t,t+1} = \frac{X_{t+1}-S_t}{S_t} - r_{t,t+1}$, the equation above turns to

$$\mathbb{E}_t\left(Y_{t,t+1}\cdot R^*_{t,t+1}\right) = 0, \tag{3}$$

stating that the conditional expected excess returns equals zero. Usually, efficiency is tested through (3), but since the model-dependent process that generates the risk-premium is not observable, the EMH can be ultimately tested only jointly with a model providing $Y_{t,\tau}$. Since Eq. (3) implies

$$\mathbb{E}_t\left(R^*_{t,t+1}\right) = -\left(1+r_{t,t+1}\right)\cdot \mathrm{Cov}\left(R^*_{t,t+1}, Y_{t,t+1}\right) \tag{4}$$

by itself the predictability of returns (i.e. the controversial failure of the random walk model) does not prove market inefficiency, since it suffices the expected conditional return to comply with (3) in order to save both efficiency and non-random walk models.

Nonetheless, the exploitation of the above relation to argue about the behavior of financial time series is somewhat controversial: while the first empirical evidence of predictability was ascribed to market inefficiency under the assumption of constant expected returns (Shiller et al. 1984; Summers 1986), further studies proposed the time-varying expected returns as an alternative explanation (Fama and French, 1988). In turn, these can be generated by time-varying risk aversion (Campbell and Cochrane 1999), long-run consumption risk (Bansal and Yaron 2004, or time-variation in risk-sharing opportunities (Lustig et al. 2005).

To say it shortly, since empirical evidence suggests that the intensity of dependence in financial time series actually changes over time, new insights are likely to come from framing the notion of efficiency in a dynamical perspective, which means to include market inefficiencies into the models rather than considering them as pathological outliers. To this aim, several models have been proposed, relaxing the hypothesis of independence or the assumption of identical distribution of the log price variations.

[1] Remind that the random process $Y_{t,\tau}$, $\tau > t$, is a stochastic discount factor (or pricing kernel) if (a) $\mathbb{P}(Y_{t,t} = 1, Y_{t,\tau} > 0) = 1$, and (b) $S_t = \mathbb{E}^{\mathbb{P}}_t[Y_{t,\tau}S_\tau]$.

In the standard Gaussian framework, stationarity requires that the process pointwise regularity, quantified by the Hölder exponent[2], be time-invariant; at the same time, independence means assuming that the Hölder parameter is almost everywhere equal to $\frac{1}{2}$.

The model that will be discussed in this chapter is defined based on processes that relax both the assumptions (independence and identical distributions), by acting on the process pointwise regularity. This is described in detail in paragraph 3. Such stochastic processes—named *multifractional Brownian motion* (mBm) (see Benassi et al. 1997; Péltier and Lévy Véhel 1995) and *multifractional Processes with Random Exponent* (MPRE) (see Ayache and Taqqu 2005)—generalize the celebrated *fractional Brownian motion* (fBm) (Kolmogorov 1940; Mandelbrot and Van Ness 1968). In turn, fBm extended the Brownian motion by replacing its scaling factor $\frac{1}{2}$ by the parameter $H \in (0, 1)$, that rules the dependence of its increments and quantifies the pointwise and global regularity. Despite their large modeling flexibility, mBm and MPRE are still largely unknown in finance, mostly because of the prejudice due to the fact that fBm admits arbitrage when $H \neq \frac{1}{2}$ (Sewell 1997) and thus violates a fundamental hypothesis of complete markets. The word *prejudice* is not out of place; in fact, while the proof of the existence of arbitrage opportunities was deduced under the constancy of the parameter H when it differs from $\frac{1}{2}$, no results are available yet for mBm and MPRE, whose functional parameter changes over time in a deterministic (mBm) or stochastic (MPRE) way. In addition, if one properly restricts the class of admissible strategies, even when $H \neq \frac{1}{2}$ the fBm fulfills the condition of No Free Lunch with Vanishing Risk (Cheridito 2003; Jarrow and Sayit 2009). Aside the arbitrage argument, the adoption of an mBm/MPRE-based modeling also augments the difficulties to infer global probabilistic properties. Despite this state of affairs, they are very promising stochastic processes for several reasons:

- a proper choice of the functional parameter renders the mBm/MPRE able to reproduce stylized facts. Some of these features will be discussed: the absence of autocorrelation in its increments and significant autocorrelation in its absolute/squared increments, the unconditional and the conditional heavy tails, the gain/loss asymmetry, the intermittency and the volatility clustering;
- the time varying Hölder exponent of mBm/MPRE measures the pointwise regularity of the process paths, but can also be interpreted as the pointwise intensity of memory. The very fact that it changes over time can provide an explanation for the many seemingly inconsistent estimates of the long-range dependence parameter (see Baillie 1996 and Henry and Zaffaroni 2003 for a survey). In fact, as noted in (Bianchi and Pianese 2008), it can be easily shown that detecting dependence or not using asymptotic estimators strongly relies on the segment of data one looks at, what basically indicates nonstationarity;
- it is not only about multifractional processes ability to replicate financial dynamic processes. Importantly, these processes are flexible providing a rationale to the

[2] The definition of Hölder exponent will be given in paragraph 2; here, we just recall that it measures the degree of irregularity of the graph of a function.

market mechanisms, since they potentially can gather multifaceted behaviors of investors' trading (in particular their reactions to financial crashes). The financial intuition for these processes will be presented in paragraph 4, while their empirical validation will be discussed in paragraph 6;
- the mBm/MPRE can be interpreted as the quantitative counterpart of qualitative models such as the BF or the AMH, in the sense that it allows to assess analytically how far markets are from efficiency at any time t. In this way, the mBm/MPRE could serve as models for the development of a new financial calculus.

Due to these reasons, the starting point of a fitting multifractional modeling relies on a reliable estimate of the functional parameter of the mBm/MPRE. This task has attracted in the last fifteen years many contributions in diverse fields. In paragraph 5 an estimator will be described which allows a reliable and timely reckoning of the pointwise regularity.

Finally, it is appropriate to underline that, in spite of their semantic affinity, multifractional processes should not be confused with the more popular multifractal models, studied for example by (Calvet et al. 1997; Arneodo 1998; Riedi 2002). In fact, multifractal models are mostly based on measures deforming the calendar (or physical) time; therefore, by definition, they belong to the family of time-change processes. Unlike these processes, the mBm is not in general a multifractal process. Ayache 2000 provides technical conditions under which the Generalized multifractional Brownian motion (which extends the mBm) can be multifractal; an analysis of the use and the limits in finance of multifractality detection techniques is proposed in (Bianchi 2005) and (Bianchi and Pianese 2007).

2 Pointwise, Local and Uniform Hölder Regularity

In the previous paragraph we have pointed out that, if the expected conditional return complies with (4), predictability of returns by itself does not prove market inefficiency. This implies that a stochastic financial time series ought to be characterized in some other manner, namely by using a process *regularity* and its link with the semimartingale condition of no arbitrage. For a discussion of this notion with respect to stochastic processes we refer to (Ayache 2013) (reference that we will follow in the definitions below), while for the general case of a function f, see (Kolwankar, K. and Lévy Véhel 2002).

Let $X(t,\omega)$ be a stochastic process with a.s. continuous and not differentiable trajectories over the real line $\mathbb{R}$. One has the following

Definition 1 (Uniform Hölder exponent) Let $J \subset \mathbb{R}$ be a non-degenerate compact interval (i.e., not empty nor made of a single point). The global Hölder regularity of the trajectory $t \mapsto X(t,\omega)$ is measured by

$$\beta_X(J,\omega) = \sup\left\{\beta \geq 0 : \sup_{t,s\in J} \frac{|X(t,\omega) - X(s,\omega)|}{|t-s|^{\beta}} < \infty\right\} \tag{5}$$

Definition 2 (Pointwise and Local Hölder Exponent) The local Hölder regularity of the trajectory $t \mapsto X(t,\omega)$ with respect to some fixed point $t \in \mathbb{R}$ can be measured through the *pointwise* or the *local* Hölder exponent, respectively defined as[3]

$$\alpha_X(t,\omega) = \sup\left\{\alpha \geq 0 : \limsup_{h\to 0} \frac{|X(t+h,\omega) - X(t,\omega)|}{|h|^{\alpha}} = 0\right\} \tag{6}$$

$$\tilde{\alpha}_X(t,\omega) = \sup\{\beta_X([a,b],\omega) : a, b \in \mathbb{R}, t \in (a,b)\} \tag{7}$$

It is always

$$\tilde{\alpha}_X(t,\omega) \leq \alpha_X(t,\omega).$$

The intuition for (6), (7), and hence for (5), is sketched in Fig. 1. It provides the geometrical meaning of the pointwise Hölder regularity quantified by the exponent α. Function X has exponent α at t_0 if, around t_0, for any positive ϵ, there exists a neighborhood of t_0, $I(t_0)$, such that, for $t \in I(t_0)$, the graph of X is included in the envelope defined by $t \mapsto X(t_0) - c|t-t_0|^{\alpha-\epsilon}$ and $t \mapsto X(t_0) + c|t-t_0|^{\alpha+\epsilon}$ (Lévy Véhel and Barriére 2008). Hence, the regularity (or smoothness) of the graph increases with α. For certain classes of stochastic processes, remarkably for Gaussian processes, by virtue of zero-one law, both the quantities are deterministic, that is there exist the non random quantities $a_X(t)$ and $b_X(t)$ such that $\mathbb{P}(a_X(t) = \alpha_X(t,\omega)) = 1$ and $\mathbb{P}(b_X(t) = \beta_X(t,\omega)) = 1$ (Ayache 2013). From a practical viewpoint α_X increases with the smoothness of the graph and it can be proved to be equal to $\frac{1}{2}$ when the process is the ordinary Brownian motion.

3 Fractional and Multifractional Gaussian Processes

In the following, three (multi)fractional processes will be discussed. Other variants exist: the step fractional Brownian motion (Benassi 2000), the multiscale fractional Brownian motion (Bertrand 2005), the sparse multifractional Brownian motion (Bertrand 2012), the Generalized multifractional Brownian motion (Ayache and Lévy Véhel 2000) (just to limit the discussion to Brownian models). Nonetheless, we start from the miliar stone representation of the fractional Brownian motion and focus on two processes-the multifractional Brownian motion (mBm) and the Multifractional Processes with Random Exponent (MPRE)-which offer an immediate financial interpretation. Before discussing their peculiarities, it is worthwhile to underline that, in spite of the assonance, these processes should not be confused with

[3] The distinction between these two ways of measuring the local regularity is not just a mathematical detail, since-as proved in (Ayache 2013)-while the trajectories of an arbitrary centered Gaussian process share the same global (uniform) as well as local Hölder regularity, the same property does not hold with respect to the pointwise Hölder regularity.

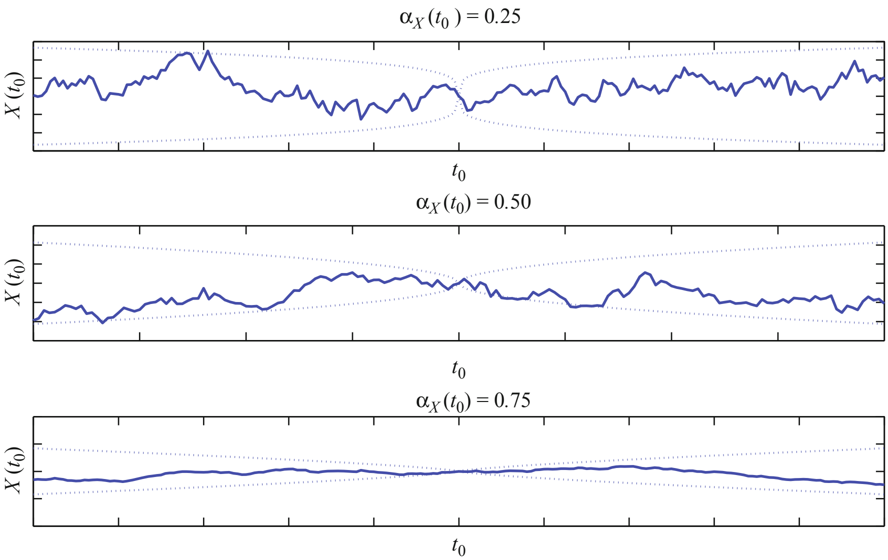


Fig. 1 Hölder regularity of the graph of a random function

the more popular multifractal models (see Calvet and Fischer 2002; Calvet and Fisher 2008; Arneodo 1998, Riedi 2002), whose use (and abuse) in finance is analyzed in (Bianchi 2005; Bianchi and Pianese 2007). In fact, while the former originate from the variability of the scale, the latter are defined in terms of the variability of the time: multifractal processes mostly deform the calendar (or physical) time, which in finance is meant to account for the time-changing number of transactions. In this sense, by definition, they belong to the family of time-changed processes.Generally, the two stochastic processes we are going to discuss are not multifractal (Ayache 2000 provides the technical conditions under which a further extension of the multifractional Brownian motion-the Generalized multifractional Brownian motion-can be multifractal).

3.1 Fractional Brownian Motion

The fractional Brownian motion $B_H(t)$ is a centered Gaussian, self-similar[4], continuous stochastic process with almost surely non differentiable sample paths and stationary increments. It was introduced in a seminal paper by Mandelbrot and Van

[4] We remind that the stochastic process $\{X(t,\omega)\}_{t\in T}$ is said self-similar of parameter H if for any $a > 0$ and $t \in T$, it is $\{X(at,\omega)\} \overset{d}{=} \left\{a^H X(t,\omega)\right\}$, where the equality holds for the finite-dimensional distributions of $\{X(t,\omega)\}$.

Ness (Mandelbrot and Van Ness 1968)[5] as a generalization of the Brownian motion obtained by replacing the value $\frac{1}{2}$ by the Hurst exponent $H \in (0, 1)$. Several representations can be given of this process; in particular, the non anticipative moving average representation lends itself to interpret the fBm as a weighted sum of the Brownian measure, in which the weights prescribe the intensity of the dependence the process is endowed with. It is defined by

$$B_H(t) = K V_H \int_{\mathbb{R}} \left((t - s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}} \right) dW(s) \tag{8}$$

where $x_+ = \max(x, 0)$, $V_H = \frac{\Gamma(2H+1)\sin(\pi H)^{\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)}$ is a normalizing factor, $K^2 = Var B_H(1)$ and, as usual, dW denotes the Brownian measure.

Representation (8), up to a multiplicative constant, is equivalent to the following harmonizable one

$$\hat{B}_H(t) = \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{i\xi\, |\xi|^{H-\frac{1}{2}}} d\hat{W}(\xi) \tag{9}$$

where $d\hat{W}$ is the Fourier transform of the Brownian measure dW, i.e. the unique complex-valued stochastic measure such that $\int_{\mathbb{R}} f(x) dW(x) = \int_{\mathbb{R}} \hat{f}(x) d\hat{W}(x)$ for all $f \in L^2(\mathbb{R})$, where $\hat{f}(x) = \int_{\mathbb{R}} e^{-i\xi x} f(x) dx$ is the Fourier transform of f.[6] A relevant property is that $\alpha_{B_H}(t) = \beta_{B_H}(\Omega) = H$, almost surely at any point t. Therefore, the trajectories of an fBm of parameter H display the same regularity (or roughness), whatever the point.

The covariance of the fBm reads as

$$\mathbb{E}\left(B_H(t) B_H(s)\right) = \frac{K^2}{2} \left(|t|^{2H} + |s|^{2H} - |t - s|^{2H} \right) \tag{11}$$

which implies that $\mathbb{E} B_H(t)^2 = \mathrm{Var}\left(B_H(t)\right) = K^2 t^{2H}$.

Unless the normalizing factor K, the variance of the increments is

$$\mathrm{Var}\big(B_H(t+1) - B_H(t)\big) = \mathrm{Var}\big(B_H(1) - B_H(0)\big) = \mathrm{Var} B_H(1)$$

[5] Actually, processes having properties similar to those displayed by the fBm were considered, at least implicitly, by (Kolmogorov 1940; Hunt 1951; Lamperti 1962; Yaglom 1958; Yaglom 1968).

[6] The fBm can be also defined in terms of random wavelet series as

$$\bar{B}(x, H) = \sum_{j=-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} 2^{-jH} \epsilon_{j,k} \left(\Psi(2^j x - k, H) - \Psi(-k, H) \right) \tag{10}$$

where $\left\{\epsilon_{j,k}\right\}_{(j,k) \in \mathbb{Z}^2}$ is a sequence of independent $N(0, 1)$ random variables; $\Psi \in C^\infty(\mathbb{R} \times (0, 1))$ is well-localized in the first variable and uniformly localized in H, which means that, for all $(n, p) \in \mathbb{N}^2$, $\sup\left\{(1 + |x|)^p |(\partial_x^{(n)} \Psi)(x, H)| : (x, H) \in \mathbb{R} \times (0, 1)\right\} < +\infty$. This representation-which allows to define the MPRE (see below) also when its random exponent depends on the Brownian measure-is almost surely uniformly convergent in (x, H) on each compact subset of $\mathbb{R} \times (0, 1)$.

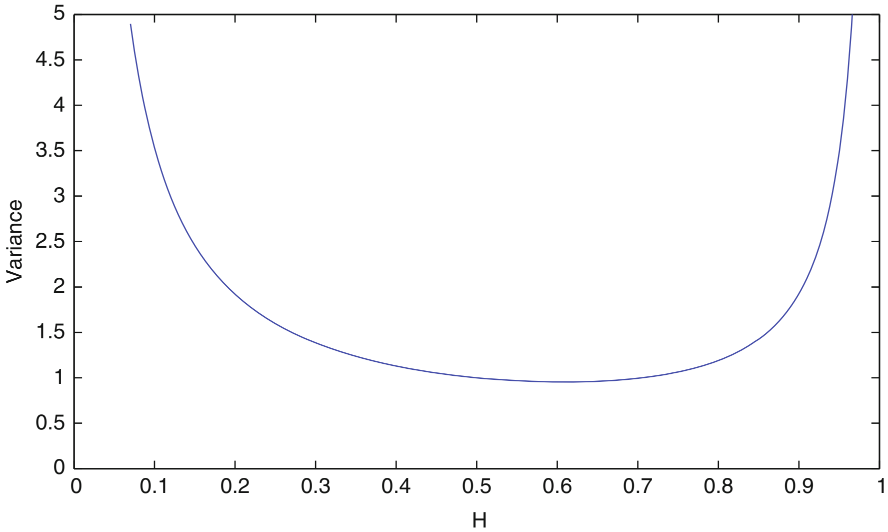


Fig. 2 Variance of the increments of fBm for different values of H

$$= \frac{\Gamma(2-2H)cos(\pi H)}{\pi H(1-2H)}$$

and behaves with respect to H as displayed in Fig. 2 (see Decreusefond and Üstünel 1999).

The stationary increment process $Y_H(t,k) = B_H(t+k) - B_H(t) \stackrel{d}{=} B_H(k) - B_H(0) = Y_H(k)$ has autocovariance function that can be easily deduced by (11). It is given by

$$\begin{aligned} \rho_H(k) = \mathbb{E}\left(Y_H(t)Y_H(t+k)\right) &= \frac{K^2}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right) \\ &= \frac{K^2}{2}\Delta^2|k|^{2H} \end{aligned} \tag{12}$$

where Δ^2 denotes the second difference. Unlike the Brownian motion, the increments of the fBm are correlated and display long range dependence; in fact, when $H > \frac{1}{2}$, $\sum_{k\in\mathbb{Z}} |\rho_H(k)| = +\infty$. Therefore, the autocovariance function (12) indicates that the larger the difference $H - \frac{1}{2}$, the larger the positive (long-range) dependence; the lower the difference $H - \frac{1}{2}$, the higher the negative dependence.

When $H = \frac{1}{2}$ (Brownian motion), the autocovariance is obviously zero whatever the lag. This behaviour reflects in the regularity of the sample paths, whose Hausdorff dimension is almost surely equal to $2 - H$, implying that for $H > \frac{1}{2}$, the paths are smoother than those of the Brownian motion and conversely for $H < \frac{1}{2}$ (see Fig. 3).

This peculiar feature is a source of problems in financial modelling. Indeed, irrespective of the integration theory one chooses, it is well-known that a continuous time market model excludes free lunch with vanishing risk if and only if the price

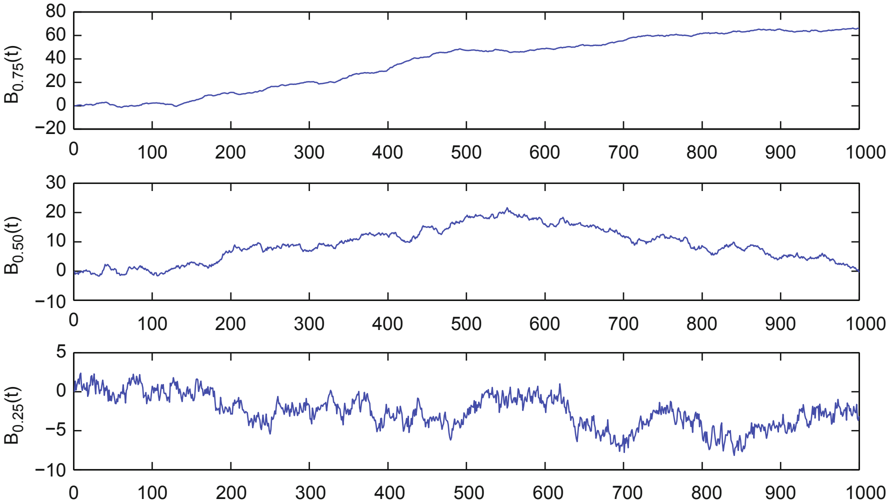


Fig. 3 Examples of fBm for different Hurst exponents

process is a semimartingale (Delbaen and Schachermayer 1994). Unfortunately, the predictability resulting from the autocovariance function implies that the fBm is not a semimartingale when $\alpha_{B_H}(t) = H \neq \frac{1}{2}$ (Rogers 1997). This result readily follows from the behaviour of the order-p variation of the fBm

$$V_{n,p} = \sum_{j=1} 2^n |B_H(j2^{-n}) - B_H((j-1)2^{-n})|^p \sim (2^n)^{1-pH}.$$

As $n \to \infty$, $V_{n,p}$ tends to zero if $p > H^{-1}$ and to infinite if $p < H^{-1}$. Since all semimartingales have well defined quadratic variation, the behaviour of $V_{n,p}$ is consistent with a semimartingale only if $H = \frac{1}{2}$. In fact, if $H > \frac{1}{2}$, one could choose $p \in (H^{-1}, 2)$ in order to let $V_{n,p}$ tend to zero in probability. This would mean that the quadratic variation of the fBm is zero and that B_H is a finite-variation process. But this is not, since for $p \in (1, H^{-1})$, almost surely $\lim_{n\to\infty} V_{n,p}$ is infinite, as well as the order-p variation (because of the scaling). Therefore B_H cannot be finite-variation. In the opposite case, if $H < \frac{1}{2}$, it suffices to choose $p > 2$ (such that $pH < 1$) to see that the order-p variation is infinite, what violates the almost-sure finiteness of the quadratic variation of a semimartingale.

Explicit arbitrage strategies can be found in (Shiryayev 1998; Dasgupta and Kallianpur 2000; Cheridito 2003 or Bende et al. 2007). Although remedies were suggested to correct the fBm to avoid arbitrage[7], it is widely recognized that the process can unlikely represent a good model of financial dynamics.

[7] The corrections mainly concerned the revision of classical definitions of arbitrage and self-financing condition (Hu and Oksendal 2003; Elliott and Van Der Hoek 2003); the regularization

In addition to arbitrage-based objections, the constancy itself of the Hölder regularity seems too restrictive to depict the complexity of financial dynamics, characterized-for example-by volatility clustering and (eventually skewed) non Gaussian unconditional distributions.

3.2 Multifractional Brownian motion

The most immediate generalization of the fBm can be obtained by replacing its exponent H by a proper Hölderian deterministic function of time $h(t)$[8]. This extension, referred to as *multifractional Brownian motion* (see Péltier and Lévy Véhel 1995; Benassi et al. 1997) can describe the dynamics of signals whose regularity changes through time. The cost for the increased flexibility of the model resides in the fact that the increments of the mBm are generally no longer stationary nor self-similar[9]. Once accounted for the function $h(t)$, the non anticipative representation of the mBm becomes

$$X_{h(t)}(t) = K\, V_{h(t)} \int_{\mathbb{R}} \left((t-s)_+^{h(t)-\frac{1}{2}} - (-s)_+^{h(t)-\frac{1}{2}} \right) dW(s), \tag{13}$$

where, as above, $V_{h(t)} = \frac{\sqrt{\Gamma(2h(t)+1) sin(\pi h(t))}}{\Gamma(h(t)+\frac{1}{2})}$ is a normalizing factor.

Using the notation as in the previous case, the harmonizable representation of the mBm is similar to (9)

$$\hat{X}_{h(t)}(t) = \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{i\xi\, |\xi|^{h(t)-\frac{1}{2}}} d\hat{W}(\xi). \tag{14}$$

However, the distributions of $X_{h(t)}(t)$ and $\hat{X}_{h(t)}(t)$ are not properly the same but just nearly the same, as pointed out by (Stoev and Taqqu 2006).

Figure 4 displays noticeably the effect of the function $h(t)$ on a path of the process, simulated using the improved (Chan and Wood algorithm 1998). Notice the increasing jaggedness for decreasing values of $h(t)$ (or, symmetrically, the increasing smoothness for increasing values).

Remark 1 If almost surely $\beta_h([0,1]) > \sup_{t\in[0,1]} h(t)$, namely the uniform Hölder exponent of function h is larger than the supremum of $h(t)$, then $\alpha_{X_{h(t)}}(t) = h(t)$ and $\beta_{X_{h(t)}(t)} = \inf_{t\in J} h(t)$, almost surely (see Benassi et al. 1997; Péltier and Lévy Véhel

of the weighting kernel of the fBm (Rogers 1997; Cheridito 2001), the introduction of transaction costs (Guasoni 2006 or of delays in transaction times Cheridito 2003).

[8] Remind that the function $h(t)$ is Hölderian of order β on each compact interval $J \subset \mathbb{R}$ if, for each $t, s \in J$ and for $c > 0$, it holds $|h(t) - h(s)| \leq c|t-s|^{\beta}$, where $\beta > \max_{t\in J} h(t)$.

[9] Indeed, (Ayache and Taqqu 2005) provide sufficient conditions for a multifractional process with random functional parameter to be self-similar (in the sense of its marginal distributions) or to have stationary increments.

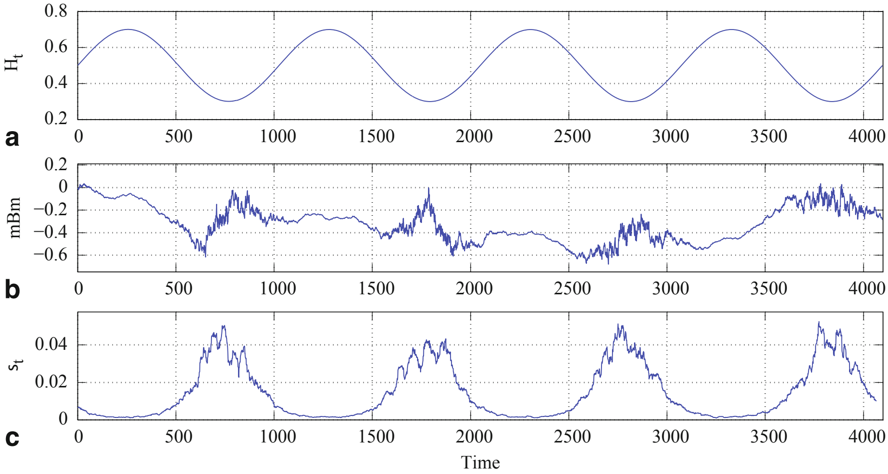


Fig. 4 Surrogated mBm. Panels: **a** sinusoidal $h(t)$; **b** surrogated process; **c** volatility process

1995). Therefore, within this class of models, under the above technical condition, the pointwise and the uniform regularity capture the time-changing volatility process that eventually causes the departure from efficiency, when the process is used to model financial data.

Remark 2 The moving average representation (13) is due to (Péltier and Lévy Véhel 1995). Benassi et al. (1997) provides the spectral representation (14), also convenient for simulation purposes. The two representations are equivalent up to a multiplicative deterministic function of time (Cohen 1999).

Remark 3 Because of the Remark 1, $h(t)$ is the pointwise Hölder exponent of the mBm at point t. This means that in a neighborhood of t, the process is asymptotically self-similar with parameter $h(t)$, in the sense stated by (Benassi et al. 1997), and denoted by $Y(t, au) = X_{h(t+au)}(t + au) - X_{h(t)}(t)$ the increment process, it holds

$$\lim_{a \to 0^+} a^{-h(t)} Y(t, au) \stackrel{d}{=} B_{h(t)}(u), u \in \mathbb{R}. \tag{15}$$

Equality (15) (see Péltier and Lévy Véhel 1995) states that at any point t there exists an fBm with parameter $h(t)$ 'tangent' to the mBm. From (11) it follows that $Var(B_{h(t)}(u)) = K^2 u^{2h(t)}$; so, recalling that the fBm is a Gaussian process, the infinitesimal increment of the mBm at time t, normalized by $a^{h(t)}$, is normally distributed with mean 0 and variance $K^2 u^{2h(t)}$ $(u \in \mathbb{R}, a \to 0^+)$. The estimator of $h(t)$ discussed in the sequel is based on this result.

Remark 4 The hölderianity of function $h : [0, \infty) \to (0, 1]$ represents a sufficient condition for the continuity of the motion. This constraint is relaxed by Ayache and Lévy Véhel 2000, who define a Gaussian process, the *Generalized Multifractional Brownian Motion*, that extends the mBm and allows the functional parameter $h(t)$ to belong to a set of functions larger than the space of the Hölder functions. The result is a continuous process with Hölder regularity given by even very irregular functions.

Remark 5 The covariance of the mBm reads as (Ayache et al. 2000)

$$\mathbb{E}(X_{h(t)}(t)X_{h(s)}(s)) = K^2 D(h(t), h(s)) \left(t^{h(t)+h(s)} + s^{h(t)+h(s)} - |t-s|^{h(t)+h(s)}\right) \tag{16}$$

where

$$D(h(t)+h(s)) = \frac{\sqrt{\Gamma(2h(t)+1)\Gamma(2h(s)+1)\sin(\pi h(t))\sin(\pi h(s))}}{2\Gamma(h(t)+h(s)+1)\sin\left(\pi \frac{h(t)+h(s)}{2}\right)}.$$

This results will be useful to justify the use of a class of estimators of the functional parameter $h(t)$.

Remark 6 Unlike the generalization considered in (Coeurjolly 2000), that assumes K to depend on time, here we maintain constant the scale parameter and explain the whole process variability in terms of its functional parameter $h(t)$.

Remark 7 An interesting feature of the mBm is that, similarly to the empirical distributions of financial returns, the unconditional distribution of its increments is high-peaked and fat-tailed. This follows from the variability of the Hölder exponent along the process' paths. More precisely, consider a multifractional Brownian motion sampled at n points on the unit time interval, with functional parameter $h(t_i)$ for $i = 0, \ldots, n$ and $t_i = \frac{i}{n}$. Since at each fixed time t_i the mBm behaves like an fBm with parameter $h(t_i)$ (see (15)), we can exploit relation (11). This implies $\mathbb{E}B_{h(t_i)}(t_i)^2 = \text{Var}\left(B_{h(t_i)}(t_i)\right) = K^2 t_i^{2h(t_i)}$, and hence-by the stationarity of the increments of the fBm-$K^2 t_i^{2h(t_i)}$ is also the variance of the t_i-lagged increments. It follows that the variance of each random variable of the sequence above defined is $K^2 n^{-2h(t_i)}$. Therefore, the variance of the unconditional distribution of the mBm above defined reads as

$$\sigma^2 = \frac{K^2}{n} \sum_{i=1}^{n} n^{-2h(t_i)} = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 \tag{17}$$

where $\sigma_i^2 = K^2 n^{-2h(t_i)}$. Finally, the unconditional density can be written as

$$f_X(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{e^{-\frac{x^2}{2\sigma_i^2}}}{\sigma_i \sqrt{2\pi}} \tag{18}$$

We can now prove that $f_X(x)$ is leptokurtic with respect to the Gaussian distribution. To this aim, consider the index of kurtosis excess $\gamma_2 = \mathbb{E}X^4/\sigma^4 - 3$. Values of γ_2 larger than zero denote that the kurtosis is larger that the one of the Gaussian distribution. Let us calculate $\mathbb{E}X^4$ by the moment-generating function:

$$m_X(t) = \mathbb{E}\left(e^{tX}\right) = \int_{\mathbb{R}} e^{tx} f_X(x) dx = \int_{\mathbb{R}} e^{tx} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_i^2}} dx =$$

$$= \frac{1}{n} \int_{\mathbb{R}} \sum_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\sigma_i^2 t)^2}{2\sigma_i^2} + \frac{\sigma_i^2 t^2}{2}} dx = \frac{1}{n} \sum_{i=1}^{n} e^{\frac{\sigma_i^2 t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\sigma_i^2 t)^2}{2\sigma_i^2}} dx$$

As the integrands in the last line are normal densities with mean $\mu_i = \sigma_i^2 t$, the integral equals one. It follows that $m_X(t) = \frac{1}{n} \sum_{i=1}^{n} e^{\frac{\sigma_i^2 t^2}{2}}$. Recalling that $\mathbb{E}(X^k) = \frac{d^k m}{dt^k}(0)$, one has

$$\mathbb{E}(X^4) = \frac{d^4\, m}{dt^4}(0) = \frac{3}{n} \sum_{i=1}^{n} \sigma_i^4$$

Substituting in the index γ_2 and solving the inequality

$$\frac{\frac{3}{n} \sum_{i=1}^{n} \sigma_i^4}{\left(\frac{1}{n} \sum_{i=1}^{n} \sigma_i^2\right)^2} - 3 > 0$$

leads to

$$\sum_{i=1}^{n} \sigma_i^4 > \frac{1}{n} \left(\sum_{i=1}^{n} \sigma_i^2 \right)^2 \tag{19}$$

Relation (19) is trivially true because of the Chebishev's inequality, stating that

$$n \left(\sum_{i=1}^{n} a_i b_i \right) \geq \left(\sum_{i=1}^{n} a_i \right) \left(\sum_{i=1}^{n} b_i \right),$$

where $n \in \mathbb{N}$ and $a_1 \leq a_2 \leq \cdots \leq a_n$, $b_1 \leq b_2 \leq \cdots \leq b_n$, with $a_i, b_i \in \mathbb{R}$.

In our case, setting $a_i = b_i = \sigma_i^2$, there is no need to assume the ordering of sequences a_i and b_i. Furthermore, since the σ_i's are all larger than zero, the inequality holds strictly.

Since the index of kurtosis is larger than zero, the unconditional distribution is leptokurtic with respect to the normal law whose variance is given by (17).

A further way to see things is to prove that the unconditional density exhibits fat tails with respect to the Gaussian density with variance (17). To this aim, it suffices to analyze the asymptotic behaviour of the ratio

$$Q(x) = \frac{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_i^2}}}{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}}.$$

Since σ^2 is the convex combination of the values σ_i^2, properly arranging these one can write

$$\sigma_1 < \cdots < \sigma_k < \sigma < \sigma_{k+1} < \ldots \sigma_n$$

from which it follows

$$Q(x) = \frac{\sigma}{n}\left(\sum_{i=1}^{k}\frac{e^{-\frac{(\sigma^2-\sigma_i^2)x^2}{2\sigma_i^2\sigma^2}}}{\sigma_i} + \sum_{i=k+1}^{n}\frac{e^{-\frac{(\sigma^2-\sigma_i^2)x^2}{2\sigma_i^2\sigma^2}}}{\sigma_i}\right)$$

and therefore

$$\lim_{x\to\infty} Q(x) = +\infty. \tag{20}$$

Limit (20) states that the function at the numerator tends to zero more slowly than the function at the denominator (i.e. the normal distribution) as x diverges, and this trivially indicates that the tails are heavier.

For practical purposes, the mBm can represent a good candidate as a model of financial dynamics, but it has a conceptual limit due to the deterministic function $h(t)$. Since this summarizes the weight that traders ascribe to the past in the sense stated by Remark 3, this is ultimately conditional to the new information that spreads into the market. Thus, there is no apparent reason to assume a deterministic functional parameter (see, e.g., Bianchi and Pianese 2007, 2008 or Bianchi and Pantanella 2013 for a discussion of this issue). For this reason it is useful to consider a different process, in which even the determinism of $h(t)$ is abandoned.

3.3 Multifractional Processes with Random Exponents

Starting from the fBm, (Ayache and Taqqu 2005) build a process in which the parameter H is replaced by a random variable or even by a stochastic process[10] $\{S(t,\omega)\}_{t\in\mathbb{R}}$ with values in the fixed interval $[a,b] \subset (0,1)$. The starting point to define the MPRE is the result provided by Papanicolaou and Sølna 2002, who showed that the stochastic integral

$$Z(t,\omega) = \int_{\mathbb{R}} \frac{e^{it\xi}-1}{i\xi\, |\xi|^{S(t)-\frac{1}{2}}} d\hat{W}(\xi), \tag{21}$$

i.e. the fBm (9) where the parameter H is replaced by the stochastic process $S(t,\omega)$, is well-defined when $S(t,\omega)$ is independent of $d\hat{W}$. In this case, the main results stated on mBm apply to $\{Z(t,\omega)\}_{t\in\mathbb{R}}$.

It is worthwhile to emphasize that the Roger's argument to prove that the fBm is not a semimartingale when $H \neq \frac{1}{2}$ does not apply to (21), as well as to the mBm defined in (13) or (14). In fact, the behaviour of the order-p variation depends on the

[10] To avoid ambiguity, when necessary, we write explicitly ω for the stochastic process.

assumptions made about the random parameter $S(t, \omega)$ or the functional parameter $h(t)$, respectively. Reasonably, this depends on how $S(t, \omega)$ is modeled; in this concern, some works show that the random parameter significantly changes through time (Costa and Vasconcelos 2003; Bertrand 2005; Bayrakta et al. 2013) and, on large samples, it is very close to $\frac{1}{2}$ (Bianchi 2005; Bianchi et al. 2013), what intuitively seems to be consistent with the arbitrage principle and indicates that the model's financial consistency is still an open question deserving further research.

Process (21) is not defined if $\{S(t, \omega)\}$ depends on $d\hat{W}$. In this case, Ayache and Taqqu consider the standard random wavelet series representation (10) and exploit its almost sure uniform convergence in (x, H) on each compact subset of $\mathbb{R} \times (0, 1)$; therefore, they replace (x, H) by $(t, S(t, \omega))$ and define the *Multifractional Process with Random Exponent* as

$$Z(t, \omega) = f_2(f_1(t)) = B_{S(t,\omega)}(t, \omega) \tag{22}$$

where

- $f_1 : [0, 1] \to [0, 1] \times [a, b]$ (i.e., $t \mapsto (t, S(t, \omega))$) and $f_2 : [0, 1] \times [a, b] \to \mathbb{R}$ (i.e., $(t, H) \mapsto B_H(t, \omega)$).
- the stochastic process $S : t \in [0, 1] \to [a, b] \subset [0, 1]$ (without loss of generality one could replace the time domain $[0, 1]$ by any compact interval);
- $\{B_H(t)\}$ is a random field for which $(t, H) \in [0, 1] \times [a, b] \subset (0, 1)$.

The construction of the MPRE does not necessarily require neither stationarity nor independence of $S(t, \omega)$ on the Brownian motion W. When independence is assumed, the MPRE recovers the main results stated for the mBm; if dependence holds, the kernel $(t - s)_+^{S(t,\omega)-1/2} - (-s)_+^{S(t,\omega)-1/2}$ is no longer adapted to the natural filtration of W and the integral is no longer defined. In this case, the standard random wavelet series representation (10) can be still used, since it does not involve the variable s.

It is worthwhile to recall four relevant features of the MPRE proved by (Ayache and Taqqu 2005):

Remark 8 The continuity of the paths of $\{S(t, \omega)\}$ implies the continuity of $\{Z(t, \omega)\}$. In addition, if S is a non-degenerate process, $Z(t, \omega)$ is not Gaussian.

Remark 9 As in the case of the mBm, if $\beta_S([0, 1]) > \sup_{t\in[0,1]} S(t, \omega)$ with probability 1, then almost surely $\alpha_Z(t, \omega) = S(t, \omega)$ at any point $t \in (0, 1)$ and $\beta_Z(J, \omega) = \inf_{t\in J} S(t, \omega)$. This means that, almost surely, the pointwise Hölder exponent of the MPRE equals its stochastic functional parameter S and the uniform Hölder exponent equals the infimum of S over J, so preserving the information in terms of pointwise regularity of the process.

Remark 10 If S is a random variable independent of W in (21) then, for $h \in [0, 1-t]$, $\{(Z(t + h) - Z(t)\} \stackrel{d}{=} \{(Z(h) - Z(0)\}$, i.e. the increments of Z form a stationary sequence.

Remark 11 If S is a stationary stochastic process independent of W in (21), then $\{Z(t, \omega)\}$ is self-similar in its marginal distributions, namely $Z(at) \stackrel{d_1}{=} a^{S(t)} Z(t)$.

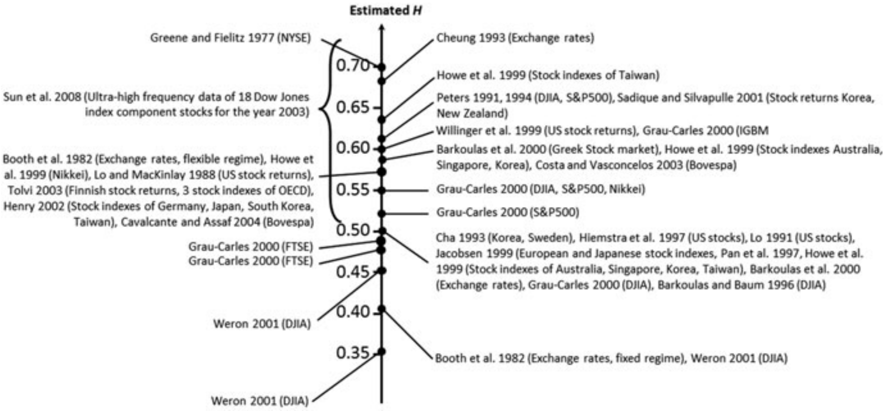


Fig. 5 Some estimates of the long-range dependence parameter H of several financial time series

4 Financial intuition for the multifractionality

Starting from its introduction at the end of 1960 s (Papanicolaou and Soølna 1968), the fBm was used as a model of financial time series and a number of contributions tried to estimate empirically its parameter H, called Hurst exponent, using different techniques (R/S analysis, Whittle's estimators, the Higuchi method, just to quote the main tools).

In particular, one of the most used estimators is based on the range statistics, whose role in the estimation of the long-run dependence has been widely studied in literature. In this respect, interesting theoretical as well as practical results are provided by Tapiero and Vallois (see Tapiero 2000; Tapiero and Vallois 2007; Vallois and Tapiero 1996; Vallois and Tapiero 1997; Vallois and Tapiero 2001; Vallois and Tapiero 2008). In detail, in (Vallois and Tapiero 1996) a modified Hurst exponent is calculated starting from run length statistics; in (Vallois and Tapiero 1997), the authors provide a definition of reliability based on a process range; in (Tapiero and Vallois 2007) the mean and the variance are explicitly calculated for a memory-based persistent process.

While several results are available from a theoretical viewpoint, from an empirical perspective many inconsistent results were obtained with respect to financial time series (see the non exhaustive list in Fig. 5), even for the same time series using different estimation methods and/or with respect to different time intervals. The unquestionable diversity of the estimates strongly indicates that a sole parameter of long-run dependence cannot seize the complexity of the price process. Indeed, the great variability of the estimates can be parsimoniously explained by assuming that the intensity of dependence changes through time (see the discussion in Bianchi and Pianese 2008). From a modeling viewpoint, if one wants to remain within the Brownian paradigm, the most immediate choice is to consider the mBm/MPRE processes. In such cases, the key parameter $h(t)$, or $S(t)$ for the MPRE, can be interpreted as summarized in Table 1.

Table 1 Interpretation of $h(t)$ (or $S(t)$)

$h(t)$	Stochastic consequence	Investors' belief	Market consequence
$> \frac{1}{2}$	Positive dependence Low variance Persistence (pos. or neg. trend)	Confidence that future information will confirm past positions	Low volatility Underreaction Overconfidence as $h(t) - \frac{1}{2}$ increases
$= \frac{1}{2}$	Independence	Past information fully discounted by prices	Efficiency
$< \frac{1}{2}$	Negative dependence High variance Antipersistence (mean reversion)	Confidence that future information will contradict past positions	High volatility Overreaction as $\frac{1}{2} - h(t)$ increases

The theoretical justification for this interpretation is provided by Remark 1 (Remark 9) and Remark 3. Since in a neighborhood of any fixed time t_0 the mBm, as well as the MPRE if S is independent of dW, behaves like an fBm of parameter equal to the pointwise functional (or stochastic) parameter, the intuition is straightforward: $h(t)$ or $S(t)$ can be read as the weight assigned by investors to past prices when they take their trading decisions. In this way, with respect to a given market, when $h(t) = \frac{1}{2}$ (independence, Brownian motion), the current price discounts all past prices and the market is efficient; $h(t) > \frac{1}{2}$ (positive long-range dependence, low variance and trends) means that the investors believe that future prices will move accordingly to past ones, no matter if in a bullish or bearish market. As a consequence, the stronger this belief the higher the difference $h(t) - \frac{1}{2}$; momentum strategies predominate and the market reacts only gradually to new information, generating what in Behavioural Finance is known as *underreaction*. On the contrary, $h(t) < \frac{1}{2}$ (negative dependence, high variance and mean reversion) denotes the investors' belief that future prices will contradict the current price. This typically occurs when (or as a consequence of) some bad news suddenly spreads into the market, triggering a quick buy-and-sell activity with profit taking (the so called touch-and-go market). Even if small, the capital gain coming from the transactions in this market phase can satisfy the investors, due to the perceived extreme unpredictability of how future information could influence the current price (*overreaction*).

This interpretation naturally suggests that the overall dynamics of a market (or even of single stocks) is nothing but a collection of local disequilibria and equilibria in the sense stated by the EMH. Nonetheless, the fact that disequilibria roughly compensate over a sufficiently long interval (see Fama 1998) does not imply that efficiency holds everywhere, but only on large samples. This is exactly the meaning of $\mathbb{E}\left(S_T | \mathcal{F}_t\right) = \mathbb{E}_t\left(S_T\right) = S_t$, in which the filtration $\mathcal{F}_t$ summarizes *the whole* past history, not just a part of it. The compensating mechanism described above works if, in the long run, $S(t)$ distributes with roughly the same masses above and below $\frac{1}{2}$. With the working hypothesis that the long-run dependence parameter (and hence the pointwise regularity) can change through time, it looks natural to link the efficiency of a market to the velocity of reabsorption of the departures from the threshold $\frac{1}{2}$.

5 Estimation of the Pointwise Regularity

The idea that the pointwise regularity of financial time series changes through time and is possibly well-behaved with respect to the threshold $\frac{1}{2}$ needs to be tested, in order to validate or at least falsify the market mechanism as outlined above. Therefore, one or more estimators of the pointwise regularity are needed; by definition, these cannot be asymptotical, as the majority of the estimators available for the long-range dependence parameter. Many authors attempted to solve the estimation problem and a short survey of these contributions could be useful to contextualize the technique that will be described hereinafter.

Using the method defined in (Benass et al.1998) for filtered white noises, (Istas and Lang 1997) and (Benassi 1998) introduce the generalized quadratic variations[11], an estimator of a continuously differentiable function, whose pro is allowing Gaussian limiting distribution with $\sqrt{N}$-rate of convergence, N being the sampling points along the process path. The result is extended by (Coeurjolly 2005), who considers an estimator based on a local estimation of the second order moment of a unique discretized filtered path. This allows to consider Hölderian functions (of arbitrary positive order) and to provide limit theorems for the functional estimators. A semi-parametric estimator for a piece-wise constant $h(t)$ is proposed in (Benassi et al.1999) and Benassi et al. 2000 with the aim to detect abrupt changes of the Hölder exponent for Gaussian processes with almost sure continuous paths. To identify the functional parameter of an even more general mBm (the Generalized mBm), (Ayache and Lévy Véhel 2004; Ayacheet al. 2005, Ayache et al. 2007) use the generalized quadratic variation and derive a central limit theorem for their estimator. With a specific look to financial applications, (Bianchi 2005) extends to the mBm the class of estimators introduced for the fBm in (Péltier and Lévy Véhel 1994) and studies its Gaussian limiting distribution with $(\sqrt{\delta}\log N)$-rate of convergence, δ and N respectively being the length of the estimation window and the number of sampling points. More recently, (Loutridis 2007) proposes an algorithm based on the scaled window variance method for estimating both global and local scaling exponents and claims its simplicity and computational efficiency with respect to other techniques. Finally, a class of consistent estimators based on convex combinations of sample quantiles of discrete variations is proposed by (Coeurjolly 2008), who also derives the almost sure convergence and the asymptotic normality of the estimators.

---

[11] Basically, with regard to the process X sampled at N times, the generalized quadratic variation is defined as $V_N = \sum_{p=0}^{N-2}\left(X\left(\frac{p+2}{N}\right) - 2X\left(\frac{p+1}{N}\right) + X\left(\frac{p}{N}\right)\right)^2$; the variation serves to define the estimator $\hat{h}_N = \frac{1}{2}\left(1 - \frac{lnV_N}{lnN}\right)$ which, under some assumptions on the function $h(t)$, satisfies $lim_{N\to+\infty}\hat{h}_N = inf_{t\in(0,1)}h(t)$ almost surely. The result stated for the infimum provides the way to estimate h itself at any point $t \in (0,1)$; in fact choosing a proper (ϵ, N)-neighborhood of t one can calculate the generalized quadratic variation $V_{\epsilon,N}$ and hence the estimator $\hat{h}_{\epsilon,N}$, that satisfies $lim_{N\to+\infty}\hat{h}_{\epsilon,N} = inf_{|s-t|<\epsilon}h(s)$ (a.s). Letting ϵ tend to zero one gets an estimate of $h(t)$.

Our approach is based on the *Absolute Moment Based Estimator* introduced by (Bianchi 2005), and refined by (Bianchi et al. 2013), as an extension of the estimator of the parameter H of the fBm defined by (Péltier and Lévy Véhel 1994). It works for the multifractional Brownian motion, and even if the case of a stochastic parameter (MPRE) hasn't been covered yet from a theoretical viewpoint, the estimation works also when the functional parameter $h(t)$ is very irregular, as it will be seen hereinafter. In order to discuss this class of estimators, let us assume that the process (13) is sampled in discrete time on the grid $t = 1, \ldots, n$. From Remark 3, one has[12].

$$X_{j+q,n} - X_{j,n} \stackrel{d}{=} \mathcal{N}\left(0, K^2 \left(\frac{q}{n-1}\right)^{2h\left(\frac{t}{n-1}\right)}\right), \tag{23}$$

for $j = t - \delta, \ldots, t - q; \quad t = \delta + 1, \ldots, n - q + 1; \quad q = 1, \ldots, \delta$. The construction of the estimator starts from the formula providing the k^{th} absolute moment of $Y \sim \mathcal{N}(0, \sigma^2)$

$$\mathbb{E}(|Y|^k) = \frac{2^{k/2}\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\sigma^k \tag{24}$$

Exploiting (23) one can define the quantity

$$\mathbb{S}^k_{\delta,q,n,K}(t) = \frac{1}{\delta - q + 1}\sum_{j=t-\delta}^{t-q} |X_{j+q,n} - X_{j,n}|^k \quad t = \delta + 1, \ldots, n$$

which, by (24), leads to

$$\mathbb{E}(\mathbb{S}^k_{\delta,q,n,K}(t)) = \mathbb{E}\left(\frac{1}{\delta - q + 1}\sum_{j=t-\delta}^{t-q} |X_{j+q,n} - X_{j,n}|^k\right)$$

[12] The variance in (23) follows from assuming a smooth $h(t)$. In fact, from (16) it follows:

$$\begin{aligned} Var\,(X_t - X_s) &= \mathbb{E}\,(X_t - X_s)^2 - (\mathbb{E}(X_t - X_s))^2 = \mathbb{E}\left(X_t^2 + X_s^2 - 2X_tX_s\right) = \\ &= t^{2h(t)} + s^{2h(s)} - D(h(t), h(s))\left(t^{h(t)+h(s)} + s^{h(t)+h(s)} - |t-s|^{h(t)+h(s)}\right) \\ &= t^{h(t)}(t^{h(t)} - D(h(t), h(s))t^{h(s)}) + s^{h(s)}(s^{h(s)} - D(h(t), h(s))s^{h(t)}) + \\ &\qquad + D(H_t, H_s)|t-s|^{h(t)+h(s)}. \end{aligned}$$

Since $\lim_{|h(t)-h(s)|\to 0} D(h(t), h(s)) = 1$, whenever $h(t) \approx h(s)$, one has $Var\,(X_t - X_s) \approx |t - s|^{2h(t)}$. Assuming that the mBm is sampled in discrete time over n points with $Var(X_n - X_0) = K^2$ entails therefore

$$Var\left(X_{\frac{t+q}{n-1}} - X_{\frac{t}{n-1}}\right) \cong K^2 \left|\frac{t+q}{n-1} - \frac{t}{n-1}\right|^{2h\left(\frac{t}{n-1}\right)} = K^2\left(\frac{q}{n-1}\right)^{2h\left(\frac{t}{n-1}\right)}$$

that the variance in (23)

$$= \frac{2^{k/2}\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} K^k \left(\frac{q}{n-1}\right)^{kh(t)} \tag{25}$$

Since the ratio

$$\frac{\mathbb{S}^k_{\delta,q,n,K}(t)}{\mathbb{E}\left(\mathbb{S}^k_{\delta,q,n,K}(t)\right)} = \frac{\sqrt{\pi}\mathbb{S}^k_{\delta,q,n,K}(t)}{2^{k/2}\Gamma\left(\frac{k+1}{2}\right) K^k \left(\frac{q}{n-1}\right)^{kh(t)}} \tag{26}$$

tends to 1 in probability as δ tends to infinity (see Bianchi 2005 for the proof), one has

$$\frac{\sqrt{\pi}\mathbb{S}^k_{\delta,q,n,K}(t)}{2^{k/2}\Gamma\left(\frac{k+1}{2}\right) K^k} \xrightarrow{P} \left(\frac{q}{n-1}\right)^{kh(t)} \tag{27}$$

and hence

$$\frac{\log\left(\sqrt{\pi}\mathbb{S}^k_{\delta,q,n,K}(t)/\left(2^{k/2}\Gamma\left(\frac{k+1}{2}\right) K^k\right)\right)}{k\log\left(\frac{q}{n-1}\right)} \xrightarrow{\text{P}} h(t). \tag{28}$$

By (28) directly follows the class of estimators

$$h^k_{\delta,q,n,K}(t) = \frac{\log\left(\frac{\sqrt{\pi}}{\delta-q+1}\sum_{j=t-\delta}^{t-1}|X_{j+q,n} - X_{j,n}|^k/\left(2^{k/2}\Gamma\left(\frac{k+1}{2}\right) K^k\right)\right)}{k\log\left(\frac{q}{n-1}\right)}. \tag{29}$$

For the estimator's distribution, it can be proved that

$$k\log\left(\frac{n-1}{q}\right)\sqrt{\delta-q+1}(h(t) - h^k_{\delta,q,n,K}(t)) \stackrel{d}{\to} \mathcal{N}\left(0, \frac{\pi}{2^k\Gamma^2\left(\frac{k+1}{2}\right)}\sigma^2\right) \tag{30}$$

where σ^2 is the limit variance of a series of normalized nonlinear functions of a stationary Gaussian sequence with slowly decaying autocorrelation function. It is worthwhile to underline that relation (30) entails a rate of convergence for the estimator equal to $O\left(\delta^{-\frac{1}{2}}(\log n)^{-1}\right)$, which provides reliable estimates even for small δ's.

Remark 12 **(Optimal q and k)** Relation (30) provides conditions useful to set two of the four parameters of the estimator, q and k. Concerning the former, as the estimator's variance grows with q, a natural choice consists in setting it equal to 1, the minimal admissible value. The optimal choice of k is a little bit more complicate to set and passes through writing the variance in (30) for $h^k_{\delta,q,n,1}(t) = H = \frac{1}{2}$ and $q = 1$. Toilsome computations show that in this case one gets[13]

$$Var(h^k_{\delta,1,n,1}(t)) = \frac{\sqrt{\pi}}{\delta k^2\log^2(n-1)\Gamma^2\left(\frac{k+1}{2}\right)}\cdot\left[\Gamma\left(\frac{2k+1}{2}\right) - \frac{1}{\sqrt{\pi}}\Gamma^2\left(\frac{k+1}{2}\right)\right]$$

[13] For $h \neq \frac{1}{2}$, the variance of $h^k_{\delta,q,n,1}(t)$ is hard to deduce because of the term σ^2 that appears in the variance of (30).

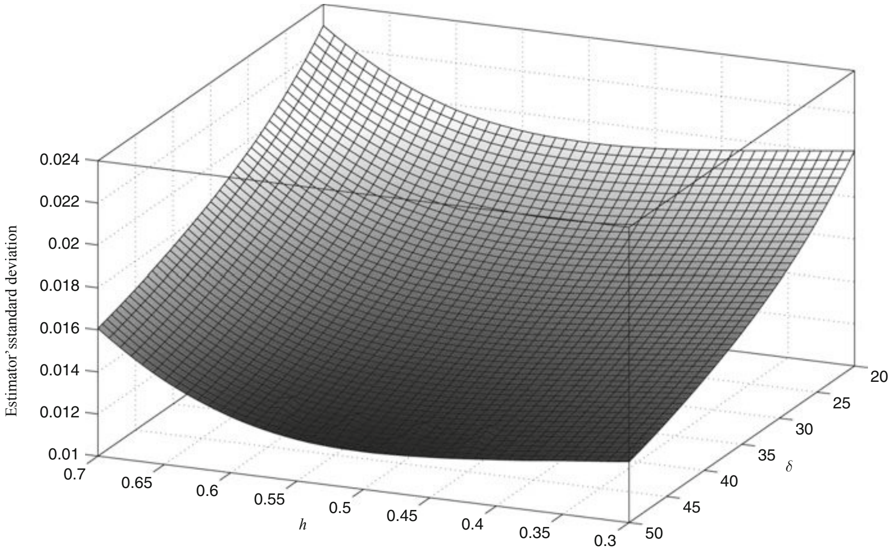


Fig. 6 Standard deviation of the estimator calculated by locally weighted smoothing quadratic regression

which, minimized, leads to the optimal k ($k = 2$).

In the general case, the variance of relationship (30) is hard to calculate due to the term σ^2. The estimator's standard deviation for different values of h and δ can be calculated by Monte Carlo simulation. Figure 6 displays the interpolating surface-using the locally weighted smoothing quadratic regression-obtained from 1,000 samples of length $N = 4,096$.

Remark 13 **(Optimal K)** As widely discussed in (Bianchi and Pianese 2013), when actual data are taken into consideration, the parameter K to be valued in (29) is generally unknown and a misleading K causes a shift of the estimated $h(t)$ sequence. In fact relation (29) can be written as follows:

$$h^k_{\delta,q,n,K}(t) = \frac{\log\left(\frac{\sqrt{\pi}\sum_{j=t-\delta}^{t-q}|X_{j+q,n}-X_{j,n}|^k}{(\delta-q+1)2^{k/2}\Gamma\left(\frac{k+1}{2}\right)}\right)}{k\log\left(\frac{q}{n-1}\right)} - \frac{\log K}{\log\left(\frac{q}{n-1}\right)}$$

Since the logarithm is slowly varying at infinity[14], the shift can be significant even when n is large. To estimate the right K from empirical data, notice that relation (23) implies the random variables collected in the set

$$V_q = \left\{X_{j+q,n} - X_{j,n} : h^k_{\delta,q,n,K}(t) \in (h^* - \epsilon, h^* + \epsilon)\right\} \tag{31}$$

[14] Function L is slowly varying at infinity if $\lim_{t\to\infty} \frac{L(\alpha t)}{L(t)} = 1$ for some $\alpha \in \mathbb{R}^+$

to be normally distributed with mean zero and variance equal to $K^2\left(\frac{q}{n-1}\right)^{2h^*}$, as $\epsilon \to 0$. So, applying (24), for each fixed h^* and for $\epsilon \to 0$, it is straightforward to obtain[15]

$$\log \mathbb{E}\left(|X_{j+q,n} - X_{j,n}|^s\right) = \log \frac{2^{s/2}\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\hat{K}^s + s\hat{h}\log\left(\frac{q}{n-1}\right) \tag{32}$$

This relation provides testable conditions to estimate K: its correct value can be evaluated through the intercept of a simple linear fit in the plane $\left(\log\frac{q}{n-1}, \log\mathbb{E}(|X_{j+q,n} - X_{j,n}|^s)\right)$ for increasing q's. In (Bianchi et al. 2013) the following steps are defined:

1. The initial estimation is run with an arbitrary K^*;
2. the empirical density function $f_{h^k_{\delta,q,n,K^*}(t)}(x)$ of the estimates is calculated;
3. the value $h^* = \arg\max f_{h^k_{\delta,q,n,K^*}(t)}(x))$ is taken as the center of the interval (whose inf and sup are the extremes of the bin h^k_{δ,q,n,K^*} belongs to) that serves to define the set in (31);
4. once V_q is settled, the log linear fit (32) is performed. The parameter K is therefore estimated by deducing $\hat{K}$ from the first term of the right-hand side of (32);
5. finally, the estimation is rerun with $\hat{K}$.

This procedure is particularly effective when the number of data in $f_{h^k_{\delta,q,n,K^*}(t)}(x)$ is suitably large for some x, so to guarantee the set V_q to have a number of elements sufficient to stabilize the log-linear regression (so, the higher the number of datapoints in the set V_q the more accurate the estimate of K).

Remark 14 **(Optimal δ)** The assumption that the observations are normally distributed with mean zero and variance in (23) holds if the parameter $h(t)$ lingers constant within the window δ. When dealing with real data, the effectiveness of this assumption strongly depends on the size of δ[16], which in principle should be allowed to vary over time. In fact, maintaining a constant length of the window implicitly means assuming that the arrival of information causes $h(t)$ to change only on a same fixed horizon. A reasonable alternative-proposed in Bianchi et al. 2013-is to change δ, preserving the local normality of the price variations. Roughly speaking, once a maximal and a minimal window lengths have been fixed (say δ_{max} and δ_{min}, respectively), one tests for normality over δ_{min}. If the normality test is passed, the window is increased by one unit and the test is looped until it fails or until δ reaches

[15] To avoid confusion with the notation, here we indicate the order of the absolute moment by s (instead of k used for $h^k_{\delta,q,n,K}(t)$).

[16] Using the two-sided Lilliefors test at different significance levels, Bianchi and Pianese 2008 show that, consistently with the mBm/MPRE model, the empirical significance values of three main stock indexes converge to the nominal ones as δ decreases. Although this suggests to maintain δ as short as possible in order to ensure the normality of data, a trade-off problem arises because the estimator's variance increases as the length of the window decreases. The discussion of this effect leads the authors in (Bianchi and Pianese 2008) to set $\delta = 30$.

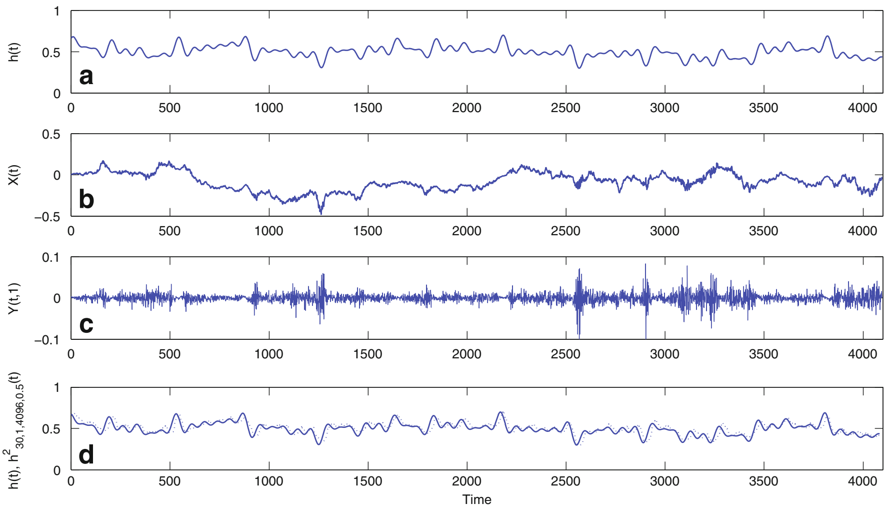


Fig. 7 Simulation and estimation of the mBm. Panel (**a**) Function $h(t)$ generated from the linear combination of sine and cosine waves of different frequencies; Panel (**b**) the resulting mBm ($X(t)$), simulated using the Wood and Chan circulant matrix method (Chan and Wood 1998); Panel (**c**) the increment process $Y(t,1) = X(t+1) - X(t)$ (notice the "bursts" of variance, typical of financial time series); Panel (**d**) the functional parameter $h(t)$ (*bold line*) and the estimated sequence $h^2_{30,1,4096,0.5}$ (*dotted line*)

δ_{max}. Clearly, with this procedure, the estimator's variance changes with δ and the sequence of estimated $h^k_{\delta,q,n,K}(t)$ can contain holes for the time subsets in which normality is rejected.

Figure 7 describes how the estimator functions. Notwithstanding the jaggedness of $h(t)$, thanks to the very good rate of convergence of the estimator, the sequence $h^2_{30,1,4096,0.5}$ succeeds to shadow the functional parameter. The results obtained with the time-changing δ are very similar and therefore they are omitted, in order to improve the readability of the Figure.

5.1 Regularity of a Portfolio

Given the estimator defined in the previous paragraph and assuming the same notation, it may be of interest to deduce the estimated regularity for a whole portfolio of N assets ${}_sX_{j,n}$, $s = 1,\ldots,N$, $j = t-\delta,\ldots,t-q$, $t = \delta+1,\ldots,n-q+1$, $q = 1,\ldots,\delta$, whose dynamics are modeled by N multifractional processes, each one characterized by its own functional parameter ${}_sh(t)$ estimated by ${}_sh^k_{\delta,q,n}(t)$. In the sequel, we will assume each process to have unit variance at time n and that $k = 2$[17].

[17] This assumption is not restrictive, since it has just been shown that $k = 2$ minimizes the estimator's variance (see Remark 14).

Denoted by α_s the weight allocated on the asset s^{th}, let

$$\Pi(t) = \sum_{s=1}^{N} {}_sX(t)\alpha_s$$

be the value of the portfolio at time t[18].

It can be proved that the portfolio's regularity estimator reads as

$$_{\Pi}h^2_{\delta,q,n}(t) = -\frac{\ln\left(\sum_{p=1}^{N}\sum_{r=1}^{N}\alpha_p\alpha_r\left(\frac{n-1}{q}\right)^{-({}_ph^2_{\delta,q,n}(t)+{}_rh^2_{\delta,q,n}(t))}\rho_{p,r,\delta}\right)}{2\ln\left(\frac{n-1}{q}\right)} \tag{33}$$

where $\rho_{p,r,\delta}$ denotes the correlation of the absolute increments of the assets p^{th} and r^{th}.

In fact, let $d(\Pi_{j,q}) = \Pi_{j+q} - \Pi_j = \sum_{s=1}^{N}\alpha_s d({}_sX_{j,q})$ denote the portfolio's increments, with $d({}_sX_{j,q}) :=_s X_{j+q} -_s X_j$. One has

$$_{\Pi}h^2_{\delta,q,n}(t) = -\frac{\ln\frac{\sum_{j=t-\delta}^{t-q}\left|\sum_{s=1}^{N}\alpha_s d({}_sX_{j,q})\right|^2}{K^2(\delta-q+1)}}{2\ln\left(\frac{n-1}{q}\right)}$$

$$= -\frac{\ln\frac{\sum_{s=1}^{N}\alpha_s^2\sum_{j=t-\delta}^{t-q}d({}_sX_{j,q})^2+2\sum_{j=t-\delta}^{t-q}\sum_{p=1}^{N-1}\sum_{r=p+1}^{N}\alpha_p\alpha_r\left|d({}_pX_{j,q})\right|\left|d({}_rX_{j,q})\right|}{K^2(\delta-q+1)}}{2\ln\left(\frac{n-1}{q}\right)}$$

Since from relation (23) it readily follows that

$$\frac{\sum_{j=t-\delta}^{t-q} d\left(X_j^2\right)}{K^2(\delta-q+1)} = \left(\frac{n-1}{q}\right)^{-2h^2_{\delta,q,n}(t)}$$

one has

$$_{\Pi}h^2_{\delta,q,n}(t) = -\frac{\ln\left(\sum_{s=1}^{N}\alpha_s^2\left(\frac{n-1}{q}\right)^{-2_sh^2_{\delta,q,n}(t)} + \frac{A}{K^2(\delta-q+1)}\right)}{2\ln\left(\frac{n-1}{q}\right)}. \tag{34}$$

[18] Likewise the notation already introduced, from now on the subscript will refer to the discrete time sampling.

where

$$A = \sum_{j=t-\delta}^{t-q} \sum_{p=1}^{N-1} \sum_{r=p+1}^{N} \alpha_p \alpha_r \left| d\left({}_p X_{j,q}\right) \right| \left| d\left({}_r X_{j,q}\right) \right|.$$

A more insightful way of writing relation (34) exploits again (23), from which one has

$$-{}_s h^2_{\delta,q,n}(t) = \frac{\ln \frac{\sum_{j=t-\delta}^{t-q} \left| d({}_s X_{j,q}) \right|^2}{K^2(\delta-q+1)}}{2 \ln\left(\frac{n-1}{q}\right)}$$

Once the above relation is written for p and r, summing up side by side we get

$$-{}_p h^2_{\delta,q,n}(t) -{}_r h^2_{\delta,q,n}(t) = \frac{\ln \frac{\sum_{j=t-\delta}^{t-q} \left| d({}_p X_{j,q}) \right|^2 \sum_{j=t-\delta}^{t-q} \left| d({}_r X_{j,q}) \right|^2}{K^4(\delta-q+1)^2}}{2 \ln\left(\frac{n-1}{q}\right)}$$

and therefore

$$\left(\frac{n-1}{q}\right)^{-\left({}_p h^2_{\delta,q,n}(t) + {}_r h^2_{\delta,q,n}(t)\right)} = \frac{\sqrt{\sum_{j=t-\delta}^{t-q} \left| d\left({}_p X_{j,q}\right) \right|^2 \sum_{j=t-\delta}^{t-q} \left| d\left({}_r X_{j,q}\right) \right|^2}}{K^2(\delta - q + 1)}.$$

Finally, setting $\rho_{p,r,\delta} := \frac{\sum_{j=t-\delta}^{t-q} \left| d({}_p X_{j,q}) \right| \left| d({}_r X_{j,q}) \right|}{\sqrt{\sum_{j=t-\delta}^{t-q} \left| d({}_p X_{j,q}) \right|^2 \sum_{j=t-\delta}^{t-q} \left| d({}_r X_{j,q}) \right|^2}}$,

$$\left(\frac{n-1}{q}\right)^{-\left({}_p h^2_{\delta,q,n}(t) + {}_r h^2_{\delta,q,n}(t)\right)} \rho_{p,r,\delta} = \frac{\sum_{j=t-\delta}^{t-q} \left| d\left({}_p X_{j,q}\right) \right| \left| d\left({}_r X_{j,q}\right) \right|}{K^2(\delta - q + 1)}$$

and by substituting in (34) one gets (33).

6 Analysis of Financial Data

As seen in the previous paragraphs, the estimation of the pointwise regularity is the core issue to assess the presence of multifractionality in financial time series. In this regard, we provide the results obtained using the Absolute Moment Based Estimators with seven main daily stock indexes, from different Countries: All Ordinaries (Australia), Dow Jones Industrial Average (USA), Footsie 100 (United Kingdom), DAX (Germany), Hang Seng (Hong Kong), MerVal (Argentine), Nikkei 225 (Japan), all

Table 2 Data set

Ticker	AORD	DJIA	FTSE	GDAXI	HSI	MERV	N225
#(obs)	5,926	5,882	5,893	5,921	5,926	5,799	5,739
Mean($\times 10^{-4}$)	1.9075	2.8620	1.6487	2.4892	3.5850	9.2457	−1.7962
St.Dev	0.0093	0.0111	0.0114	0.0146	0.0167	0.0265	0.0155
Skewness	−0.5173	−0.2165	−0.1034	−0.1108	−0.0043	0.7466	−0.0929
Kurtosis	9.2583	10.8882	8.9866	7.5653	12.0977	12.8407	8.4302

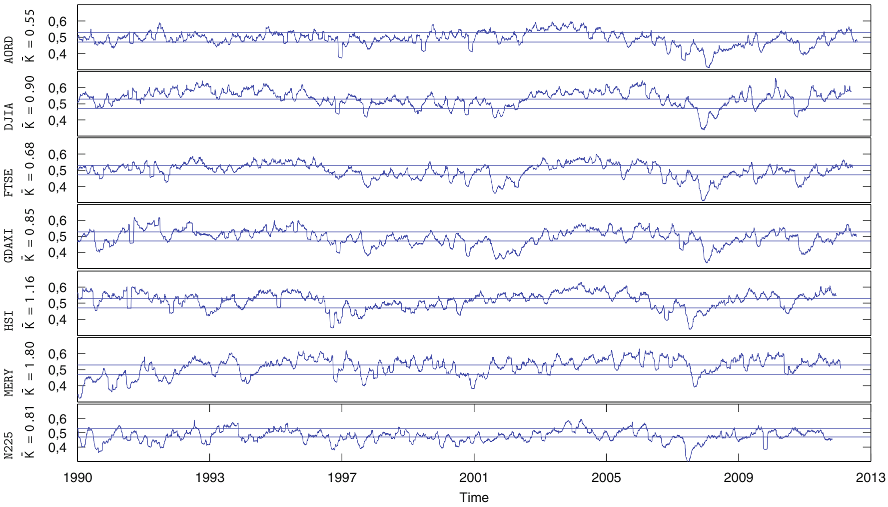


Fig. 8 Estimation of the pointwise regularity (parameters: $\delta = 30$, $q = 1$ and $k = 2$)

examined from January 1st, 1990 to April 30th, 2013 (Table 2 reports the lengths of the series along with their main distributional characteristics).

Figure 8 displays the pointwise regularity estimated by means of relation (29), with $\delta = 30$, $q = 1$ and $k = 2$[19]. The two straight lines provide the confidence interval around $\frac{1}{2}$ at a $p-$level of 5 % (see Remark 12). When the estimated $h(t)$ lies outside the acceptance region, the series displays a pointwise regularity significantly different from $\frac{1}{2}$. The figure clearly shows that the departures from this value are ubiquitous and generally last for not negligible time spans, as opposed to what the efficient market theories claim. A clear example is provided by the Dow Jones from 2003 to 2007, period during which the pointwise regularity was always significantly above $\frac{1}{2}$.

Table 3 summarizes the main distribution parameters of the estimated $h(t)$ of each index, along with the results of the Dickey Fuller test. Several interesting issues deserve a discussion.

[19] Using a time changing δ (see Remark 14) produces very similar results, hence to save the homogeneity of the results we have chosen a fixed window of proper length.

Table 3 Main distribution parameters of $h^2_{30,1,n,K}$

	AORD	DJIA	FTSE	GDAXI	HSI	MERV	N225
Mean	0.491	0.534	0.496	0.493	0.517	0.520	0.476
Median	0.494	0.539	0.501	0.499	0.525	0.530	0.478
St. Dev	0.044	0.050	0.049	0.051	0.051	0.055	0.043
Skewness	−0.642	−0.697	−0.707	−0.478	−0.669	−0.729	−0.417
Kurtosis	4.172	3.737	3.464	2.978	3.250	3.155	3.904
$\hat{\Phi}(z_1)$[a]	0.2763	0.1091	0.2794	0.2987	0.1964	0.1941	0.4344
$\hat{\Phi}(z_2) - \hat{\Phi}(z_1)$	0.5363	0.3127	0.4499	0.4416	0.3423	0.3057	0.4661
$1 - \hat{\Phi}(z_2)$	0.1874	0.5782	0.2707	0.2597	0.4614	0.5002	0.0994
ADF test							
Stat	−4.4301	−4.0108	−3.6949	−3.9739	−4.029	−4.6337	−4.631
pValue	0.0026	0.0089	0.0231	0.0098	0.0084	1.00E-03	1.00E-03

[a] $\hat{\Phi}(z_1)$ denotes the empirical cumulative distribution function of $h(t)$ calculated in $z_1 = \hat{\Phi}^{-1}(0.025) \simeq 0.4708$ and $z_2 = \hat{\Phi}^{-1}(0.975) \simeq 0.5292$, respectively the lower and the upper thresholds of the 95 % confidence interval of the estimator

Long-Term Mean Value It is really self-consistent that the overall mean of the estimated pointwise regularity is close to $\frac{1}{2}$ for all the stock indexes. Together with the mean-reverting behaviour of the pointwise regularity itself, this facet suggests an overall efficiency obtained by the compensation of local inefficiencies of opposite sign (below and above $\frac{1}{2}$, respectively). The potential to harmonize the no arbitrage principle with the turbulence that indeed affects the real financial markets is evident: if a sole value of the pointwise Hölder exponent is estimated on the whole time series using any asymptotic estimator, one would conclude that it roughly equals $\frac{1}{2}$, the sole value consistent with the absence of arbitrage. Actually, a finer examination leads to surface the complex nature of the price process that an analysis with a coarser resolution cannot seize. To say it with Bayaraktar et al. 2013, "*for long-term economic models, it is important to look at time-varying rates of long-range dependence, where the variation is caused by global economic factors, or regime changes. At some times, the market may be efficient, for example in a bullish exuberant economy like the late 1990s (...), while at other times, the Joseph effect may be prominent, such as during a recession or period of economic nervousness as in the early 1990s*".

Stationarity Significantly, the Augmented Dickey-Fuller test (see Table 3) indicates that the functional parameter $h(t)$ is trend stationary for each time series here analyzed. In principle, there is no apparent reason to take this result for granted; in addition, the mean reversion of the pointwise regularity to approximately $\frac{1}{2}$ is very pronounced. This indicates that markets generally tend to correct inefficiencies very quickly, even if periods exist where inefficiency prevails. A potentially relevant consequence of both the stationarity and the long run average equal to $\frac{1}{2}$ can simplify the models, whose nonstationarity ultimately depends on the nonstationarity of $h(t)$. In fact, if the pointwise regularity could be modeled by a random variable or a stationary process, then the resulting MPRE would be more tractable in terms of global probabilistic properties because of Remarks 10 and 11.

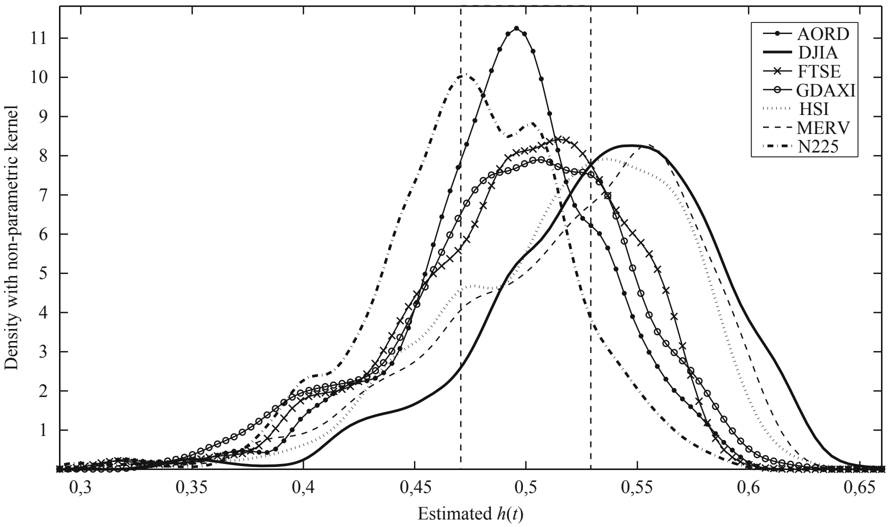


Fig. 9 Empirical non parametric density estimation of $h^2_{30,1,n,K}$. The vertical dashed bands denote the 95 % confidence interval

Magnitude of the Swings As displayed by Fig. 9, the swings of the estimated pointwise Hölder exponent around $\frac{1}{2}$ are mostly bounded between 0.4 and 0.6, against a confidence interval that for the given lengths is approximately equal to $[0.4708, 0.5292]$. The statistical significance of the departures from the value $\frac{1}{2}$ is therefore out of discussion, since in almost all the cases more than the fifty percent of the distribution mass is outside the confidence interval $\left[\hat{\Phi}^{-1}(.025), \hat{\Phi}^{-1}(.975)\right]$, even approaching the seventy percent for the Dow Jones, the Hang Seng and the Merval. Also notice that periods can be observed in which the pointwise regularity overstays significantly above $\frac{1}{2}$ (bubbles), but generally the larger the distance the faster the return to the central value. Again, this behaviour looks self-consistent in that it describes the tendency of the markets to move towards the equilibrium, even if this is nothing but a special case of a much more complex dynamics.

Well-Behaved Residuals The sequence $h^k_{\delta,q,n,K}(t)$ can be used to calculate the residuals $r(t) = \frac{dX(t)}{K(n-1)^{-h^k_{\delta,q,n,K}(t)}}$ of the increment process $dX(t) := \ln(X(t+1)/X(t))$. If the model and the estimator work, we expect $r(t)$ to be i.i.d. standard normal. In this regard, it is well-known that heavy tails and volatility clustering, the very motivations for the use for example of GARCH models, are expected to disappear once the returns are normalized by the level of volatility, making their distribution Gaussian. Nevertheless, under the classical ARMA-GARCH modeling, such scaled returns are still generally heavy tailed and show extremal dependence, whose strength reduces only as extreme levels increase. This unpleasant result, which makes questionable the capability of such models to capture entirely the variation in volatility, is generally addressed by modeling the residuals by some *ad hoc* distribution. The most popular distributions used to this aim are the (eventually asymmetric) Student's t, the generalized Pareto, the normal inverse Gaussian distribution or the double exponential, as a particular case of the generalized error distribution when the tail thickness parameter equals 1. Despite the many efforts devoted to the analysis of this topic, the choice of such distributions appears somewhat arbitrary. So, one may wonder whether multifractional modeling succeeds in improving the behaviour of the residuals. At a first glance, it seems so. In fact, Fig. 10 displays the box plots of the p-values calculated by the Jarque-Bera normality test (Jarque and Bera 1987)[20] for different samplings of the sequence of residuals (from approximately one trading month to approximately one trading year). The p-value generally stays above 0.05, which means that normality cannot be rejected at 5 % significance level. The worst cases occur for the indexes Dow Jones and Merval; for both, normality is mostly rejected starting from a span of about six trading months.

Table 4 displays the average value of the first four moments of the residuals; the average is calculated with respect to the number of samplings obtained dividing the number of data in the whole series by each fixed time span. Notice that all the moments are close to those of a normal i.i.d. distribution; nonetheless, the distributions are systematically lightly left-skewed, what will be discussed hereinafter.

Finally, for each index the sample autocorrelation functions of both the residuals and the squared residuals were calculated. It is quite evident that neither the residuals nor their squared transforms are autocorrelated at the 5 % significance level (see Fig. 11).

Autocorrelation of the Pointwise Regularity Parameter For each analyzed time series, the estimated sequences of the regularity parameters display a strong positive and slowly decaying autocorrelation (see Fig. 12). It should be noted that the high level of autocorrelation is not significant up to lag 30, as a consequence of the spurious effect of the overlapping induced by the estimation window δ. Anyway, over this value, the autocorrelation remains high $(0.35 - 0.50)$ and decays slowly, losing statistical significance only for very large lags (well above one trading year). This behaviour suggests two insights: (a) the level of pointwise regularity is strongly conditional to its past values. In other words, given the interpretation provided in Table 1, markets seem to preserve for a long time the memory of their own assessments in terms of the weight attributed to the past information; (b) the positive autocorrelation can contribute to limit the domain of the functions (mBm) or processes (MPRE) that can model the dynamics of the pointwise regularity itself.

Asymmetry of the Pointwise Regularity The estimates of $h(t)$ reflect the asymmetric role played by the information process. As claimed previously, all the distributions of the functional parameter are left skewed, what indicates that the low values in the

---

[20] The Jarque-Bera test quantifies the distance from the Gaussian distribution in terms of skewness and kurtosis and then computes a single p value using the sum of these discrepancies.

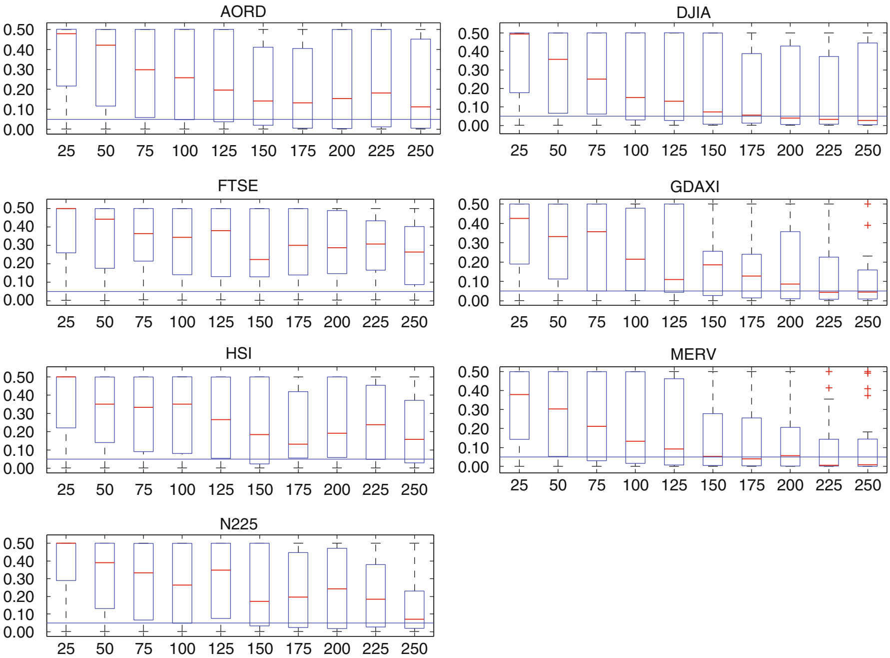


Fig. 10 Box plots of the p-values from the Jarque-Bera normality test of the residuals. On each box, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually

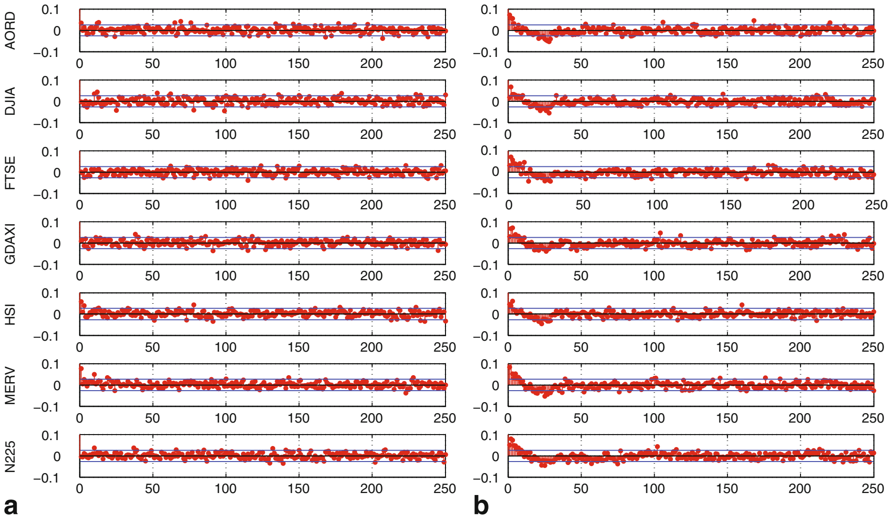


Fig. 11 Sample ACF of the **a** residuals and **b** squared residuals

Table 4 Main distribution parameters of the residuals

	25	50	75	100	125	150	175	200	225	250
					AORD					
Mean	0.029	0.033	0.031	0.029	0.032	0.033	0.027	0.029	0.031	0.028
St.Dev	0.993	1.004	1.006	1.008	1.009	1.009	1.009	1.010	1.010	1.008
Skewness	−0.103	−0.190	−0.214	−0.232	−0.244	−0.244	−0.238	−0.240	−0.252	−0.239
Kurtosis	2.941	3.294	3.383	3.452	3.462	3.516	3.483	3.466	3.500	3.485
					DJIA					
Mean	0.042	0.041	0.043	0.038	0.039	0.042	0.041	0.040	0.041	0.039
St.Dev	1.000	1.011	1.013	1.014	1.013	1.016	1.014	1.013	1.014	1.014
Skewness	−0.116	−0.170	−0.214	−0.216	−0.219	−0.243	−0.226	−0.235	−0.235	−0.233
Kurtosis	3.076	3.419	3.515	3.566	3.545	3.651	3.598	3.604	3.557	3.588
					FTSE					
Mean	0.026	0.022	0.023	0.022	0.020	0.022	0.018	0.022	0.022	0.018
St.Dev	1.003	1.011	1.014	1.014	1.015	1.016	1.015	1.014	1.014	1.015
Skewness	−0.034	−0.091	−0.109	−0.118	−0.141	−0.138	−0.136	−0.140	−0.137	−0.146
Kurtosis	2.760	2.994	3.002	3.100	3.080	3.093	3.083	3.116	3.088	3.087
					GDAXI					
Mean	0.030	0.028	0.028	0.028	0.028	0.027	0.029	0.027	0.027	0.028
St.Dev	0.992	1.003	1.006	1.006	1.008	1.008	1.009	1.009	1.011	1.010
Skewness	−0.098	−0.161	−0.199	−0.201	−0.206	−0.232	−0.221	−0.229	−0.232	−0.227
Kurtosis	2.930	3.193	3.361	3.392	3.431	3.516	3.477	3.503	3.530	3.503
					HSI					
Mean	0.035	0.038	0.036	0.038	0.037	0.038	0.036	0.037	0.037	0.037
St.Dev	1.006	1.016	1.023	1.020	1.024	1.022	1.021	1.021	1.022	1.022
Skewness	−0.045	−0.104	−0.152	−0.146	−0.151	−0.143	−0.144	−0.146	−0.158	−0.158
Kurtosis	3.101	3.356	3.520	3.558	3.563	3.543	3.542	3.528	3.579	3.557
					MERV					
Mean	0.030	0.029	0.029	0.028	0.029	0.028	0.027	0.027	0.028	0.029
St.Dev	0.996	1.012	1.018	1.019	1.021	1.023	1.022	1.022	1.022	1.023
Skewness	−0.097	−0.146	−0.165	−0.180	−0.182	−0.171	−0.168	−0.185	−0.171	−0.181
Kurtosis	3.183	3.597	3.674	3.717	3.762	3.785	3.768	3.801	3.799	3.848
					N225					
Mean	−0.018	−0.019	−0.016	−0.018	−0.019	−0.017	−0.021	−0.025	−0.021	−0.025
St.Dev	1.011	1.021	1.027	1.027	1.027	1.028	1.028	1.028	1.028	1.029
Skewness	0.001	−0.065	−0.065	−0.111	−0.092	−0.108	−0.116	−0.125	−0.123	−0.134
Kurtosis	2.884	3.205	3.304	3.355	3.408	3.409	3.461	3.454	3.462	3.489

pointwise regularity carry more weight than the high values. This is very consistent with the financial interpretation summarized in Table 1: the pointwise regularity is a proxy of the confidence the traders nourish in the past. In this view, we expect the two following implications:

- the increase of trust (revealed by increasing values of $h(t)$) is by definition a slow and gradual process, made of step-by-step reinforcements that can be frustrated suddenly by the occurrence of even a single shocking news. As shown by the distributions of Fig. 9, this is right what one observes as to the estimated pointwise regularity functions;
- since the higher $h(t)$ the smoother the price process (the more evident the trend), the higher $h(t)$ the more the past information weighs in the traders' assessment

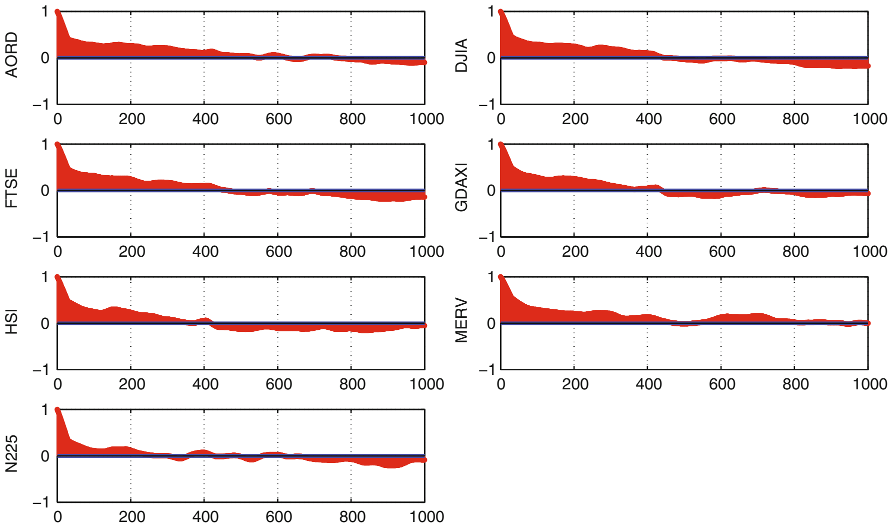


Fig. 12 Sample ACF of the estimated pointwise regularity

of future prices, regardless the direction of the trend. Therefore, large and sudden downward movements-that is the destruction of the "memory" or the "trust in the past"-are likely to take place when $h(t)$ is larger than $\frac{1}{2}$. What is more, the larger $h(t)$ the higher the conditional probability of a heavy downward variation.

The fact that very large and sudden downward variations tend to occur only when the Hölder exponent is much larger than $\frac{1}{2}$, while the reverse (large upward and more unstable variations) is much less frequent and tends to appear only immediately after the falls, is well-rendered by the behaviour of the conditional average variations of the pointwise regularity. More precisely, setting for notational simplicity $\hat{h}(t) = h^k_{\delta,q,n,K}(t)$ and denoting by $\hat{h}_m = \min\{\hat{h}(t)\}$ and $\hat{h}_M = \max\{\hat{h}(t)\}$ the minimum and the maximum of the estimated pointwise regularity, one can estimate the conditional average variations through the following steps:

- for any $\varepsilon \in \mathbb{R}^+$ small enough and any fixed $h \in [\hat{h}_m + \varepsilon, \hat{h}_M - \varepsilon]$, define the set $T_{(h,\varepsilon)} = \{t : \hat{h}(t) \in [h - \varepsilon, h + \varepsilon)\}$ as the set collecting all the times for which the corresponding estimates of $h(t)$ belong to the interval centered on h;
- for each $t \in T_{(h,\varepsilon)}$ and any fixed integer f identifying the number of trading days ahead with respect to t, calculate $\Delta_t(h, f) = \hat{h}(t + f) - h$;
- finally, calculate the average $\bar{\Delta}(h, f) = \#(T_{(h,\varepsilon)})^{-1} \sum_{t \in T_{(h,\varepsilon)}} \Delta_t(h, f)$, where as usual, $\#(X)$ denotes the number of elements of the set X.

The average conditional variations $\bar{\Delta}(h, f)$ are displayed in Fig. 13 over an horizon of one trading month ($f = 25$) only for two indexes (Footsie 100 and Hang Seng), because the output is nearly the same for all the examined series. From the perspective of behavioural finance, the conditional variations well describe both the underreaction and the overreaction biases: as to the first, large values of h tend to be

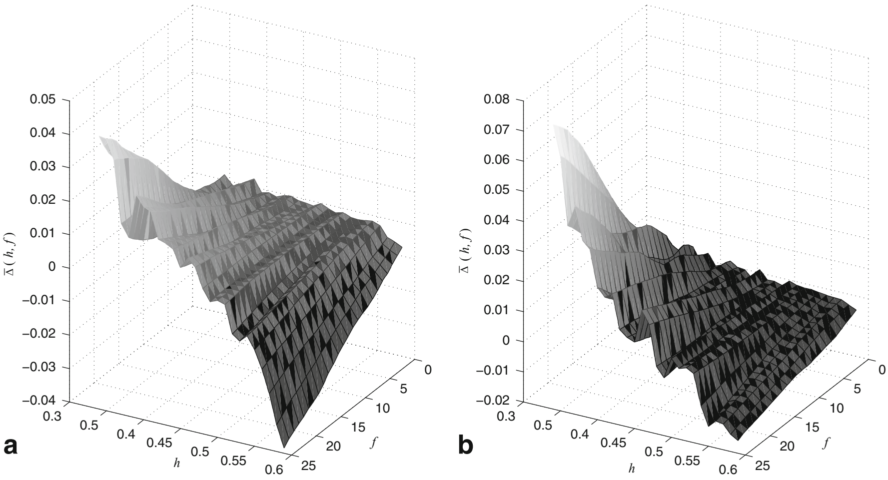


Fig. 13 Average variation of the pointwise regularity conditional to the estimated $h(t)$, on a number of f trading days to come. Panel **a** Footsie 100; Panel **b** Hang Seng

followed by small negative variations towards the equilibrium, as far as h itself is not too large (in this case the correction is usually heavy and abrupt); as to the second, low values of h, typical of financial crises, tend to be followed by wider and more unstable positive variations (the amplitude of the swings in Fig. 13 tend to diminish as h increases).

It is worthwhile to underline that this is can be seen as a stylized fact, since by construction the conditional average variations are calculated with respect to the whole series, and therefore they summarize reactions induced by even very different market conditions.

Pointwise Regularity of the Financial Crises In proximity of financial crises, the estimated pointwise regularity falls even to 0.35. The reason why the financial crises do not simply reset the 'memory' $\left(h(t) = \frac{1}{2}\right)$, but induce low values of $h(t)$ can be well explained in behavioural terms: when markets experience dramatic shocks (whose intensity can be estimated by the difference $\frac{1}{2} - h(t)$), traders typically react by trading frenetically and settle for reduced margins. Said differently, the collapse of the trust induces the traders to take profits as soon as possible. If this behaviour is widespread in the market, it causes the fall in the price and thus the asset becomes attractive for the next buyer, whose demand will increase the price again in a cyclical mechanism. This touch-and-go market produces antipersistence, which is the statistical counterpart of very low values of $h(t)$.

7 Further Developments

There is a growing awareness that fundamental models of stock prices are limited due to market complexity induced by several factors such as globalisation, the quantity and the quality of traded derivatives and information asymmetry and its management, just to quote the main ones.

Traders and investors increasingly are questioning the choice of models we use.

A potential good candidate model should be able to reduce in a coherent frame all empirical evidence that, collected by the stylized facts, do not agree with the current models. As a second step, a good candidate should be sufficiently formalized from a mathematical viewpoint to allow for the asset pricing, even of complex or illiquid financial derivatives. This has motivated research in behavioural finance such as the Lo's Adaptive Market Hypothesis; albeit their conceptual content is well grounded, these efforts are still insufficient.

In the previous paragraphs, we have introduced the main analytical properties of the multifractional processes and described how they can be used to model financial time series, in order to make the paradigm of Efficient Markets consistent with the stylized facts and, ultimately, with the biases brought to light by the Behavioural Finance. The synthesis that the multifractional stochastic processes provide for these two apparently opposite worlds benefits of a very rigorous and parsimonious mathematical framework. The parsimony resides in their functional parameter, which can be read as the time-changing weight that traders ascribe to past prices. A null weight indicates local efficiency (that is, a locally null autocorrelation function of the price variations); a positive weight denotes a positive inefficiency (that is, positive autocorrelation, trend and underreaction); finally, a negative weight reflects a locally negative inefficiency (i.e., negative autocorrelation, mean reversion and overreaction).

Under the assumption of multifractionality, the empirical estimates show that the functional parameter of several stock indexes fluctuates around $\frac{1}{2}$, that is the null weight, but even long time intervals exist where it lies very far from this mean value, in both the directions. These outcomes strongly suggest that the alternation of inefficiencies of opposite sign leads to long-term efficiency. But what generally matters is not just the long-term, and in this respect the main advantage offered by the model is to allow to assess, at each given time, how far the market is from efficiency and how probable a correction is.

Once an extensive empirical evidence will be produced about the consistency between the model and the actual financial time series (as to this, Bianchi and Pianese 2008; Bianchi and Pantanella 2010; Bianchi and Pantanella 2011; Bianchi et al. 2013) are promising contributions), the results will concur to model the functional parameter in order to face the problem of the asset pricing in the multifractional framework.

References

Arneodo, A., E. Bacry, and J. F. Muzy. 1998. Random cascades on wavelet dyadic trees. *Journal of Mathematical Physics* 39:4142–4164.

Ayache, A.. 2000. The generalized multifractional Brownian motion can be multifractal. *Publications du Laboratoire de statistique et probabilités* 22.

Ayache, A. 2013. Continuous gaussian multifractional processes with random pointwise Hölder regularity. *Journal of Theoretical Probability* 26:72–93.

Ayache, A., and J. Lévy Véhel. 2000. The generalized multifractional brownian motion. *Statistical Inference for Stochastic Processes* 3:7–18.

Ayache, A., and J. Lévy Véhel. 2004. On the identification of the pointwise Hölder exponent of the generalized multifractional Brownian motion. *Stochastic Processes and their Applications* 111:119–156.

Ayache, A., and M.S. Taqqu. 2005. Multifractional processes with random exponent. *Publicacions Matemàtiques* 49:459–486.

Ayache, A., S. Cohen, and J. Lévy Véhel. 2000. The covariance structure of multifractional Brownian motion, with application to long range dependence. In *ICASSP '00: Proceedings of the Acoustics, Speech, and Signal Processing, 2000. on IEEE International Conference*, IEEE Computer Society, Washington, DC, USA, ISBN 0-7803-6293-4, pp. 3810–3813, doi:http://dx.doi.org/10.1109/ICASSP.2000.860233.

Ayache, A., S. Benassi, S. Cohen, and J. Lévy Véhel. 2005. Regularity and identification of generalized multifractional Gaussian processes. *Séminaire de Probabilités XXXVIII, Lecture Notes in Math* 1857:290–312.

Ayache, A., P. R. Bertrand, and J. Lévy Véhel. 2007. A central limit theorem for the quadratic variations of the step fractional Brownian motion. *Statistical Inference for Stochastic Processes* 10:1–27.

Baillie, R. T.. 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73:5–59.

Bansal, R., and A. Yaron. 2004. Risks for the long-run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59:1481–1509.

Bayraktar, E., U. Horst, and R. Sircar. 2006. A limit theorem for financial markets with inert investors. *Mathematics of Operations Research* 31:789–710.

Benassi, A., S. Jaffard, and D. Roux. 1997. Gaussian processes and pseudodifferential elliptic operators. *Revista Mathematica Iberoamericana* 13:19–89.

Benassi, A., S. Cohen, and J. Istas. 1998. Identifying the multifractional function of a Gaussian process. *Statistics and Probability Letters* 39:337–345.

Benassi, A., S. Cohen, J. Istas, and S. Jaffard. 1998. Identification of filtered white noises. *Stochastic Processes and their Applications* 75:31–49.

Benassi, A., P. Bertrand, S. Cohen, and J. Istas. 1999. Identification d'un processus gaussien multifractionnaire avec des ruptures sur la fonction d'échelle. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* 329:435–440.

Benassi, Albert, Pierre Bertrand, Serge Cohen, and Jacques Istas. 2000. Identification of the hurst index of a step fractional brownian motion. *Statistical Inference for Stochastic Processes* 3:101–111.

Bender, C., T. Sottinen, and E. Valkeila. 2007. Arbitrage with fractional Brownian motion? *Theory of Stochastic Processes* 13:23–34.

Bertrand, Pierre. 2005. Financial modelling by multiscale fractional Brownian motion. In *Fractals in Engineering*, eds. Jacques Lévy Véhel and Evelyne Lutton, 181–196. London: Springer.

Bertrand, Pierre, Abdelkader Hamdouni, and Samia Khadhraoui. 2012. Modelling NASDAQ series by sparse multifractional Brownian motion. *Methodology and Computing in Applied Probability* 14:107–124.

Bianchi, S.. 2005. A cautionary note on the detection of multifractal scaling in finance and economics. *Applied Economics Letters* 12:775–780.

Bianchi, S. 2005. Pathwise identification of the memory function of the multifractional Brownian motion with application to finance. *International Journal of Theoretical and Applied Finance* 8:255–281.

Bianchi, S., and A. Pianese. 2007. Modelling stock price movements: Multifractality or multifractionality? *Quantitative Finance* 19:301–319.

Bianchi, S., and A. Pianese. 2008. Multifractional properties of stock indices decomposed by filtering their pointwise Hölder regularity. *International Journal of Theoretical and Applied Finance* 11:567–595.

Bianchi, S., and A. Pantanella. 2010. Pointwise regularity exponents and market cross-correlations. *International Review of Business Research Papers* 6:39–51.

Bianchi, S., and A. Pantanella. 2011. Pointwise regularity exponents and well-behaved residuals in stock markets. *International Journal of Trade, Economics and Finance* 2:52–60.

Bianchi, S., A. Pantanella, and A. Pianese. 2013. Modeling stock prices by multifractional Brownian motion: An improved estimation of the pointwise regularity. *Quantitative Finance* 13:1317–1330.

Calvet, L., and A. Fischer. 2002. Multifractality in asset returns: Theory and evidence. *The Review of Economics and Statistics* 84:381–406.

Calvet, L., A. Fischer, and B. B. Mandelbrot. 1997. A multifractal model of asset returns. Cowles Foundation Discussion Paper, Yale University, 1164.

Calvet, Laurent E., and Adlai J. Fisher. 2008. *Multifractal volatility: Theory, forecasting, and pricing*. Elsevier: Academic Press.

Campbell, J.Y., and J. H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107:205–251.

Chan, G., and A. Wood. 1998. Simulation of multifractional Brownian motion. *Proceedings in Computational Statistics*, pp. 233–238.

Cheridito, P. 2001. Regularizing fractional Brownian motion with a view towards stock price modelling. Ph.D. thesis, Eidgenöss Technische Hochschule Zürich, Swiss Federal Institute of Technology.

Cheridito, P. 2003. Arbitrage in fractional brownian motion models. *Finance and Stochastics* 7:533–553.

Coeurjolly, J. F. 2000. Inférence statistique pour les mouvements browniens fractionnaires et multifractionnaires, Ph. D. Thesis.

Coeurjolly, J. F. 2005. Identification of the multifractional Brownian motion. *Bernoulli* 11: 987–1008.

Coeurjolly, J. F.. 2008. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Annals of Statistics* 36:1404–1434.

Cohen, S. 1999. Chapter I. In *From self-similarity to local self-similarity: The estimation problem*, pages 3–16. ed. M. Dekking, J. Lévy-Véhel, E. Lutton, and C. Tricot, 3–16. Springer Berlin/Heidelberg

Cont, R. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1:223–236.

Costa, Rogério L., and G. L. Vasconcelos. 2003. Long-range correlations and nonstationarity in the Brazilian stock market. *Physica A: Statistical Mechanics and its Applications* 329:231–248.

Dasgupta, A., and G. Kallianpur. 2000. Arbitrage opportunities for a class of Gladyshev processes. *Applied Mathematics and Optimization* 41:377–385.

Decreusefond, L., and S. A. Üstünel. 1999. Stochastic analysis of the fractional brownian Motion. *Potential Analysis* 10:177–214.

Delbaen, F., and W. Schachermayer. 1994. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300:463–520.

Elliott, R. J., and J. Van Der Hoek. 2003. A general fractional white noise theory and applications to finance. *Mathematical Finance* 13:301–330.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25:383–417.

Fama, E.F., and K.R. French. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 22:246–273.

Fama, E. F.. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49 (3): 283–306.

Guasoni, P.. 2006. No arbitrage under transaction cost, with fractional Brownian motion and beyond. *Mathematical Finance* 16:569–582.

Henry, M., and P. Zaffaroni. 2003. Chapter III In *The long range dependence paradigm for macroeconomics and finance*, ed. Birkhäuser P. Doukhan, G. Oppenheim, and M.S. Taqqu, 417–438.

Hu, Y., and B. Øksendal. 2003. Fractional white noise calculus and applications to finance. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 6:1–32.

Hunt, G. A.. 1951. Random fourier transforms. *Transactions of the American Mathematica* 71:38–69.

Istas, J., and G. Lang. 1997. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l'Institut Henri Poincare (B) Probability and Statistics* 33:407–436.

Jarque, C. M., and A. K. Bera. 1987. A test for normality of observations and regression residuals. *International Statistical Review* 55:163–172.

Protter, P., R. A. Jarrow, and H. Sayit. 2009. No arbitrage without semimartingales. *The Annals of Applied Probability* 19:596–616.

Kahneman, D., and A. Tversky 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263–292.

Kolmogorov, A. N. 1940. Wienershe spiralen und einige andere interessante Kurven im Hilbertishen Raum. *DAS of the URSS (Nat. Sciences)* 26:115–118.

Kolwankar, K. M. and J. Lévy Véhel. 2002. A time domain characterization of the fine local regularity of functions. *Journal of Fourier Analysis and Applications* 8:319–334.

Lamperti, J. 1962. Semi-stable stochastic process. Transactions of The American Mathematical Society 104:62–78.

Lévy Véhel, J., and O. Barriére. 2008. Local Hölder regularity-based modeling of RR intervals. *Proceedings of the 21th IEEE international symposium on computer-based medical systems*, pp. 75–80.

Lo, A. W. 2004. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management* 30:15–29.

Lo, A. W. 2005. Reconciling efficient markets with behavioral finance: The adaptive market hypothesis. *Journal of Investment Consulting* 7 (2): 21–44.

Lo, A. W.. 2012. Adaptive markets and the new world order. *Financial Analysts Journal* 68:18–29.

Loutridis, S. J.. 2007. An algorithm for the characterization of time-series based on local regularity. *Physica A Statistical Mechanics and its Applications* 381:383–398.

Lustig, H. and S. Van Nieuwerburgh. 2005. Consumption Insurance and risk premia: An empirical perspective. *Journal of Finance* 60:1167–1219.

Mandelbrot, B. B., and J. W. Van Ness. 1968. Fractional brownian motions, Fractional noises and applications. *SIAM Review* 10:422–437.

Papanicolaou, G. C., and K. Sbibølna. 2002. Wavelet based estimation of local Kolmogorov turbulence. In *Long-range dependence: Theory and applications*. Birkhäuser, pp. 473–506.

Péltier, R.S., and J. Lévy Véhel. 1994. A new method for estimating the parameter of fractional Brownian motion. Technical report, CCSd/HAL : E-articles server (based on gBUS) [http://hal.ccsd.cnrs.fr/oai/oai.php] (France), http://hal.inria.fr/inria-00074279/en/.

Péltier, R. S., and J. Lévy Véhel. 1995. Multifractional Brownian Motion: Definition and preliminary results, Technical report, INRIA, citeseer. ist.psu.edu/peltier95multifractional.html.

Riedi, R. H. 2002. In *Multifractal processes*, ed. Birkhäuser P. Doukhan. G. Oppenheim, and M. Taqqu, pp. 625–715.

Rogers, L. C. G. 1997. Arbitrage with fractional Brownian motion. *Mathematical Finance* 7:95–105.

Sewell, M. 2007. Behavioural finance. http://www.behaviouralfinance.net/behavioural-finance.pdf.

Hersh, Shefrin, and Meir Statman. 1994. Behavioral capital asset pricing theory. *The Journal of Financial and Quantitative Analysis* 29:323–349.

Hersh Shefrin and Meir Statman. 2000. Behavioral portfolio theory. *The Journal of Financial and Quantitative Analysis* 35:127–151.

Shefrin, H. M. and R. H. Thaler. 1988. Hyothesis The Behavioral life-cycle. *Economic Inquiry* 26:609–643.

Shiller, R. J. 1984. Stock prices and social dynamics. Technical report, Brookings Papers on Economic Activity.

Shiryayev, A. N. 1998. On arbitrage and replication for fractal models. Technical report, Department of Mathematical Sciences, University of Aarhus, Denmark.

Stoev, A. S., and M. S. Taqqu. 2006. How rich is the class of multifractional Brownian motions? *Stochastic Processes and their Applications* 116:200–221.

Summers, L. H.. 1986. Does the stock market rationally reflect fundamental values? *Journal of Finance* 41:591–601.

Tapiero, Charles. 2010. *Risk finance and asset pricing*. New York: Wiley.

Tapiero, C. S. 2013. *Re-Engineering risks and the future of finance*, 233–250. London: EUROMONEY Books.

Tapiero, C.S., and P. Vallois. 2000. The inter-event range process and testing for chaos in time series. *Neural Network World* 10 (1–2) 89–99.

Tapiero, C. S., and P. Vallois. 2007. Memory-based persistence in a counting random walk process. *Physica A: Statistical Mechanics and its Applications* 386 (1): 303–317.

Vallois, P., and C. S. Tapiero. 1996. Run length statistics and the Hurst exponent in random and birth-death random walks. *Chaos, Solitons & Fractals* 7 (9): 1333–1341.

Vallois, P., and C. S. Tapiero. 1997. Range reliability in random walks, Mathematical methods of operations research. *Physica A: Statistical Mechanics and its Applications* 45 (3): 325–345.

Vallois, P., and C. S. Tapiero. 2001. The range inter-event process in a symmetric birth death random walk. *Applied Stochastic Models in Business and Industry* 17 (3): 293–306.

Vallois, P., and C. S. Tapiero. 2008. Volatility estimators and the inverse range process in a random volatility random walk and Wiener processes. *Physica A: Statistical Mechanics and its Applications* 387 (11): 2565–2574.

Yaglom, A. M. 1958. Correlation theory of processes with random stationary n^{th} increments. *American Mathmathical Society Translations*. 2:87–141.

Yaglom, A. M. 1965. Stationary gaussian processes satisfying the strong mixing condition and best predictable functionals. In *Proceedings international research seminar*. p. 241–252.

Yen, G. 2008. Efficient market hypothesis (EMH): Past, present and future. *Review of Pacific Basin Financial Markets and Policies* 11:305–329.

Financial Analytics and A Binomial Pricing Model

Charles S. Tapiero and Jiangyi Qi

1 Introduction

The Walras, Arrow and Debreu economic and financial equilibrium models have set the foundations for pricing in an equilibrium (complete) competitive financial and theoretical markets (Arrow 1951a, 1951b, 1963; Debreu 1959; Arrow and Debreu 1960; Lucas 1978). These models presume that events are memory-less, as well as a number of assumptions that are not always met in practice. Real financial data (such as the S&P, the Euro-Dollar exchange rate, etc.), indicate that financial rates of returns are not necessarily normally distributed as may be affected by past and observed returns (i.e. have memory, be auto-correlated and increasing or decreasing as a function of observed returns). For example, over a period of 759 consecutive days, we found that the probability of a price increase on the S&P will follow a price increase was 0.4774 while the probability of such an increase following an actual price decline in the previous day was 0.50964. By the same token, the probability of a price decline following a price increase was 0.5254 while a decline following a decline was found to have a probability of 0.4820. Analysis of other time series (such as on intraday data) has indicated more pronounced results. Over shorter periods of time (51 days), these probabilities were found to be even more pronounced (0.5925 and 0.4583) and (0.4074 and 0.54166). These observations lead to some financial traders to devise trading strategies based on persistent random walks (see for example Damien, at www.cetcapital.com) and profit from a financial markets incompleteness arising due to a mis-specification of underlying financial models.

The intent of this paper is to consider basic elements of the binomial pricing model by including a memory-based and a learning model. While most of the results presented in this paper are commonly known (for example, the Arrow-Debreu pricing model and the Bayesian learning model), we consider a short memory, regime switching model. That is, unlike Markov switching pricing models where one move

C. S. Tapiero (✉) · J. Qi
Department of Finance and Risk Engineering, New York University Polytechnic
School of Engineering, Brooklyn, New York, USA
e-mail: cst262@nyu.edu

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,*
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2_8,

from one pricing model to another in (an exogenous) probability, we define a model switching using the actual observed information of a price model at a given time. To simplify the presentation, we assume in fact that when a model history (limited for presentation purposes to one past period) is say of a price increase and then the price decrease, this will lead to a switch to another pricing model and vice versa. Such an approach is motivated by the belief that financial markets are characterize by long trends with fundamental underlying factors while future expectations (and thus pricing models) are sensitive to news, fears and financial assessments that lead to price models to switch from one type to another. For example, are new fears regarding financial markets altering future expectations and therefore their evolution? Most events might not alter the basic probabilities laws and randomness that determine the occurrence of subsequent events. But some events might have such an effect.

The traditional and statistical credibility theory approach commonly used in actuarial science is based on an approach that evaluates the objectivity and the subjectivity of a risk source and devises a statistical "learning" mechanism that allows the updating of the underlying claim probability. Using Bayesian statistics for example, credibility theory divides risk events into a number of classes each with a propensity for the event to occur and which are updated using subjective prior estimates of risk classes and its accrued experience—the process history observed. The goal of credibility theory is then to set up an experience rating system (and thus with a statistical memory). Unlike credibility theory, the short memory approach presumes that observed information may alter in probability the underlying random process defining the probability law of a subsequent event. Explicitly, while credibility theory seeks to integrate "experience" in estimating the propensity of a process future's states and its probabilities, short memory models are an inherent property of the process that determines, conditionally on observed information, its future process evolution. The credibility approach however, will use the fact that an event occurred to revise the probability (say by applying Bayes Theorem) to estimate the future probability of that event with a credibly estimated probability (since it is based on more information). In this sense these approaches differ fundamentally. The examples we shall use in this paper seeks to outline their differences.

In fact, what is memory and what is time? These are two intrinsically and fundamental related concepts that underlie financial modeling. Time is often defined implicitly in terms of a relative time line to organize events and their theories. Dynamic models for example (such as price processes) are merely models sought to organize a hypothetical evolution of prices along a given time line. Extending this time line beyond a current time provides means to forecast future events and prices. Time is then defined as a sequence of points, denoted by the year, the month, the week, the day, the hours, the seconds and the microseconds. Each time line may have different properties, explanatory and predictive powers and each embeds past events and their effects on current ones—the memory. The greater the time line interval, the more a memory is embedded in the evolution of the processes' parameters it seeks to represent. These elements lead to difficult problems that seek to account for theories and their models are constructed to negate their effects. For example, in fundamental finance option prices are defined strictly in terms of a model (rational expectations) that has the current option price to be an expectation (under a specific probability measure) of a future price. The past may or may not be relevant to such expectations is then simply done away with. Modeling memory is therefore both a challenging modeling and analytical problem. We circumvent its challenges by a mathematical transformation of the underlying data we have, by a dimensional expansion of an underlying model we construct and by endogenizing the elements that we use to define what we "mean" by memory and how future states and expectations are applied to define their present consequences. The problems we use in this paper outline and compares some of these approaches using the simple binomial model.

2 Financial Models and Memory

Modeling memory processes takes on many form. The following examples outline their differences.

1. No memory in which the past and the future have no effect on current states. Future estimates are then defined in terms of a filtration which summarizes all the relevant information relative to that state. In this sense, "memory" is embedded on the definition of a filtration changing over time as information is revealed, accumulted and interpreted.
2. Markovian memory where all past states are summarized by the last state attained. Their incremental occurrence may therefore be statistically independent.
3. Statistical and Bayesain memory. These models use Bayes Theorem to update future state estimates as new information is revealed. In this sense, it provides a statistical definition for a "new filtration" based on data sets defined at any particular instant of time.
4. Long run Memory underlie a family of models where the volatility of underlying processes does not grow linearly but increases over time at a nonlinear rate. Models such as Levy processes, limit infinite variance processes etc. are such models. These models are not considered in this paper.
5. Short memory processes unlike previous memory models are "regime swtiching" stochastic models reacting to observed information. A variety of derived models may be constructed to better represent (using appropriate data and statistical estimation techniques) the causal factors underlying the evolution of events. Unlike, Bayesian models, short run models are information specific that trigger a changing process rather than its statistical estimates.

For example, does an increase in a market price depend statistically on a previous growth? Does a current market growth correlate with a series of price increases? Such dependence, if existing, is often used by traders to predict market increases and trade on a momentum of prices. Dependence is then endogenous, expressing the time behavior of assets, responding to their "memory", namely their past history of growth and decline in stock prices. Such dependence leads to processes that can be called persistent, anti-persistent. In the first case, the occurrence of a random event determines the random occurrence of subsequent events. When these events are self-reinforcing, we shall call these "persistent" when they are counter reinforcing (for example, as it is presumed in mean reverting processes), these will be called "anti-persistent. In some models, extensive use of a Bayesian approach is used, where information continuously updates prior estimates.

Below, we consider first how memory is accounted for in fundamental finance. This is the case, since financial prices are defined by an equilibrium implied in the exchange of buyers and sellers in financial markets, assumed "complete" and thereby defined in terms of certain conditions we shall elaborate below. In particular, the Arrow-Debreu fundamental model for assets pricing is based on the assumption that at any equilibrium time, the future is predictable and therefore its future well defined by a state and its price. The models we use are then extended within binomial models.

3 The Arrow-Debreu Model and Memory

Memory models (or a lack of it) are implied in both financial and econometric models. It assumes many forms spanning synthetic probability models where probability of future states are an artificial construct (as it is the case in the Arrow-Debreu, definition of risk neutral distributions), random walk models where events occur independently of their past unless structured in a parametric and mathematical model, Bayesian and adaptive models where information is continuously translated into a "statistical memory", long-run and short run memories etc.

In an Arrow-Debreu pricing model for example, future states, say 2 such states $S_i,\ i = 1,\ 2$ are priced explicitly by investors' exchange—some buying a future state and some selling that future state. In equilibrium, each state has then price, say $\pi_i, \quad i = 1,\ 2$. Over a "complete set of potential future states", all potential and future states define a predictable market which is a future risk model, unlike state uncertainty (in a Knight sense) where future states may be unknown. Such a situation is one aspect (future states predictability) that differentiates between risk and uncertainty models in finance and thus, complete and incomplete markets. For example, expectation that a stock price increases by a fixed rate to S_1 with $S_1 = S_0(1 + h)$ or just decrease to S_2 by a fixed loss rate, or $S_2 = S_0(1 - \ell)$ defines a 2-states predictable future. In addition, the Arrow-Debreu model (as it is the case in all complete markets finance models), assumes a market equilibrium for each future state implying that the demand and supply for a stock future state define uniquely its price. Thus, in the Arrow-Debreu model, buying these future states for S_0 initially means that an exchange occurred, one paying S_0 and the other selling the future prospect at $S_1\pi_1 + S_2\pi_2$ (since once an investor has paid S_0 he owns as well the future states and therefore, $S_0 = S_1\pi_1 + S_2\pi_2$). Any additional information provides an opportunity to reduce the parameters defined by such a pricing model. Explicitly, define $\pi = \pi_1 + \pi_2$ and let,

$$S_0 = \pi\left(S_1\frac{\pi_1}{\pi} + S_2\frac{\pi_2}{\pi}\right) = \pi(S_1 p_1 + S_2(1 - p_1)).$$

Since there is no risk and uncertainty in such transactions and since the future price occurs a period later, its present price, discounted at a risk free rate R_f (since there is no risk and since the price of such an investment is only 1 \$) the equation above is equivalent to earning the risk free rate at all state prices, or:

$$1 = \pi((1 + R_f)p_1 + (1 + R_f)(1 - p_1)) = \pi(1 + R_f)(p_1 + (1 - p_1)) = \pi(1 + R_f)$$

And $1/(1 + R_f) = \pi$ which leads to:

$$S_0 = \pi(S_1 p_1 + S_2(1 - p_1)) = \frac{1}{1 + R_f}(S_1 p_1 + S_2(1 - p_1)),$$

$$p_1 = \frac{\pi_1}{\pi_1 + \pi_2} \quad \text{and} \quad p_2 = \frac{\pi_2}{\pi_1 + \pi_2}$$

In equilibrium, *one pays what one wants to get and both the buyer and the seller have agreed on the unique price of their exchange.* Therefore if we were to pay instead B_0 dollar and obtain for sure some amount, B_1, then $B_0 = B_1\pi_1 + B_1\pi_2 = B_1(\pi_1 + \pi_2)$, thus $\pi_1 + \pi_2 = B_0/B_1$. As a result we obtain a relative pricing model—pricing one asset (say a stock) relative to another (a bond). The price of the stock is thus conditional on "information" which is in this case, expressed by the price of a risk free asset, or:

$$S_0 = \frac{B_0}{B_1}(S_1 p_1 + S_2 p_2) \quad \text{or} \quad p_1 = \frac{1}{S_1 - S_2}\left(S_0 \frac{B_1}{B_0} - S_2\right)$$

Further, if $C_i = Max(S_i - K, 0),\ i = 1,\ 2$ is the price of a call option, then, the current price of such an option is:

$$C_0 = \frac{B_0}{B_1}(C_1 p_1 + C_2 p_2) = \frac{B_0}{B_1}(Max(S_1 - K, 0)p_1 + Max(S_2 - K, 0)p_2)$$

And therefore,

$$C_0 \frac{B_1}{B_0} = \left(Max(S_1 - K, 0)\frac{1}{S_1 - S_2}\left(S_0 \frac{B_1}{B_0} - S_2\right) + Max\,(S_2 - K, 0)\,\frac{1}{S_1 - S_2}\left(S_1 - S_0 \frac{B_1}{B_0}\right)\right)$$

Of course, if a stock price, its current call option price and say a put option whose exercise price is Q, then we have a system of two equations in the two implied future states. In other words, the second equation is:

$$P_0 \frac{B_1}{B_0} = \left(Max(Q - S_1, 0)\frac{1}{S_1 - S_2}\left(S_0 \frac{B_1}{B_0} - S_2\right) + Max\,(Q - S_2, 0)\,\frac{1}{S_1 - S_2}\left(S_1 - S_0 \frac{B_1}{B_0}\right)\right)$$

In this sense, the Arrow-Debreu is an endogenous model with parameters defined by the financial information which investors or financial analysts have.

The connotation “probability” to p_1 and p_2 is therefore artificial. In this sense, there is no memory and there is no future but just a present summarized by exchanges in a market which is always in equilibrium and fair—with one price for the two states (S_1, S_2), and a price relative to one state asset whose price is a risk free bond whose future price is B_1. As a result, the existence of memory or memory models that do not replicate the basic exchange outlined above (albeit in far more complex situations) are therefore pricing models that do not conform to the fundamental finance theory of pricing. Such models are therefore off-equilibrium prices. For example, in markets with few and powerful (rich) financial agents (e.g. Big Banks, Big Funds and Big Financial Firms), the interactions as well as the weight each of these agents can impose on financial markets will alter the price that each agents will be paying for financial assets. In this case, prices are not unique and markets are “incomplete”.

Finance in the “real world” is in fact besieged by many future states, some unknown, some known but cannot be priced, etc. and cannot therefore replicate the kind of exchange that occur as stated above which will lead to a unique (and equilibrium) price. Financial analytics consists then in reconciling real finance with its theoretical underpinning—either exactly (in which case, our financial models define what we call complete markets) or approximately, in which case they “transform” what is an incomplete market financial model, into an approximate one. In these cases, since market prices exist in fact “only in the present”, memory and anticipation of future states are mechanisms that reduce both past and future states into a single present one (see Tapiero 2013).

Example: Pricing in a Complete Market The following example outlines this approach. No memory in a modeling sense means that two subsequent events are independent of one another. In addition, financial models assume the predictability of future states. For example, let a stock price be given by future states modeled as a lognormal model:

$$\frac{d\tilde{S}(t)}{\tilde{S}(t)} = d\tilde{R}(t) \text{ where } d\tilde{R}(t) = \alpha dt + \sigma dW(t),\ S(0) = S_0 > 0$$

Where $W(t)$ is a Brownian motion, consisting of independently and identically standard normally distributed random variables (of mean 0 and variance 1). In this model, the price at time $(t + dt)$, $S(t + dt)$, depends necessarily on the price of at time t and is therefore “dependent” on its previous value. This occurs in two ways, first due to a trend rate α and due to the random occurrence of the Brownian motion $\tilde{W}(t)$ whose stock price multiplying factor is the volatility σ. However the stock price rate of return, defined by $d\tilde{R}(t) = \frac{d\tilde{S}(t)}{\tilde{S}(t)}$ is necessarily memory-less since rates of returns are independent and time linear with expected normal rates of returns with mean αt and variance $\sigma^2 t$. Such a model, called the Lognormal model for asset prices underlies many financial applications withy complete markets. For example, it is easy to show by Ito’s calculus that:

$$\ln \tilde{S}(t) = \left(\alpha - \frac{1}{2}\sigma^2\right) t + \sigma W(t),\ S(0) = S_0 > 0 \text{ and } \tilde{S}(t) = S_0 e^{(\alpha - \frac{1}{2}\sigma^2)t + \sigma W(t)}$$

Adding and subtracting a risk free rate payment, we have by definition:

$$\tilde{S}(t) = S_0 e^{\left(-R_f+\alpha-\frac{1}{2}\sigma^2+R_f\right)t+\sigma W(t)} = S_0 e^{\left(+R_f-\frac{1}{2}\sigma^2\right)t+\sigma W(t)+(\alpha-R_f)t}$$
$$= S_0 e^{\left(+R_f-\frac{1}{2}\sigma^2\right)t+\sigma\left(W(t)+\left(\frac{\alpha-R_f}{\sigma}\right)t\right)}$$

Thus, interpreting the risk premium rate ($\frac{\alpha-R_f}{\sigma}$) that one pays to remove the effect of future price risk (however not to remove the underlying future price risks, and therefore it is merely an insurance price against future returns risks). We may define a "riskless probability model", (in the sense that the price of risk has been accounted for by the appropriate probability measure selected) which we shall call a probability measure Q with:

$$W^Q(t) = W(t) + \left(\frac{\alpha - R_f}{\sigma}\right)t$$

Thus, with respect to this probability measure the following pricing model results:

$$\tilde{S}(t) = S_0 e^{\left(R_f-\frac{1}{2}\sigma^2\right)t+\sigma W^Q(t)} \quad \text{or} \quad \frac{d\tilde{S}(t)}{\tilde{S}(t)} = R_f dt + \sigma dW^Q(t), S(0) > 0$$

This probability measure (coined the Q probability measure) defines a Martingale. Namely, a price process with the following (stable and equilibrium) quantitative pricing model:

$$e^{-R_f(0)} S_0 = e^{-R_f*(1)} E^Q \tilde{S}(1) = \ldots = e^{-R_f*(t)} E^Q \tilde{S}(t)$$

This corresponds to the binomial case treated earlier, defining a pricing Martingale with respect to the state prices $p_1 = \frac{\pi_1}{\pi_1+\pi_2}$ and $p_2 = \frac{\pi_2}{\pi_1+\pi_2}$ with $(p_1,\ p_2)$ defining the probabilities of this probability measure. In this context,

$$e^{-R_f*t} E^Q \tilde{S}(t) = \ldots = e^{-R_f*(T)} E^Q \tilde{S}(T) \quad \text{or} \quad E^Q \tilde{S}(t) = e^{-R_f(T-t)} E^Q \tilde{S}(T)$$

However, since at time t, we can observe uniquely the price, $\tilde{S}(t) \equiv S(t)$ and as a result,

$$S(t) = e^{-R_f(T-t)} E^Q \left\{ \tilde{S}(T) \,|\Im_t \right\}$$

Where $\Im_t$ denotes a filtration, with expectation at time t conditional on such filtration. A filtration stands for all the information available and commonly shared at time t. In this sense, it embeds past information as well which has led financial agents to act on financial markets to buy or to sell and therefore reach a price equilibrium which is implied by the probability measure Q, named in this special case (as it is defined with respect and based on the information about the risk free bond market) a "risk neutral probability measure". The Arrow-Debreu thus defines an important pricing approach with a model which is utility free and therefore based on an implied behavior of financial markets rather than explicit financial agents. Financial analytic pricing models are then designed to account for deviations from such a model, but yet be approximately consistent with such an approach.

4 Financial Prices, Filtration and Memory

Financial prices at time t are defined conditionally on information available at time t. When time changes, the information-filtration may of course change and therefore, future prices are necessarily revised to reflect a current and observed price to maintain the completeness of the financial model. Consider two instants of time, then:

$$S(t) = e^{-R_f(1)} E^Q \left\{ \tilde{S}(t+1) \,|\Im_t \right\} = \ldots = e^{-R_f(T-t)} E^Q \left\{ \tilde{S}(T) \,|\Im_t \right\}$$
$$S(t+1) = e^{-R_f(1)} E^Q \left\{ \tilde{S}(t+2) \,|\Im_{t+1} \right\} = \ldots = e^{-R_f(T-(t-1))} E^Q \left\{ \tilde{S}(T) \,|\Im_{t+1} \right\}$$

Note that the filtration has changed from time t to time t + 1. Further, the price at time t + 1 need not be defined by states defined at time t and therefore, the price of a state at time t measured with respect to states defined at time t + 2 and the price at time t + 1 defined with respect to the states at time t + 2 need not be identical. A filtration, resulting from both a re-definition of future states and a "new economic environment" contribute to changing prices. In fact, such a "model revision" depends on the information one use and how we use it. Since future state prices and economic news and information alter investors decisions, past information combined with a new one, imply that the filtration of an Arrow-Debreu pricing model implies a memory of some sort. In this sense, observed prices $S(t)$ and $S(t+1)$ for a future price $S(t+2)$ are not necessarily based on the same "financial model" (since they are not based on the same information regarding the states and the prices of $\tilde{S}(t+2)$).

$$S(t) = e^{-R_f(2)} E^Q \left\{ \tilde{S}(t+2) \,|\Im_t \right\} \text{ and } S(t+1) = e^{-R_f(1)} E^Q \left\{ \tilde{S}(t+2) \,|\Im_{t+1} \right\}$$

For example, in a binomial model, we have the "risk neutral probabilities":

$$p_{1,t}^Q = \frac{\pi_{1,t}}{\pi_{1,t} + \pi_{2,t}} \quad \text{and} \quad p_{2,t}^Q = \frac{\pi_{2,t}}{\pi_{1,t} + \pi_{2,t}} \quad \text{at time t with state prices} \quad (\pi_{1,t}, \pi_{2,t})$$

where $\pi_{1,t} \equiv (\pi_{1,t} \,|\Im_t)$, $\Im_t \equiv \{S_{t-1}, S_{t-2}, \ldots\ldots\}$ is a state defined conditionally on some filtration (information) which is necessarily past, future, or external to the process. A pricing model that simplify its definition is thus based on specific assumption regarding the information we use and how it is used. Given the standardization of these states into risk neutral probabilities, we note that past prices are implied in the risk neutral probability since:

$$p_{1,t}^Q \equiv \left(p_{1,t}^Q \,|S_{t-1},\ S_{t-2}, \ldots \right) \quad \text{as well as} \quad p_{1,t+1}^Q \equiv \left(p_{1,t+1}^Q \,|S_t, S_{t-1},\ S_{t-2}, \ldots \right)$$

And at time t + 1

$$\left(p_{1,t+1}^Q \,|S_t, S_{t-1},\ S_{t-2}, \ldots \right) = \frac{\pi_{1,t+1}}{\pi_{1,t+1} + \pi_{2,t+1}} \quad \text{with state prices} \quad (\pi_{1,t+1}, \pi_{2,t+1})$$

And therefore, at time t:

$$S(t) = \frac{1}{1+R_f} \left(p_{1,t}^Q S_h(t+1) + p_{2,t}^Q S_\ell(t+1) \right)$$

Setting, $S_h(t+1) = S(t)(1+h_{t+1}),\ S_\ell(t+1) = S(t)(1-\ell_{t+1})$, we have:

$$1 = \frac{1}{1+R_f}\left(p_{1,t}^{Q}(1+h_{t+1}) + p_{2,t}^{Q}(1-\ell_{t+1})\right), \quad \text{with filtration } \Im_t \text{ at time t}$$

$$1 = \frac{1}{1+R_f}\left(p_{1,t+1}^{Q}(1+h_{t+2}) + p_{2,t+1}^{Q}(1-\ell_{t+2})\right), \quad \text{with filtration } \Im_{t+1} \text{ at time t+1}$$

Or

$$p_{1,t}^{Q} = \frac{R_f+\ell_{t+1}}{h_t+\ell_{t+1}},\ p_{1,t+1}^{Q} = \frac{R_f+\ell_{t+2}}{h_{t+2}+\ell_{t+2}}, \quad \text{with filtrations } \Im_t,\ \Im_{t+1}$$

Of course, if additional information regarding, say option prices with known strike prices at time t + 2, then at times t and t + 1, we have:

$$C_t\left(\tilde{S}_{t+1}, K_1\right) = \frac{1}{1+R_f}\left(p_{1,t}^{Q} C_{t+1}\big(S_t(1+h_{t+1}), K_1\big) + p_{2,t}^{Q} C_{t+1}\big(S_t(1-\ell_{t+1}), K_1\big)\right)$$

Replacing the implied risk neutral probability, we have two equations in two states at time t + 1, h_{t+1} and ℓ_{t+1} denoting the rate of returns (whether gain or loss) at time t + 1

$$C_t\left(\tilde{S}_{t+1}, K_1\right) = \frac{1}{1+R_f}\left\{\begin{array}{l}\dfrac{R_f+\ell_{t+1}}{h_{t+1}+\ell_{t+1}}\big(C_{t+1}\left(S_t\left(1+h_{t+1}\right),\ K_1\right) \\ -\,C_{t+1}\left(S_t\left(1-\ell_{t+1}\right), K_1\right)\big) \\ +\,C_{t+1}\left(S_t\left(1-\ell_{t+1}\right), K_1\right)\end{array}\right\}$$

$$C_t\left(\tilde{S}_{t+1}, K_2\right) = \frac{1}{1+R_f}\left\{\begin{array}{l}\dfrac{R_f+\ell_{t+1}}{h_{t+1}+\ell_{t+1}}\big(C_{t+1}\left(S_t\left(1+h_{t+1}\right), K_2\right) \\ -\,C_{t+1}\left(S_t\left(1-\ell_{t+1}\right), K_2\right)\big) \\ +\,C_{t+1}\left(S_t\left(1-\ell_{t+1}\right), K_2\right)\end{array}\right\}$$

where K_1 and K_2 are two strikes of two call options we have used in this example. Of course given the future states ℓ_{t+1} and h_{t+1}, the risk neutral probabilities in the underlying asset are well defined and theoretical option prices at time t can be calculated. At time t + 1, an instant of time later we have similarly,

$$C_{t+1}\left(\tilde{S}_{t+2}, K_1\right) = \frac{1}{1+R_f}\left\{\begin{array}{l}\dfrac{R_f+\ell_{t+2}}{h_{t+2}+\ell_{t+2}}\big(C_{t+2}\left(S_{t+1}\left(1+h_{t+2}\right), K_1\right) \\ -\,C_{t+2}\left(S_{t+1}\left(1-\ell_{t+2}\right), K_1\right)\big) \\ +\,C_{t+2}\left(S_{t+1}\left(1-\ell_{t+2}\right), K_1\right)\end{array}\right\}$$

$$C_{t+1}\left(\tilde{S}_{t+2}, K_2\right) = \frac{1}{1+R_f}\left\{\begin{array}{l}\dfrac{R_f+\ell_{t+2}}{h_{t+2}+\ell_{t+2}}\big(C_{t+2}\left(S_{t+1}\left(1+h_{t+2}\right), K_2\right) \\ -\,C_{t+2}\left(S_{t+1}\left(1-\ell_{t+2}\right), K_2\right)\big) \\ +\,C_{t+2}\left(S_{t+2}\left(1-\ell_{t+2}\right), K_2\right)\end{array}\right\}$$

Since, $p_{1,t}^{Q} = \frac{R_f+\ell_{t+1}}{h_{t+1}+\ell_{t+1}}$ and $p_{1,t+1}^{Q} = \frac{R_f+\ell_{t+2}}{h_{t+2}+\ell_{t+2}}$, the implied risk neutral probabilities $(p_{1,t}^{Q}, p_{1,t+1}^{Q})$ may define an autocorrelation and therefore a memory anchored in their common information or in the use of common information to estimate the future states at time t + 1 and at time t + 2. Is there then a relationship between the probabilities $p^Q_{1,t}$ and $p^Q_{1,t+1}$? While, $p^Q_{1,t}$ is implied in the known future states (h_{t+1}, ℓ_{t+1}) and its filtration at time t + 1, these same states at time t are embedded in a filtration at time t − 1. As a result, we can consider these elements as probabilities implied in our temporal observations of time, providing a series of probabilities $p^Q_{1,1}, p^Q_{1,2}, p^Q_{1,3}, \ldots p^Q_{1,t}$ implied by the information we have regarding these prices. Given a set of prices, are these probabilities dependent? Do they reveal a trend? Is there an autocorrelation between these probabilities (that would reveal an inter-temporal dependence)? Etc. Such a problem is considered in the next section. Below a number of extensions are considered. These include the effects of filtration (additional information at a given time) and short memory as will be developed subsequently) that leads to a price process as a function of past information. Such approaches are then based on the statistical realization of information which contributes to the adjustment of financial markets to be consistent with the Arrow-Debreu equilibrium model. A specific case is considered below which assumes that information is defined by the proportional number of times a stock price has increased or decreased (or both) consecutively over a given time period. Of course, if a series increases or decreases is random, the probability that consecutive "risk neutral" probabilities are dependent is null and vice versa. Namely, consecutive risk neutral probabilities are dependent.

5 A Multivariate Pricing Models with Short Memory

For simplicity, let $\tilde{p}^Q_t \in \{1,0\}$ with $p^Q_{1,t}$ the probability that the stock price has increased (or equal 1) and $p^Q_{2,t} = 1 - p^Q_{1,t}$. We define similarly $\tilde{p}^Q_{t+1} \in \{1,0\}$ and therefore assume that $\left\{\tilde{p}^Q_t, \tilde{p}^Q_{t+1}\right\}$are two random variables defined by a bi-variate Bernoulli process. Consider next two consecutive periods. There are then four states $\langle\{1,1\},\{1,0\},\{0,1\},\{0,0\}\rangle$. If these probabilities are statistically independent, their joint distribution is their product $(\tilde{p}^Q_t \tilde{p}^Q_{t+1})$. If they are dependent, then they define a bi-variate bernoulli joint probability distribution (and evidently, for n samples, these define a bi-variate binomial probability distribution):

$$\wp\left(\tilde{p}^Q_t, \tilde{p}^Q_{t+1}\right) = \begin{cases} \tilde{p}^Q_t = 1, \tilde{p}^Q_{t+1} = 1 & \text{probability } p_{11} \\ \tilde{p}^Q_t = 1, \tilde{p}^Q_{t+1} = 0 & \text{probability } p_{10} \\ \tilde{p}^Q_t = 0, \tilde{p}^Q_{t+1} = 1 & \text{probability } p_{01} \\ \tilde{p}^Q_t = 0, \tilde{p}^Q_{t+1} = 0 & \text{probability } p_{00} \end{cases}$$

Or,

	t + 1 Price increase	t + 1 Price decrease
t Price increase	p_{11}	p_{10}
t price decrease	p_{01}	p_{00}

While their joint distribution is:

$$\wp(\tilde{p}_t^Q, \tilde{p}_{t+1}^Q) = \{p_{11}\}^{\tilde{p}_t^Q \tilde{p}_{t+1}^Q}\{p_{10}\}^{\tilde{p}_t^Q(1-\tilde{p}_{t+1}^Q)}\{p_{01}\}^{(1-\tilde{p}_t^Q)\tilde{p}_{t+1}^Q}\{p_{00}\}^{(1-\tilde{p}_t^Q)(1-\tilde{p}_{t+1}^Q)},$$
$$\text{with } (\tilde{p}_t^Q, \tilde{p}_{t+1}^Q) \in [1,0] \text{ and } p_{11} + p_{10} + p_{01} + p_{00} = 1$$

The probability generating function is $P^*(z_1, z_2) = \sum_{\tilde{p}_t^Q=1,0} \sum_{\tilde{p}_{t+1}^Q=1,0} \wp(\tilde{p}_t^Q, \tilde{p}_{t+1}^Q) z_1^{\tilde{p}_t^Q} z_2^{\tilde{p}_{t+1}^Q}$ and therefore,

$$P^*(z_1, z_2) = p_{00} + p_{10}z_1 + p_{01}z_2 + p_{11}\, z_1 z_2$$

The moments of such a distribution are thus estimated by deriving the PGF at $z_1 = 1, z_2 = 1$ leading to:

$$E(\tilde{p}_t^Q) = p_{10} + p_{11}, \operatorname{var}(\tilde{p}_t^Q) = (p_{10} + p_{11})(1 - (p_{10} + p_{11}))$$
$$E(\tilde{p}_{t+1}^Q) = p_{01} + p_{11}, \operatorname{var}(\tilde{p}_{t+1}^Q) = (p_{01} + p_{11})(1 - (p_{01} + p_{11}))$$

and $E(\tilde{p}_t^Q \tilde{p}_{t+1}^Q) = p_{11}$ or $p_{11} = 1 - p_{00} - p_{10} - p_{01}$. Further, setting $1 - p_{00} - p_{01} = p_1$ and $1 - p_{00} - p_{10} = p_2$ we obtain two Bernoulli distributions:

$$E(\tilde{p}_t^Q) = p_1 = 1 - p_{00} - p_{01}, \operatorname{var}(\tilde{p}_t^Q) = p_1(1 - p_1)$$
$$E(\tilde{p}_{t+1}^Q) = p_2 = 1 - p_{00} - p_{10}, \operatorname{var}(\tilde{p}_{t+1}^Q) = p_2(1 - p_2)$$

As well as: $E(\tilde{p}_t^Q \tilde{p}_{t+1}^Q) = p_{11} = E(\tilde{p}_t^Q)E(\tilde{p}_{t+1}^Q) + \rho\sqrt{\operatorname{var}(\tilde{p}_t^Q)\operatorname{var}(\tilde{p}_{t+1}^Q)}$ where ρ is the correlation of the two implied risk neutral probabilities at time t and t + 1. Or,

$$\rho = \frac{p_{11} - (1 - p_{00} - p_{01})(1 - p_{00} - p_{10})}{\sqrt{(1 - p_{00} - p_{01})(p_{00} + p_{10})(1 - p_{00} - p_{10})(p_{00} + p_{10})}}$$

Thus, $E(\tilde{p}_t^Q \tilde{p}_{t+1}^Q) = p_{11} = p_1 - p_{10} = p_2 - p_{01} = p_1 p_2 + \rho\sqrt{p_1 p_2(1 - p_1)(1 - p_2)}$ and finally:

$$\operatorname{cov}(\tilde{p}_t^Q, \tilde{p}_{t+1}^Q) = \rho - p_1 p_2 = \rho - (1 - p_{00} - p_{01})(1 - p_{00} - p_{10})$$

Assume for example, over a10 periods, the following record of increases and decreases in a stock price,

$$\langle\{1,0\},\{0,1\},\{1,1\},\{1,1\},\{1,1\},\{1,0\},\{0,1\},\{1,0\},\{0,0\},\{0,1\}\rangle$$

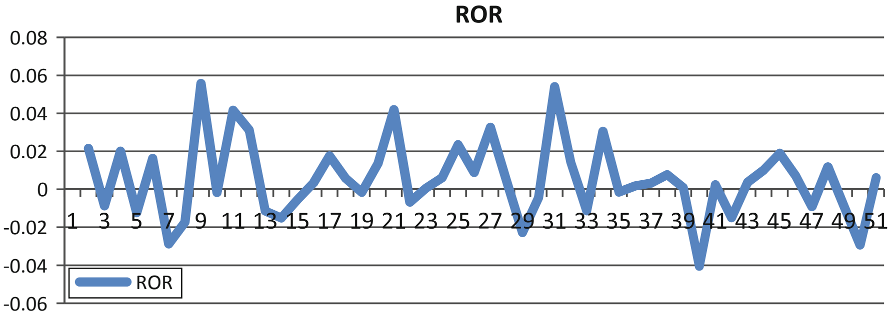
Or

	t + 1 Price increase	t + 1 Price decrease
t Price increase	$p_{11} = 0.3$	$p_{10} = 0.3$
t price decrease	$p_{01} = 0.3$	$p_{00} = 0.1$

The average time the price increased or decreased are equal and equal 0.4 while: $\rho = -0.25$.

Similarly, we have considered a number of stocks and indexes over various periods of time and data sets and have obtained the following results, indicating mostly a correlation between the probabilities of consecutive, increase or decrease.

For Apple's stock rate of return over a given period in 2005, we used the following rates of returns



Consecutive increases and decreases indicated (for 100 days) the following statistic 01 11 10. . . with

	t + 1 Price increase	t + 1 Price decrease
t Price increase	$p_{11} = 0.333$	$p_{10} = 0.2745$
t price decrease	$p_{01} = 0.29411$	$p_{00} = 0.09803922$

Over 100 days in APPLE's stock in 2005.

In this case,

$$E(\tilde{p}_t^Q) = p_1 = 1 - 0.098 - 0.2941 = 0.6079, \text{var}\,(\tilde{p}_t^Q) = 0.2383$$

$$E(\tilde{p}_{t+1}^Q) = p_2 = 1 - 0.098 - 0.2741 = 0.6279, \text{var}\,(\tilde{p}_{t+1}^Q) = 0.2336$$

And $E(\tilde{p}_t^Q \tilde{p}_{t+1}^Q) = p_{11} = 0.3334 = (0.6079)(0.6279) + \rho\sqrt{(0.2383)(0.2336)}$, and therefore, a negative correlation $\rho = -0{,}2046$.

Similar results, indicating a dependence are obtained using data on a variety of indexes using different time periods. The results below summarize a number of estimates on both indexes and stocks using intraday and day data. IBM intraday data in 1, 5 and 10 min indicates that dependence is more pronounced in 1 min data while it is reduced as the time interval increases. This observation reinforces the advantages that intraday high frequency traders have over day traders or long investors.

	t + 1 Price increase	t + 1 Price decrease
IBM: Intraday data, 1, 5 and 10 min		
t Price increase	$p_{11}(1\,\text{min}) = 0.4437$	$p_{11}(5\,\text{min}) = 0.4694$
	$p_{11}(10\,\text{min}) = 0.5067$	$p_{10}(1\,\text{min}) = 0.5562$
	$p_{10}(5\,\text{min}) = 0.5305$	$p_{10}(10\,\text{min}) = 0.4932$
t Price decrease	$p_{01}(1\,\text{min}) = 0.4989$	$p_{01}(5\,\text{min}) = 0.5373$
	$p_{01}(10\,\text{min}) = 0.5034$	$p_{00}(1\,\text{min}) = 0.5010$
	$p_{00}(5\,\text{min}) = 0.4626$	$p_{00}(10\,\text{min}) = 0.4965$
SPX index: day and weekly data		
t Price increase	$p_{11}(1\text{day}) = 0.4616$	$p_{11}(1\text{week}) = 0.5722$
	$p_{10}(1\text{day}) = 0.5188$	$p_{10}(1\text{week}) = 0.4227$
t price decrease	$p_{01}(1\text{day}) = 0.5846$	$p_{01}(1\text{week}) = 0.477$
	$p_{00}(1\text{day}) = 0.4153$	$p_{00}(1\text{week}) = 0.5230$
Shanghai composite index: 1 day		
t Price increase	$p_{11} = 0.5866$	$p_{10} = 0.41338$
t Price decrease	$p_{01} = 0.5341$	$p_{00} = 0.4658$
Aapp intraday 1 min and 5 min data		
t Price increase	$p_{11}(1\,\text{min}) = 0.456$	$p_{11}(5\,\text{min}) = 0.4911$
	$p_{10}(1\,\text{min}) = 0.544$	$p_{10}(5\,\text{min}) = 0.5088$
t Price decrease	$p_{01}(1\,\text{min}) = 0.5156$	$p_{01}(5\,\text{min}) = 0.5112$
	$p_{00}(1\,\text{min}) = 0.4843$	$p_{00}(5\,\text{min}) = 0.4887$

Risk neutral probability distributions (in the binomial model we have used) thus imply future states prices and therefore the belief of a typical agent, representative of the financial market, to know these two states and thus their proportional value. Further it assumes as well that all agents in a financial market will buy or sell at the same expected price.

Example: Extending the Memory to Two Periods Extending the memory to two periods leads to a Markov model defined by the following matrix for which appropriate probabilities at time t are defined conditional on a price increasing or decreasing next:

	(1)++	(2)−+	(3)+−	(4)−
(1)++	$p^t_{1,1}$	0	$1 - p^t_{1,1}$	0
(2)−+	$p^t_{2,1}$	0	$1 - p^t_{2,1}$	0
(3)+−	0	$p^t_{3,2}$	0	$1 - p^t_{3,2}$
(4)−	0	$p^t_{4,2}$	0	$1 - p^t_{4,2}$

Where ++ means that the price has increased in the subsequent two periods while + − means that the price has first increased and the decrease in the previous two periods. Thus, $p^t_{1,1}$ is the probability that the price will continue to increase while $p^t_{3,1} = 1 - p^t_{1,1}$ is the probability that it will decrease at the subsequent period. For simplicity, assume a stationary state. Then the (long run) probability to be in any of these information states $p^t_1, p^t_2, p^t_3, p^t_4$ with $\sum_{i=1}^{4} p_i = 1$ are given by the ergodic probabilities using the current state transition probabilities $p^t_{1,1}$, $p^t_{2,1}$, $p^t_{3,2}$ and $p^t_{4,2}$. At the limit, the long run probabilities are given by (in our example case, by setting $p^{t+1}_i = p^t_i = \hat{p}_i$ and therefore

$$p_1^{t+1} = p_1^t p_{11}^t + p_2^t p_{21}^t;\ p_2^{t+1} = p_3^t p_{32}^t + p_4^t p_{42}^t$$
$$p_3^{t+1} = p_1^t \left(1 - p_{11}^t\right) + p_2^t \left(1 - p_{21}^t\right),\ p_4^{t+1} = p_3^t \left(1 - p_{32}^t\right) + p_4^t \left(1 - p_{42}^t\right)$$

and thus (with $p_4 = 1 - p_1^t - 2p_2^t$):

$$p_1^t = \left(\frac{p_{42}^t p_{21}^t}{p_{42}^t p_{21}^t + \left(1 + 2p_{42}^t - p_{32}^t\right)\left(1 - p_{11}^t\right)}\right),$$
$$p_2^t = p_3^t = \frac{p_{42}^t \left(1 - p_{11}^t\right)}{\left(1 + 2p_{42}^t - p_{32}^t\right)\left(1 - p_{11}^t\right) + p_{42}^t p_{21}^t}$$

A binomial pricing model is given in this case by:

$$1 = \frac{1}{1+R_f}\left(p_{11}^t\left(1 + h_{t+1}^1\right) + \left(1 - p_{1,1}^t\right)\left(1 - \ell_{t+1}^1\right)\right),$$
$$1 = \frac{1}{1+R_f}\left(p_{21}^t\left(1 + h_{t+1}^2\right) + \left(1 - p_{21}^t\right)\left(1 - \ell_{t+2}^2\right)\right),$$
$$1 = \frac{1}{1+R_f}\left(p_{32}^t\left(1 + h_{t+1}^3\right) + \left(1 - p_{32}^t\right)\left(1 - \ell_{t+2}^3\right)\right),$$
$$1 = \frac{1}{1+R_f}\left(p_{42}^t\left(1 + h_{t+1}^4\right) + \left(1 - p_{42}^t\right)\left(1 - \ell_{t+2}^4\right)\right)$$

However, since the probability of being in any of these 4 states is given by $p_1^t, p_2^t, p_3^t, p_4^t$ we have:

$$\begin{aligned}1 = {} & \frac{p_1^t}{1+R_f}\left(p_{11}^t\left(1 + h_{t+1}^1\right) + \left(1 - p_{1,1}^t\right)\left(1 - \ell_{t+1}^1\right)\right)\\ & + \frac{p_1^2}{1+R_f}\left(p_{21}^t\left(1 + h_{t+1}^2\right) + \left(1 - p_{21}^t\right)\left(1 - \ell_{t+2}^2\right)\right),\\ & + \frac{p_3^t}{1+R_f}\left(p_{32}^t\left(1 + h_{t+1}^3\right) + \left(1 - p_{32}^t\right)\left(1 - \ell_{t+2}^3\right)\right)\\ & + \frac{p_4^t}{1+R_f}\left(p_{42}^t\left(1 + h_{t+1}^4\right) + \left(1 - p_{42}^t\right)\left(1 - \ell_{t+2}^4\right)\right)\end{aligned}$$

where each of the information states $p_i^t,\ i = 1{,}2,3{,}4$ are given as a function of the transition probabilities as stated above. These probabilities are necessarily dependent as they are defined in terms of common probabilities.

6 A Binomial Model and Bayesian Learning

Assume at present that there are many buyers and sellers that share a common expected risk neutral probability such that at two consecutive times $E(\tilde{p}_t^Q) = \frac{\alpha_t}{\alpha_t+\beta_t}$ and $E(\tilde{p}_{t+1}^Q) = \frac{\alpha_{t+1}}{\alpha_{t+1}+\beta_{t+1}}$ which we assume for simplicity to be approximated by a bivariate Beta probability distribution with variance:

$$\text{var}\left(\tilde{p}_t^Q\right) = \frac{\alpha_t \beta_t}{(\alpha_t + \beta_t)^2(\alpha_t + \beta_t + 1)},$$
$$\text{var}\left(\tilde{p}_{t+1}^Q\right) = \frac{\alpha_{t+1} \beta_{t+1}}{(\alpha_{t+1} + \beta_{t+1})^2(\alpha_{t+1} + \beta_{t+1} + 1)}$$

And therefore using categorical data of consecutive increases or decreases in prices, we have then 4 equations for moments estimates of $\alpha_t, \beta_t, \alpha_{t+1}, \beta_{t+1}$:

$$E(\tilde{p}_t^Q) = 0.6079 = \frac{\alpha_t}{\alpha_t + \beta_t},\ E(\tilde{p}_{t+1}^Q) = 0.6279 = \frac{\alpha_{t+1}}{\alpha_{t+1} + \beta_{t+1}},$$

and

$$\text{var}\,(\tilde{p}_t^Q) = 0.2383 = \frac{\alpha_t \beta_t}{(\alpha_t + \beta_t)^2(\alpha_t + \beta_t + 1)},$$
$$\text{var}(\tilde{p}_{t+1}^Q) = 0.2336 = \frac{\alpha_{t+1} \beta_{t+1}}{(\alpha_{t+1} + \beta_{t+1})^2(\alpha_{t+1} + \beta_{t+1} + 1)}$$

Or, $\beta = \alpha_t(1 - E(\tilde{p}_t^Q))/E(\tilde{p}_t^Q)$ and therefore $\alpha_t = \frac{E(\tilde{p}_t^Q)(1-E(\tilde{p}_t^Q))}{\text{var}\,(\tilde{p}_t^Q)} - 1$. These lead to $\alpha_t = 0.0002416$, $\beta_t = 0.00015611$ and $\alpha_{t+1} = 0.000178$, $\beta_{t+1} = 0.00014317$ with $\rho = -0{,}2046$ which provides an indication to a bi-variate Beta probability distribution. Such distributions are however varied, assuming various forms (see for example, Olkin and Liu 2003; Nadajarah and Kotz 2005 as well as Arnold and Ng 2011 for a number of such distributions as well as their moments estimation).

Generally for a K vector of multivariate Bernoulli probability distributions (note the change in notation) denoting a time series of increases or decreases in a stock price:

$$p(x_1, x_2, \ldots., x_K) = \{p_{11\ldots1}\}^{\prod_{j=1}^{K} x_j} \{p_{011..1}\}^{(1-x_1)\prod_{j=1}^{K} x_j} \ldots\ldots \{p_{00.....0}\}^{\prod_{j=1}^{K}(1-x_j)}$$

Such a multivariate distribution can then be used for a pricing model of a K-length memory time series.

Co-variation in the implied risk neutral probabilities *does not mean however that markets are incomplete, but indicate that the market is changing from period to period due to new information (filtration effects) or potentially, future states that define the whole set of potential future prices.*

If we consider information to be defined by the last past periods, say at t − 1, t and t + 1 we have then 8 states ($2^3 = 8$) probabilities. Their probability generating function, thus:

$$P^*(z_1, z_2, z_3) = \big\{p_{000} + p_{001}z_3 + p_{010}z_2 + p_{011}z_2z_3 + p_{100}z_1 + p_{101}z_1z_3 + p_{110}z_1z_2 + p_{111}z_1z_2z_3\big\}$$

To calculate their estimate, mean, and variance and autocorrelation is a simple exercise. Considering n samples, its PGF is $[P^*(z_1, z_2, z_3)]^n$ which we can use to calculate the moments of their distribution. For the bi-variate case, we obtain a multivariate binomial distribution:

$$B(x_1 = i, x_2 = j) = \sum_{k=0} \frac{n!}{i!(j-i)!(k-i)!(n-j-k+i)!} (\hat{p}_{11})^i (\hat{p}_{10})^{k-i} (\hat{p}_{01})^{k-j} (\hat{p}_{00})^{n-k-j+i}$$

Which calculates the number of times a stock price has assumed consecutive values.

Further, when the probabilities are assumed unknown or random, mixture probabilities models can be constructed as we shall briefly outline using some examples.

Generally for any arbitrary length, we let $(\tilde{x}_1, \tilde{x}_2, \ldots.., \tilde{x}_n)$ be, n bi-variate random variables, defined as follows (see also . . .):

$$\begin{aligned}
&f(\tilde{x}_i = 1 \,|p_{1,i}) = p_{1,i}\\
&f(\tilde{x}_i = 1, \tilde{x}_j = 1 \,|p_{2,ij}) = p_{2,ij}, \quad i, j = 1, \ldots n; \quad i \neq j\\
&\ldots\ldots\\
&f(\tilde{x}_i = 1, \tilde{x}_j = 1, \ldots\ldots, \tilde{x}_n = 1 \,|p_{k,i,j..n}) = p_{k,i,j..n}
\end{aligned}$$

where $\{p_k\}$ is a parameter set (similar to that defined above). Since this is a binary variable, we have:

$$\begin{aligned}
&E(\tilde{x}_i \,|p_{1i}) = p_{1i}, \; E(\tilde{x}_i\tilde{x}_j \,|p_{2,ij}\,) = p_{2,ij};\\
&\operatorname{cov}(\tilde{x}_i, \tilde{x}_j \,|p_{1i}, p_{1j} p_{2,ij}) = p_{2,ij} - p_{1i} p_{1j}; \quad i \neq j
\end{aligned}$$

and

$$\rho_{ij}(\tilde{x}_i, \tilde{x}_j) = \frac{p_{2,ij} - p_{1i} p_{1j}}{\sqrt{p_{1i}(1 - p_{1i}) p_{1j}(1 - p_{1j})}}$$

An Example: A Bayesian Learning Model Say that information is a counting process of the past times a stock price has increased and the number of times it has decreased. In this case, under complete markets and at two consecutive instants of time, we have:

$$\begin{aligned}
S(t) &= \frac{1}{1+R_f} E^Q \left\{\tilde{S}(t+1) \,|\Im_t\right\} = \ldots\ldots = \frac{1}{(1+R_f)^{T-t}} E^Q \left\{\tilde{S}(T) \,|\Im_t\right\}\\
S(t+1) &= \frac{1}{1+R_f} E^Q \left\{\tilde{S}(t+2) \,|\Im_{t+1}\right\} = \ldots\ldots = \frac{1}{(1+R_f)^{T-(t+1)}} E^Q \left\{\tilde{S}(T) \,|\Im_{t+1}\right\}
\end{aligned}$$

Where $\Im_t$ denotes the filtration of past events (the memory) of the number of times r_t an asset has increased in price and therefore, the number of times that it has decreased is $t - r_t$. Therefore, $S(t\,|r_t\,)$is a price which is defined conditionally on the number of events r_t and $t - r_t$. As a result, the future price at time t, the price is:

$$S(t\,|r_t\,) = E_{\tilde{x}_t}\left\{e^{-R_f(1)}E^Q\left\{\tilde{S}(t+1\,|r_t+\tilde{x}_t\,)\right\}\right\},\quad \tilde{x}_t = 1{,}0$$

where

$$\tilde{x}_t = \begin{cases} r_t+1 & w.p.\ \tilde{\theta}_t \\ r_t & w.p.\ 1\text{-}\tilde{\theta}_t \end{cases}$$

where "$\tilde{\theta}_t$" is the probability that the return have increased at time t + 1. Let the expectation $E(\tilde{\theta}_t) = p_t^Q$ denote the risk neutral probability, or:

$$S(t\,|r_t\,) = \left\{\frac{1}{1+R_f}E^Q\left\{E_{\tilde{\theta}_t}\tilde{S}(t+1\,|r_t+1\,)\tilde{\theta}_t + E_{\tilde{\theta}_t}\tilde{S}(t+1\,|r_t\,)(1-\tilde{\theta}_t)\right\}\right\},$$
$$\tilde{x}_t = 1{,}0$$

If $\tilde{S}(t+1\,|r_t+i\,)$ and $\tilde{\theta}_t$ are statistically independent, then, if the probability $\tilde{\theta}_t$ has a Beta probability distribution with parameters "α_t" and "β_t", or: $\tilde{\theta}_t\tilde{\ }B_t(\alpha_t,\beta_t)$ then $p_t^Q = \alpha_t/(\alpha_t,\beta_t)$ and therefore, "α_t" and "β_t" define proportional estimates of the future assets states, or

$$S(t\,|r_t\,) = \left\{e^{-R_f(1)}E^Q\left\{\tilde{S}(t+1\,|r_t+1)\frac{\alpha_t}{\alpha_t+\beta_t} + \tilde{S}(t+1\,|r_t)\frac{\beta_t}{\alpha_t+\beta_t}\right\}\right\}$$

Of course, the parameters α_t and β_t de not define the state preferences. However in expectation they provide an implied risk neutral probability. In this case, as information is revealed, the future probability $\tilde{\theta}_t$ and its distribution may be revised once the observation that a stock price has increased or decreased is recorded.

If a gain occurs, then, the Bernoulli-beta process yields a posterior beta distribution with parameters $B_{t+1}(\alpha_{t+1},\beta_{t+1}) = B_{t+1}(1+\alpha_t,\beta_t)$ However, if a decrease occurs, then the posterior distribution is also beta with $B_{t+1}(\alpha_{t+1},\beta_{t+1}) = B_{t+1}(\alpha_t, 1+\beta_t)$. In this sense, the updating scheme can be represented by the Beta distribution $B_{t+1}(\tilde{\theta}_{t+1}+\alpha_t, 1-\tilde{\theta}_{t+1}+\beta_t), \tilde{\theta}_{t+1} = 1{,}0$. For example, say that we observe the following increases and decreases in prices:

$$(r_t, t-r_t; t, \tilde{\theta}_t = 1); (r_{t+1} = r_t+1, t-(r_t+1), t+1; \tilde{\theta}_{t+1} = 0);$$
$$(r_{t+2} = r_t+2, t+2-(r_t+2), t+2; \tilde{\theta}_{t+2} = 1);\ldots$$

For example, at time t, we observed r_t times that the stocks increased and therefore $t - r_t$times that the stock decreased in this case $\tilde{\theta}_t \sim B_t(\alpha_t,\beta_t)$, and at time t, $\theta_t = 1$. In this case, at the next period t + 1 we have therefore $r_{t+1} = r_t$+1 increases while $t+1-r_{t+1}$ decreases where the probability of an increase has now the Beta probability distribution $\tilde{\theta}_{t+1} \sim B_{t+1}(1+\alpha_t,\beta_t)$. In this case the stock price decreased which leads to the observation $\tilde{\theta}_{t+1} = 0$ and therefore to the next probability $\tilde{\theta}_{t+2} \sim$

$B_{t+1}(1+\alpha_t, 1+\beta_t)$. The following price models with $\tilde{\theta}_t = 1, \tilde{\theta}_{t+1} = 1$ and $\tilde{\theta}_{t+2} = 0$ thus results. In this case, set, at time t + 1, state prices $(\pi_{1,t+1}, \pi_{2,t+1})$ and therefore their risk neutral probabilities are:

$$p_{1,t}^{Q} = \frac{\pi_{1,t+1}(r_t)}{\pi_{1,t+1}(r_t) + \pi_{2,t+1}(t - r_t)} = \frac{\alpha_t}{\alpha_t + \beta_t}$$

and

$$p_{2,t}^{Q} = \frac{\pi_{1,t+1}(\ell_t = t - r_t)}{\pi_{1,t+1}(r_t) + \pi_{2,t+1}(t - r_t)} = \frac{\beta_t}{\alpha_t + \beta_t}$$

Of course as data accumulates over time, say with the following observations for price increases and decreases, we note the following changes in the "risk neutral probabilities":

$$p_{1,t}^{Q} = \frac{\pi_{1,t+1}(r_t)}{\pi_{1,t+1}(r_t) + \pi_{2,t+1}(t - r_t)} = \frac{\alpha_t}{\alpha_t + \beta_t}, \quad \text{Price increased at t, t+1}$$

$$p_{1,t+1}^{Q} = \frac{\pi_{1,t+2}(r_t + 1)}{\pi_{1,t+2}(r_t + 1) + \pi_{2,t+2}(t + 1 - r_t)} = \frac{1 + \alpha_t}{1 + \alpha_t + \beta_t}, \quad \text{Price increased at t+1, t+2}$$

$$p_{1,t+2}^{Q} = \frac{\pi_{1,t+3}(r_t + 2)}{\pi_{1,t+2}(r_t + 2) + \pi_{2,t+2}(t + 2 - r_t)} = \frac{2 + \alpha_t}{2 + \alpha_t + \beta_t}, \quad \text{Price decrease at t+2, t+3}$$

$$p_{1,t+3}^{Q} = \frac{\pi_{1,t+4}(r_t + 2)}{\pi_{1,t+2}(r_t + 2) + \pi_{2,t+2}(t + 3 - (r_t + 1))} = \frac{2 + \alpha_t}{3 + \alpha_t + \beta_t}, \text{Price decrease at t+2, t+3}$$

and so on. Using the data we calculate using the Apple stock rates of returns,

$$1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, \ldots .$$

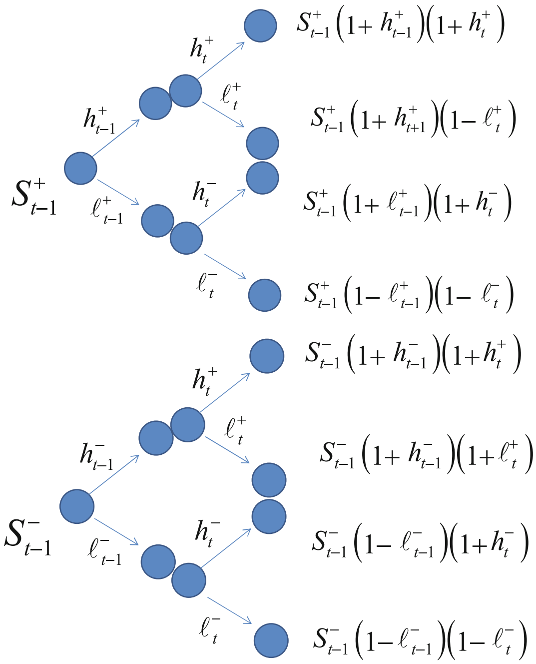
we find the following of parameters estimates, indicating the changing relative state prices defining the risk neutral probability:

$$p_{1,t}^{Q}, t = 0,1,2,3,4,... = \left\{ \begin{array}{l} \dfrac{\alpha_0}{\alpha_0 + \beta_0}; \dfrac{1 + \alpha_0}{1 + \alpha_0 + \beta_0}; \dfrac{2 + \alpha_0}{2 + \alpha_0 + \beta_0}; \dfrac{2 + \alpha_0}{3 + \alpha_0 + \beta_0}; \dfrac{3 + \alpha_0}{4 + \alpha_0 + \beta_0}; \\ \dfrac{3 + \alpha_0}{5 + \alpha_0 + \beta_0}; \dfrac{4 + \alpha_0}{6 + \alpha_0 + \beta_0}; \dfrac{4 + \alpha_0}{7 + \alpha_0 + \beta_0}{}' \dfrac{4 + \alpha_0}{8 + \alpha_0 + \beta_0}; \ldots\ldots \end{array} \right\}$$

Considering the two periods and assuming that α_0 and β_0 are negligible compared to 4, we have, a rough estimate of the risk neutral probabilities $\frac{4+\alpha_0}{7+\alpha_0+\beta_0} \approx 0.5714$ and $\frac{4+\alpha_0}{8+\alpha_0+\beta_0} \approx 0.500\ldots$ and thus to changing patterns in risk neutral probabilities.

7 Arrow-Debreu and Short Memory

Short memory presumes that information can alter future probability processes. Thus, unlike statistical dependence, an underlying process may switch to be a different process based on the information observed at a given time. For example, when counting the number of claims that an insured has over time, there is an increased interest in assessing whether a claim generates more claims, or rather a claim has reduced the insured propensity to claim. Similarly, if a stock price has increased, what are the effects of this increase on it increasing next or decreasing? This observation underlies trading momentum based strategies as well as pair trading strategies. While in the latter case, one generally implies the statistical relationships between two asset prices, they seek in fact to define a causal relationship between one asset and the other. By the same token, a driver who has an accident may be more careful in the year that follows the accident but be less concerned in a subsequent year when there were no accidents to report etc. For example, we consider a two periods and one period memory binomial model with stock prices S_t^+ and S_t^- denoting the time t and the and an index "+" or "−" to point out that in the previous period stock prices increased or decreased. In this case, assume that there are probabilities defined by the bivariate process $(x_{1,t}^Q, x_{2,t}^Q)$ with probabilities $(p_{t,ij}^Q)$. In this case, a two consecutive process is represented graphically (in case at the current time t − 1, the previous outcome has been an increase in price as well as in case the previous period there was a price decrease):



In both cases a representation of these 8 possibilities at time t − 1 are given in the table below:

	$x_{1,t}^Q = 1$	$x_{1,t}^Q = 0$
$\left(x_{1,t-1}^Q = 1\right)$	$\left(p_{t,11}^{Q+}; p_{t,11}^{Q-}\right)$	$\left(p_{t,10}^{Q+}; p_{t,10}^{Q-}\right)$
	$\left(h_{t-1}^+, h_t^+\right); \left(h_{t-1}^-, h_t^+\right)$	$\left(h_{t-1}^+, \ell_t^+\right); \left(h_{t-1}^-, \ell_t^+\right)$
$\left(x_{1,t-1}^Q = 0\right)$	$\left(p_{t,01}^{Q+}; p_{t,01}^{Q-}\right)$	$\left(p_{t,00}^{Q+}; p_{t,00}^{Q-}\right)$
	$\left(\ell_{t-1}^+, h_t^+\right); \left(h_{t-1}^- \ell_t^-\right)$	$\left(\ell_{t-1}^+, \ell_t^-\right); \left(\ell_{t-1}^- \ell_t^-\right)$

and therefore, assuming that there are risk neutral probabilities at time t − 1 to account for both periods starting at S_{t-1}^+ or S_{t-1}^- over two periods, we have:

$$S_{t-1}^+ = \frac{1}{(1+R_f)^2}\begin{Bmatrix} (p_{t,11}^{Q+} S_{t-1}^+ (1+h_{t-1}^+)(1+h_t^+) + p_{t,10}^{Q+} S_{t-1}^+ (1+h_{t-1}^+)(1-\ell_t^+)) \\ (p_{t,01}^{Q+} S_{t-1}^+ (1-\ell_{t-1}^+)(1+h_{t-1}^-) + p_{t,00}^{Q+} S_{t-1}^+ (1-\ell_{t-1}^+)(1-\ell_t^-)) \end{Bmatrix}$$

Or

$$1 = \frac{1}{(1+R_f)^2}\begin{Bmatrix} p_{t,11}^{Q+}(1+h_{t-1}^+)(1+h_t^+) + p_{t,10}^{Q+}(1+h_{t-1}^+)(1-\ell_t^+) \\ p_{t,01}^{Q+}(1-\ell_{t-1}^+)(1+h_{t-1}^-) + p_{t,00}^{Q+}(1-\ell_{t-1}^+)(1-\ell_t^-) \end{Bmatrix}; \text{ Starting at } S_{t-1}^+$$

As well, starting at S_{t-1}^-:

$$1 = \frac{1}{(1+R_f)}^2\begin{Bmatrix} p_{t,11}^{Q-}(1+h_{t-1}^-)(1+h_t^+) + p_{t,10}^{Q-}(1+h_{t-1}^-)(1-\ell_t^+) \\ p_{t,01}^{Q-}(1-\ell_{t-1}^-)(1+h_{t-1}^-) + p_{t,00}^{Q-}(1-\ell_{t-1}^-)(1-\ell_t^-) \end{Bmatrix}; \text{ Starting at } S_{t-1}^-$$

Where the correlation ρ^+ of the two implied risk neutral probabilities at time t and t + 1 are in case at time t − 1, the price is S_{t-1}^+:

$$\rho^+ = \frac{p_{t,11}^{Q+} - (1-p_{t,00}^{Q+} - p_{t,01}^{Q+})(1-p_{t,00}^{Q+} - p_{t,10}^{Q+})}{\sqrt{(1-p_{t,00}^{Q+} - p_{t,01}^{Q+})(p_{t,00}^{Q+} + p_{t,10}^{Q+})(1-p_{t,00}^{Q+} - p_{t,10}^{Q+})(p_{t,00}^{Q+} + p_{t,10}^{Q+})}}$$

And similarly when starting at S_{t-1}^- we calculate the correlation ρ^-. Note that if initially at time t − 1 we are in state +, then the probability of being in states + or − at time t, and at time t + 1, denote by $P(S_{t+1}^+ \left| S_{t-1}^+ \right.)$ and $P(S_{t+1}^- \left| S_{t-1}^+ \right.)$, etc. the probabilities of being in states + or – are:

$$P(S_t^+ \left| S_{t-1}^+ \right.) = p_{t-1,}^{Q+}; P(S_t^- \left| S_{t-1}^+ \right.) = 1 - p_{t-1}^{Q+};$$
$$P(S_t^+ \left| S_{t-1}^- \right.) = p_{t-1,}^{Q-}; P(S_t^- \left| S_{t-1}^+ \right.) = 1 - p_{t-1}^{Q-};$$

As well as:

$$P(S_{t+1}^+ \left| S_{t-1}^+ \right.) = p_{t-1}^{Q+} p_{t,}^{Q+} + (1-p_{t-1}^{Q+}) p_{t,}^{Q-};$$
$$P(S_{t+1}^- \left| S_{t-1}^+ \right.) = (1-p_{t-1}^{Q+})(1-p_t^{Q-}) + p_{t-1}^{Q+}(1-p_t^{Q+})$$

$$P(S_{t+1}^{+} \,|S_{t-1}^{-}) = p_{t-1}^{Q-}p_{t,}^{Q+} + (1 - p_{t-1}^{Q-})p_{t,}^{Q+};$$
$$P(S_{t+1}^{-} \,|S_{t-1}^{-}) = (1 - p_{t-1}^{Q-})(1 - p_{t}^{Q-}) + p_{t-1}^{Q-}(1 - p_{t}^{Q+})$$

And so on for subsequent periods of time. These equations clearly highlight the interdependence of movements in the probabilities of being in any one state. As a result, if, we were at time t − 2, in one of the two states + or −, namely: $(S_{t-2}^{+}, S_{t-2}^{-})$ with probabilities $P(S_{t-2}^{+} \,|S_{t-1}^{+}, \ldots .)$ and $P(S_{t-2}^{-} \,|S_{t-1}^{-}, \ldots .)$ then $P(S_{t-1}^{+}) = p_{t-2,}^{Q+}P(S_{t-2}^{+}) + p_{t-2}^{Q-}P(S_{t-2}^{-})$; $P(S_{t-1}^{-}) = (1 - p_{t-2}^{Q-})P(S_{t-2}^{-}) + (1 - p_{t-2}^{Q+})P(S_{t-2}^{+})$ which provides a recursive set of equations to a prior state where the process state can be observed for sure.

In terms risk neutral probabilities, future prices are random given by:

$$\tilde{S}_{t+1}^{+} = \begin{cases} S_{t-1}^{+}(1 + h_{t-1}^{+})(1 + h_{t}^{+}) & wp\ p_{t,11}^{Q+} \\ S_{t-1}^{+}(1 + h_{t-1}^{+})(1 - \ell_{t}^{+}) & wp\ p_{t,10}^{Q+} \\ S_{t-1}^{+}(1 - \ell_{t-1}^{+})(1 + h_{t-1}^{-}) & wp\ p_{t,01}^{Q+} \\ S_{t-1}^{+}(1 - \ell_{t-1}^{+})(1 - \ell_{t}^{-}) & wp\ p_{t,00}^{Q+} \end{cases}$$

and

$$\tilde{S}_{t+1}^{-} = \begin{cases} S_{t-1}^{-}(1 + h_{t-1}^{-})(1 + h_{t}^{+}) & wp\ p_{t,11}^{Q-} \\ S_{t-1}^{-}(1 + h_{t-1}^{-})(1 - \ell_{t}^{+}) & wp\ p_{t,10}^{Q-} \\ S_{t-1}^{-}(1 - \ell_{t-1}^{-})(1 + h_{t-1}^{-}) & wp\ p_{t,01}^{Q-} \\ S_{t-1}^{-}(1 - \ell_{t-1}^{-})(1 - \ell_{t}^{-}) & wp\ p_{t,00}^{Q-} \end{cases}$$

Thus, depending in the past states, the information required to assess future price on the basis of a current observed state, increases immensely, rendering the process of completing a memory based pricing model much more difficult to contend with. In a binomial model therefore, the assumption of no memory is a prerequisite to reduce the information needed to price a future return (of say a future stock price) to that of the currently observed single option price (and vice versa). Thus the price of an option, say a call option with strike K, yields:

$$C_{t+1}^{+}(S_{t+1}^{+}K \,|S_{t-1}^{+}) = \begin{cases} Max\left\{S_{t-1}^{+}(1 + h_{t-1}^{+})(1 + h_{t}^{+}) - K, 0\right\} & wp\ p_{t,11}^{Q+} \\ Max\left\{S_{t-1}^{+}(1 + h_{t-1}^{+})(1 - \ell_{t}^{+}) - K, 0\right\} & wp\ p_{t,10}^{Q+} \\ Max\left\{S_{t-1}^{+}(1 - \ell_{t-1}^{+})(1 + h_{t-1}^{-}) - K, 0\right\} & wp\ p_{t,01}^{Q+} \\ Max\left\{S_{t-1}^{+}(1 - \ell_{t-1}^{+})(1 - \ell_{t}^{-}) - K, 0\right\} & wp\ p_{t,00}^{Q+} \end{cases}$$

And

$$C_{t+1}^{-}(\tilde{S}_{t+1}, K \,|S_{t-1}^{-}) = \begin{cases} Max\left\{S_{t-1}^{-}(1 + h_{t-1}^{-})(1 + h_{t}^{+}) - K, 0\right\} & wp\ p_{t,11}^{Q-} \\ Max\left\{S_{t-1}^{-}(1 + h_{t-1}^{-})(1 - \ell_{t}^{+}) - K, 0\right\} & wp\ p_{t,10}^{Q-} \\ Max\left\{S_{t-1}^{-}(1 - \ell_{t-1}^{-})(1 + h_{t-1}^{-}) - K, 0\right\} & wp\ p_{t,01}^{Q-} \\ Max\left\{S_{t-1}^{-}(1 - \ell_{t-1}^{-})(1 - \ell_{t}^{-}) - K, 0\right\} & wp\ p_{t,00}^{Q-} \end{cases}$$

For three forward periods, the future price (and thus also the option price) are reduced to 8 potential future states rather than four, Further when the memory increases, the number of equations increases as well in both their numbers and their complexity (dependence). In this case, for a stock price three periods hence, with a short memory of one period the following state prices are defined:

$$\tilde{S}^{+}_{t+2} = \begin{cases} S^{+}_{t-1}(1+h^{+}_{t-1})(1+h^{+}_{t})(1+h^{+}_{t+1}) \\ S^{+}_{t-1}(1+h^{+}_{t-1})(1+h^{+}_{t})(1-\ell^{+}_{t+1}) \\ S^{+}_{t-1}(1+h^{+}_{t-1})(1-\ell^{+}_{t})(1+h^{-}_{t+1}) \\ S^{+}_{t-1}(1+h^{+}_{t-1})(1-\ell^{+}_{t})(1-\ell^{-}_{t+1}) \\ S^{+}_{t-1}(1-\ell^{+}_{t-1})(1+h^{-}_{t})(1+h^{+}_{t}) \\ S^{+}_{t-1}(1-\ell^{+}_{t-1})(1+h^{-}_{t})(1-\ell^{+}_{t}) \\ S^{+}_{t-1}(1-\ell^{+}_{t-1})(1-\ell^{-}_{t})(1+h^{-}_{t}) \\ S^{+}_{t-1}(1-\ell^{+}_{t-1})(1-\ell^{-}_{t})(1-\ell^{-}_{t}) \end{cases}$$

Of course, these equations have numerous parameters and therefore cannot be resolved uniquely unless an extensive data set on option prices can be assembled. The implications of the foregoing discussion is then that market "completeness" is primarily an informational problem rather than a theoretical or physical issue. Complete markets is this sense is a model that provides an appreciable means to information reduction to parameters that are in fact observables or can be inferred from observable information.

Further, if we were to assume that whether price increase or decrease (defined in theory by the ratio of future state prices) or the actual future state prices, is also equivalent as one can be defined in terms of the other. For example, if we observe the Shanghai memory based probabilities of the Shanghai composite index to increase or decrease over time, as indicated earlier and reproduced below, we can then define an appropriate set of states to define such a market to be complete, or:

Shanghai composite index: 1 day (see above)

	t + 1 Price increase	t + 1 Price decrease
t Price increase	$p_{11} = 0.5866$	$p_{10} = 0.41338$
t Price decrease	$p_{01} = 0.5341$	$p_{00} = 0.4658$

And as a result, calculate the implied rates of returns to complete the system of equations. A solution of these equations then provides a "solution" that meets the theoretical properties of market completeness with a one period memory.

Consider again

$$1 = \frac{1}{(1+R_f)^2} \begin{cases} (1+h^{+}_{t-1})(1+h^{+}_{t}) & p^{Q+}_{t,11} \\ (1+h^{+}_{t-1})(1-\ell^{+}_{t}) & p^{Q+}_{t,10} \\ (1-\ell^{+}_{t-1})(1+h^{-}_{t-1}) & p^{Q+}_{t,01} \\ (1-\ell^{+}_{t-1})(1-\ell^{-}_{t}) & p^{Q+}_{t,00} \end{cases} \text{at } S^{+}_{t-1},$$

$$1 = \frac{1}{(1+R_f)^2}\begin{cases}(1+h_{t-1}^{-})(1+h_t^{+}) & p_{t,11}^{Q-}\\(1+h_{t-1}^{-})(1-\ell_t^{+}) & p_{t,10}^{Q-}\\(1-\ell_{t-1}^{-})(1+h_{t-1}^{-}) & p_{t,01}^{Q-}\\(1-\ell_{t-1}^{-})(1-\ell_t^{-}) & p_{t,00}^{Q-}\end{cases} \text{ at } S_{t-1}^{-}$$

And let the lognormal process:

$$d\ln S(t) = \left(\alpha - \frac{1}{2}\sigma^2\right)dt + \sigma dW(t), S(0) > 0$$

And the binomial approximation

$$\Delta \ln S(t) = \begin{cases}\left(\left(\alpha - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}\right) & wp\frac{1}{2}\\\left(\left(\alpha - \frac{1}{2}\sigma^2\right)\Delta t - \sigma\sqrt{\Delta t}\right) & wp\frac{1}{2}\end{cases}$$

With $E(\Delta \ln S(t)) = \left(\alpha - \frac{1}{2}\sigma^2\right)\Delta t$ and $\sqrt{\text{var}(\Delta \ln S(t))} = \sigma\sqrt{\Delta t}$. Thus, for two binomial processes:

$$\Delta \ln S^{+}(t) = \begin{cases}\left(\left(\alpha^{+} - \frac{1}{2}\sigma^{+2}\right)\Delta t + \sigma^{+}\sqrt{\Delta t}\right) & wp\frac{1}{2}\\\left(\left(\alpha^{+} - \frac{1}{2}\sigma^{+2}\right)\Delta t - \sigma^{+}\sqrt{\Delta t}\right) & wp\frac{1}{2}\end{cases},$$

$$\Delta \ln S^{-}(t) = \begin{cases}\left(\left(\alpha^{-} - \frac{1}{2}\sigma^{-2}\right)\Delta t + \sigma^{-}\sqrt{\Delta t}\right) & wp\frac{1}{2}\\\left(\left(\alpha^{-} - \frac{1}{2}\sigma^{-2}\right)\Delta t - \sigma^{-}\sqrt{\Delta t}\right) & wp\frac{1}{2}\end{cases}$$

For example, say that $\alpha^{+} = 0.05$ and $\sigma^{+} = 2\alpha^{+}$, and setting $\Delta t = 1$, then

$$h^{+} = \left(\alpha^{+} - \frac{1}{2}\sigma^{+2}\right) + \sigma^{+} = 0.145 \text{ and } -\ell^{+} = -\left(\left(\alpha^{+} - \frac{1}{2}\sigma^{+2}\right) - \sigma^{+}\right) = 0.145$$

While for $\alpha^{-} = 0.04$ and $\sigma^{-} = \alpha^{-}$ we have: $h^{-} = 0.072$ and $-\ell^{-} = -(0.032 - 0.04) = 0.008$.

If we set the conditional probabilities: μ^{+-} as the probability of being in state "$+$" and switching to state "$-$" and μ^{-+}, the probability of switching to state $+$ from a "$-$" states, then, the following information dependent mixture process results for the logormal model:

$$1 = \frac{1}{(1+R_f)^2}\begin{cases}(1+h_{t-1}^{+})(1+h_t^{+}) & \mu_{t+1}^{+}p_{t,11}^{Q+} = \mu_{t-1}^{+}\mu_t^{+}p_{t,1}^{Q+}\,p_{t+1,1}^{Q+}\\(1+h_{t-1}^{+})(1-\ell_t^{+}) & \mu_{t+1}^{-}p_{t,10}^{Q+} = \mu_{t-1}^{+}\mu_t^{-}p_{t,1}^{Q+}(1-p_{t+1,1}^{Q+})\\(1-\ell_{t-1}^{+})(1+h_{t-1}^{-}) & \mu_{t+1}^{+}p_{t,01}^{Q+} = \mu_{t-1}^{+}\mu_t^{-}(1-p_{t,1}^{Q+})(1-p_{t+1,1}^{Q-})\\(1-\ell_{t-1}^{+})(1-\ell_t^{-}) & \mu_{t+1}^{-}p_{t,00}^{Q-} = \mu_{t-1}^{+}\mu_t^{-}(1-p_{t,1}^{Q+})p_{t+1,1}^{Q-}\\(1+h_{t-1}^{-})(1+h_t^{+}) & \mu_{t+1}^{+}p_{t,11}^{Q-} = \mu_{t-1}^{-}\mu_t^{+}(1-p_{t,1}^{Q-})p_{t+1,1}^{Q+}\\(1+h_{t-1}^{-})(1-\ell_t^{+}) & \mu_{t+1}^{-}p_{t,10}^{Q-} = \mu_{t-1}^{-}\mu_t^{-}(1-p_{t,1}^{Q-})p_{t+1,1}^{Q-}\\(1-\ell_{t-1}^{-})(1+h_{t-1}^{-}) & \mu_{t+1}^{+}p_{t,01}^{Q-} = \mu_{t-1}^{-}\mu_t^{-}p_{t,1}^{Q-}(1-p_{t+1,1}^{Q-})\\(1-\ell_{t-1}^{-})(1-\ell_t^{-}) & \mu_{t+1}^{-}p_{t,00}^{Q-} = \mu_{t-1}^{-}\mu_t^{-}p_{t,1}^{Q-}\,p_{t+1,1}^{Q-}\end{cases}$$

To calculate these probabilities and reduce this pricing model to a treatable model we proceed as follows. After one period, we have:

$$\begin{bmatrix} \mu_t{}^+ \\ \mu_t{}^- \end{bmatrix} = \begin{bmatrix} p_{t,1}^{Q+} & 1 - p_{t,1}^{Q-} \\ 1 - p_{t,1}^{Q+} & p_{t,1}^{Q-} \end{bmatrix} \begin{bmatrix} \mu_{t-1}{}^+ \\ \mu_{t-1}{}^- \end{bmatrix}$$

and

$$\begin{bmatrix} \mu_{t+1}{}^+ \\ \mu_{t+1}{}^- \end{bmatrix} = \begin{bmatrix} p_{t+1,1}^{Q+} & 1 - p_{t+1,1}^{Q-} \\ 1 - p_{t+1,1}^{Q+} & p_{t+1,1}^{Q-} \end{bmatrix} \begin{bmatrix} \mu_t{}^+ \\ \mu_t{}^- \end{bmatrix}$$

And therefore,

$$\begin{bmatrix} \mu_t{}^+ \\ \mu_t{}^- \end{bmatrix} = \begin{bmatrix} p_{t,1}^{Q+} \mu_{t-1}{}^+ + \left(1 - p_{t,1}^{Q-}\right) \mu_{t-1}{}^- \\ \left(1 - p_{t,1}^{Q+}\right) \mu_{t-1}{}^+ + p_{t,1}^{Q-} \mu_{t-1}{}^- \end{bmatrix}$$

As well as

$$\begin{bmatrix} \mu_{t+1}{}^+ \\ \mu_{t+1}{}^- \end{bmatrix} = \begin{bmatrix} p_{t+1,1}^{Q+} & 1 - p_{t+1,1}^{Q-} \\ 1 - p_{t+1,1}^{Q+} & p_{t+1,1}^{Q-} \end{bmatrix} \begin{bmatrix} p_{t,1}^{Q+} \mu_{t-1}{}^+ + \left(1 - p_{t,1}^{Q-}\right) \mu_{t-1}{}^- \\ \left(1 - p_{t,1}^{Q+}\right) \mu_{t-1}{}^+ + p_{t,1}^{Q-} \mu_{t-1}{}^- \end{bmatrix},$$

and

$$\begin{bmatrix} \mu_{t+1}{}^+ \\ \mu_{t+1}{}^- \end{bmatrix} = \begin{bmatrix} \left(p_{t+1,1}^{Q+} p_{t,1}^{Q+} + \left(1 - p_{t+1,1}^{Q-}\right)\left(1 - p_{t,1}^{Q+}\right)\right) \mu_{t-1}{}^+ \\ + \left(p_{t+1,1}^{Q+}\left(1 - p_{t,1}^{Q-}\right) + \left(1 - p_{t+1,1}^{Q-}\right) p_{t,1}^{Q-}\right) \mu_{t-1}{}^- \\ \left(\left(1 - p_{t+1,1}^{Q+}\right) p_{t,1}^{Q+} + p_{t+1,1}^{Q-}\left(1 - p_{t,1}^{Q+}\right)\right) \mu_{t-1}{}^+ \\ + \left(\left(1 - p_{t+1,1}^{Q+}\right)\left(1 - p_{t,1}^{Q-}\right) + p_{t+1,1}^{Q-} p_{t,1}^{Q-}\right) \mu_{t-1}{}^- \end{bmatrix}$$

Now since:

$$\begin{aligned} p_{t+1,1}^{Q+} p_{t,1}^{Q+} &= p_{t,11}^{Q+}; \quad \left(1 - p_{t+1,1}^{Q-}\right)\left(1 - p_{t,1}^{Q+}\right) = p_{t,00}^{Q+}; p_{t+1,1}^{Q+}\left(1 - p_{t,1}^{Q-}\right) \\ &= p_{t,01}^{Q-}; \left(1 - p_{t+1,1}^{Q-}\right) p_{t,1}^{Q-} = p_{t,10}^{Q-}\left(1 - p_{t+1,1}^{Q+}\right) p_{t,1}^{Q+} \\ &= p_{t,10}^{Q+}; \quad p_{t+1,1}^{Q-}\left(1 - p_{t,1}^{Q+}\right) = p_{t,01}^{Q+}; \quad \left(1 - p_{t+1,1}^{Q+}\right)\left(1 - p_{t,1}^{Q-}\right) \\ &= p_{t,00}^{Q-}; \quad p_{t+1,1}^{Q-} p_{t,1}^{Q-} = p_{t,10}^{Q-} \end{aligned}$$

We have:

$$\begin{bmatrix} \mu_{t+1}{}^+ \\ \mu_{t+1}{}^- \end{bmatrix} = \begin{bmatrix} \left(p_{t,11}^{Q+} + p_{t,00}^{Q+}\right) & \left(p_{t,01}^{Q-} + p_{t,10}^{Q-}\right) \\ \left(p_{t,10}^{Q+} + p_{t,01}^{Q+}\right) & \left(p_{t,00}^{Q-} + p_{t,10}^{Q-}\right) \end{bmatrix} \begin{bmatrix} \mu_{t-1}{}^+ \\ \mu_{t-1}{}^- \end{bmatrix}$$

For example, if at time t − 1, $\mu_{t-1}{}^+ = 1{,}0$, then $\mu_{t-1}{}^- = 0{,}1$ and

$$\begin{bmatrix} \mu_{t+1}{}^+ \\ \mu_{t+1}{}^- \end{bmatrix} = \begin{bmatrix} p_{t,11}^{Q+} + p_{t,00}^{Q+} \\ p_{t,10}^{Q+} + p_{t,01}^{Q+} \end{bmatrix}, \ \mu_{t-1}{}^+ = 1 \text{ and } \begin{bmatrix} \mu_{t+1}{}^+ \\ \mu_{t+1}{}^- \end{bmatrix} = \begin{bmatrix} p_{t,01}^{Q-} + p_{t,10}^{Q-} \\ p_{t,00}^{Q-} + p_{t,10}^{Q-} \end{bmatrix}, \ \mu_{t-1}{}^- = 1$$

Generally, a risk neutral framework provides the price:

$$1 = \frac{1}{(1+R_f)^2}\begin{pmatrix}(1+h_{t-1}^{+})(1+h_t^{+})\mu_{t+1}^{+}p_{t,11}^{Q+} + (1+h_{t-1}^{+})(1-\ell_t^{+})\mu_{t+1}^{-}p_{t,10}^{Q+} + \\ (1-\ell_{t-1}^{+})(1+h_{t-1}^{-})\mu_{t+1}^{+}p_{t,01}^{Q+} + (1-\ell_{t-1}^{+})(1-\ell_t^{-})\mu_{t+1}^{-}p_{t,00}^{Q+} + \\ (1+h_{t-1}^{-})(1+h_t^{+})\mu_{t+1}^{+}p_{t,11}^{Q-} + (1+h_{t-1}^{-})(1-\ell_t^{+})\mu_{t+1}^{-}p_{t,10}^{Q-} + \\ (1-\ell_{t-1}^{-})(1+h_{t-1}^{-})\mu_{t+1}^{+}p_{t,01}^{Q-} + (1-\ell_{t-1}^{-})(1-\ell_t^{-})\mu_{t+1}^{-}p_{t,00}^{Q-}\end{pmatrix}$$

Inserting ${\mu_{t+1}}^{+}$ and ${\mu_{t+1}}^{-}$ we obtain a binomial pricing model which is a function of the implied probabilities $p_{t,11}^{Q+}, p_{t,10}^{Q+}, p_{t,01}^{Q+}, p_{t,00}^{Q+}$ and $p_{t,11}^{Q-}, p_{t,10}^{Q-}, p_{t,01}^{Q-}, p_{t,00}^{Q-}$ where $p_{t,11}^{Q+} + p_{t,10}^{Q+} + p_{t,01}^{Q+} + p_{t,00}^{Q+} = 1$ as well as $p_{t,11}^{Q-} + p_{t,10}^{Q-} + p_{t,01}^{Q-} + p_{t,00}^{Q-} = 1$. However if the price process starts at states " + " or a state "−", we have:

$$1 = \frac{1}{\left(1+R_f\right)^2}\begin{pmatrix}\left(1+h_{t-1}^{+}\right)\left(1+h_t^{+}\right)\mu_{t+1}^{+}p_{t,11}^{Q+} + \left(1+h_{t-1}^{+}\right) \\ \left(1-\ell_t^{+}\right)\mu_{t+1}^{-}p_{t,10}^{Q+} + \\ \left(1-\ell_{t-1}^{+}\right)\left(1+h_{t-1}^{-}\right)\mu_{t+1}^{+}p_{t,01}^{Q+} + \left(1-\ell_{t-1}^{+}\right) \\ \left(1-\ell_t^{-}\right)\mu_{t+1}^{-}p_{t,00}^{Q+}\end{pmatrix} \text{ at a state " + " or}$$

$$1 = \frac{1}{\left(1+R_f\right)^2}\begin{pmatrix}\left(1+h_{t-1}^{+}\right)\left(1+h_t^{+}\right)\left(p_{t,11}^{Q+} + p_{t,00}^{Q+}\right) \\ p_{t,11}^{Q+} + \left(1+h_{t-1}^{+}\right)\left(1-\ell_t^{+}\right)\left(p_{t,10}^{Q+} + p_{t,01}^{Q+}\right)p_{t,10}^{Q+} + \\ \left(1-\ell_{t-1}^{+}\right)\left(1+h_{t-1}^{-}\right)\left(\left(p_{t,11}^{Q+} + p_{t,00}^{Q+}\right)\right) \\ p_{t,01}^{Q+} + \left(1-\ell_{t-1}^{+}\right)\left(1-\ell_t^{-}\right)\left(p_{t,10}^{Q+} + p_{t,01}^{Q+}\right)p_{t,00}^{Q+}\end{pmatrix} \text{ at a state " + "}$$

$$1 = \frac{1}{\left(1+R_f\right)^2}\begin{pmatrix}\left(1+h_{t-1}^{-}\right)\left(1+h_t^{+}\right)\mu_{t+1}^{+}p_{t,11}^{Q-} + \left(1+h_{t-1}^{-}\right) \\ \left(1-\ell_t^{+}\right)\mu_{t+1}^{-}p_{t,10}^{Q-} + \\ \left(1-\ell_{t-1}^{-}\right)\left(1+h_{t-1}^{-}\right)\mu_{t+1}^{+}p_{t,01}^{Q-} + \left(1-\ell_{t-1}^{-}\right) \\ \left(1-\ell_t^{-}\right)\mu_{t+1}^{-}p_{t,00}^{Q-}\end{pmatrix} \text{ at a state "−" or}$$

$$1 = \frac{1}{(1+R_f)^2}\begin{pmatrix}(1+h_{t-1}^{-})(1+h_t^{+})(p_{t,01}^{Q-} + p_{t,10}^{Q-})p_{t,11}^{Q-} + (1+h_{t-1}^{-})(1-\ell_t^{+}) \\ (p_{t,00}^{Q-} + p_{t,10}^{Q-})p_{t,10}^{Q-} + \\ (1-\ell_{t-1}^{-})(1+h_{t-1}^{-})(p_{t,01}^{Q-} + p_{t,10}^{Q-})p_{t,01}^{Q-} + (1-\ell_{t-1}^{-})(1-\ell_t^{-}) \\ (p_{t,00}^{Q-} + p_{t,10}^{Q-})p_{t,00}^{Q-}\end{pmatrix}$$

Where the autocorrelation in the " + " case, ρ^{+} is:

$$\rho^{+} = \frac{p_{t,11}^{Q+} - \left(1 - p_{t,00}^{Q+} - p_{t,01}^{Q+}\right)\left(1 - p_{t,00}^{Q+} - p_{t,10}^{Q+}\right)}{\sqrt{\left(1 - p_{t,00}^{Q+} - p_{t,01}^{Q+}\right)\left(p_{t,00}^{Q+} + p_{t,10}^{Q+}\right)\left(1 - p_{t,00}^{Q+} - p_{t,10}^{Q+}\right)\left(p_{t,00}^{Q+} + p_{t,10}^{Q+}\right)}}$$

8 Conclusion

A person who cannot change his mind in light of overwhelming evidence is similar to a financial process that does not alter future expectations in light of observed market prices. When evidence of any sort is revealed, a price process may or may not switch to another pricing model. Short memory in this case is defined by observed past prices and their effects on price models switching. Of course, such an approach differs from Markov switching models which are defined by a process switching from one model to another according to a Markov chain. Although we have considered a one period information, extensions to multiple periods information as well as summarizing past data statistically may also be considered and will be considered in subsequent research papers. These problems are far more complex to analyze although simulation of such processes and the application of numerical techniques may provide some insights on the effects of information on price models switching.

Although the first sections of the paper have dealt with cases that are mostly known and based on Arrow- and Debreu model for rational expectations in pricing, the short memory case provides some new results, pursuing developments of Vallois and Tapiero (2007), Tapiero and Vallois 1996. Other references include Viswanathan et al.1999, Weiss 1994, 2002, Weiss and Rubin 1983, as well as Toth 1986, Claes and Broeck 1987, Ferguson and Bazant 2005, Hermann and Vallois 2010, Masoliver et al. 2003, Patlak 1953, Pottier 1996.

References

Arnold, B. C., and Hon Keung Tony Ng. 2011. Flexible bivariate beta distributions. *Journal of Multivariate Analysis* 102:1194–1202.

Arrow, K. J. 1951a. *Social choice and individual values*. New York: Wiley.

Arrow, K. J. 1951b. Alternative approaches to the theory of choice in risk-taking situations. *Econometrica* 19:404–437.

Arrow, K. J. 1963. *Aspects of the theory of risk bearing*. YRJO Jahnsson lectures. Chicago: Markham Publishing Company (also in 1971 *Essays in the Theory of Risk Bearing*).

Arrow, K. J., and Gerard Debreu. 1954. Existence of an equilibrium for a competitive economy. *Econometrica* 22:265–290.

Claes, I., and C. Van den Broeck. 1987. Random walk with persistence. *Journal of Statistical Physics* 49 (1–2): 383–392.

Campbell, Damien. www.cetcapital.com.

Debreu, Gerard. 1959. *The theory of value: An axiomatic analysis of economic equilibrium*. New York: Wiley.

Debreu, Gerard. 1960. Economics under uncertainty. *Économie Appliquée* 13:111–116.

Ferguson, A., and M. Z. Bazant. 2005. Lecture 19, Polymer models: Persistence and self-avoidance. MIT Lecture Notes, 14 April 2005.

Hermann, S., and P. Vallois. 2010. From persistent random walk to the telegraph noise. *Stochastics and Dynamics* 10 (2): 161–196.

Hurst, H. E. 1951. Long terms Engineers 116:77.

Lucas, R. 1978. Asset prices in an exchange economy. *Econometrica* 46:1429–1446.

Masoliver, J., M. Montero, and G. H. Weiss. 2003. Continuous time random walk model for financial distributions. *Physics Review E* 67:1–9.

Nadarajah, A., and S. Kotz. 2005. Some bivariate beta distributions. *Statistics* 39 (5): 457–466.
Olkin, I., and R. Liu. 2003. A bivariate beta distribution. *Statistics and Probability Letters* 62:407–412.
Patlak, Clifford. 1953. Random walk with persistence and external bias. *Bulletin of Mathematical Biology* 15 (3): 311–338
Pottier, Noelle. 1996. Analytic study of the effect of persistence on a one-dimensional biased random walk. *Physica A* 230:563–576
Tapiero, Charles S. 2013. *Engineering risks and finance*. Berlin: Springer.
Tapiero, C. S., and P. Vallois. 1996. Run length statistics and the Hurst exponent in random and birth-death random walks. *Chaos, Solitons and Fractals* 7 (9): 1333–1341.
Toth, Balinth. 1986. Persistent random walks in random environments. *Probability Theory and Related Fields* 71:615–625.
Vallois, P., and C. S. Tapiero. 2007. Memory-based persistence in a counting random walk process. *Physica A* 386:303–317.
Viswanathan, G. M., S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. Reposo, and H. E. Stanley. 1999. Optimizing the success of random searches. *Nature* 401:911–914.
Weiss, G. H. 1994. *Aspects and applications of the random walk*. Amsterdam: North-Holland.
Weiss, G. H. 2002. Some applications of persistent random walks and the telegrapher's equation. *Physica A* 311:381–410.
Weiss, G. H., and R. J. Rubin. 1983. Random walks: theory and selected applications. *Adv Chemical Physics* 52:363–505.

Index

A. Bensoussan et al. (eds.), *Future Perspectives in Risk Models and Finance,*
International Series in Operations Research & Management Science 211,
DOI 10.1007/978-3-319-07524-2, © Springer International Publishing Switzerland 2015