Sequentially Forecasting Economic Indices Using Mixture Linear Combinations of EP Distributions

Gianna Agrò[1], Frank Lad[2] and Giuseppe Sanfilippo[1]
[1] *University of Palermo* and [2] *University of Canterbury*

Abstract: This article displays an application of the statistical method motivated by Bruno de Finetti's operational subjective theory of probability. We use exchangeable forecasting distributions based on mixtures of linear combinations of exponential power (EP) distributions to forecast the sequence of daily rates of return from the Dow-Jones index of stock prices over a 20 year period. The operational subjective statistical method for comparing distributions is quite different from that commonly used in data analysis, because it rejects the basic tenets underlying the practice of hypothesis testing. In its place, proper scoring rules for forecast distributions are used to assess the values of various forecasting strategies. Using a logarithmic scoring rule, we find that a mixture linear combination of EP distributions scores markedly better than does a simple mixture over the EP family, which scores much better than does a simple Normal mixture. Surprisingly, a mixture over a linear combination of three Normal distributions also makes a substantial improvement over a simple Normal mixture, although it does not quite match the performance of even the simple EP mixture. All substantive forecasting improvements become most marked after extreme tail phenomena were actually observed in the sequence, in particular after the abrupt drop in market prices in October, 1987. However, the improvements continue to be apparent over the long haul of 1985-2006 which has seen a number of extreme price changes. This result is supported by an analysis of the Negentropies embedded in the forecasting distributions, and a proper scoring analysis of these Negentropies as well.

Key words: Dow-Jones index, exponential power distributions, fat tails, logarithmic scoring rule, mixture distributions, partial exchangeability, proper scoring rules, subjective probability, subjectivist statistical methods.

1. Introduction

It is now widely recognised that a sequence of return rates from statistical indices of stock or bond prices observed at almost any time frequency, short or long, generates a histogram that is more centrally peaked and displays fatter

tails than is well described by a Normal distribution. Results in articles by Timmerman (1995) and Mantegna and Stanley (1995) have been supported in investigations of many specific phenomena such as the work of Lim et al (2006, 1998) on exchange rates and currency options, to cite only one example. Many hundreds of related research reports circulate on the web sites "arXiv.org" and "gloriamundi.org". Both sites provide a clearing ground for much collaborative research propagated by physicists with economists. Motivation for understanding the fat-tail phenomenon can be achieved by realising that actual trades whose prices are recorded in the sequence are trades between only two parties (one or both of them perhaps being a managed group of portfolio holders). The parties who actually make an exchange exhibit opinions or utilities that must be extreme in some way relative to the other parties who did not engage in that transaction. Thus, the conditions of the central limit theorem do not really apply to the definition of trading prices, which are not sums. When conditions arise that might motivate a trade, the size of the expected price change required by the parties to break the inertia of the "hold" option can be substantial.

Concomitant with this recognition of empirical tail behaviour have been renewed investigations of the properties of the family of Exponential Power (EP) distributions. This family parameterises the exponent on $|(x - \mu)/\sigma_p|^p$ in the Normal density as a variable, p. When p is less than 2, family members display fatter tails in their densities than does the Normal. The early work of Subbotin (1923) has been extended with varying terminologies in works such as Box and Tiao (1953); Vianelli (1963); Agrò (1995, 1999); and Choy and Walker (2003).

A third important contemporary development in statistical forecasting strategies has been stimulated by the understanding of exchangeability as a judgement of symmetry, and the representation of exchangeable distributions via mixture distributions. This has stemmed from the original result of de Finetti (1937), the didactic article of Heath and Sudderth (1976), and systematic expositions as in the text of Lad (1996). More than 200 refereed journal articles on exchangeability have been published during the past thirty years, as can be found through the *Current Index to Statistics*. A common misinterpretation in statistical applications presumes that the data sequence is generated by independent emanations from some particular member of the family of distributions over which exchangeable mixtures are constructed. It is common to use a maximum likelihood or even Bayesian estimates of the parameters underlying the mixture, and to forecast the continuing data sequence using such estimates. The subjectivist understanding of the issue counters such procedures with the suggestion that there is no "true distribution" generating the data. Rather, the feature of exchangeability and its representation via a mixture distribution may be a property of a forecaster's uncertain assessment of the historical processes that are measured by the data

sequence. There need be no presumption that the mixing function over p, μ and σ_p in an EP mixture, for example, will degenerate over time onto a single point. Rather, the need for mixtures is a fact of life, and the use of mixture distributions to forecast throughout the sequence is usually appropriate.

This article unifies these three themes of research in an applied sequential scoring analysis of daily closing values of the Dow-Jones Industrial Stock Price Average. Since our primary purpose is to exhibit the method of data analysis motivated by Bruno de Finetti's operational subjective theory of probability, we begin in Section two with a brief primer on methodological implications of the subjective point of view. We itemise a few important concepts that will be used in our analysis for readers who are new to them, and we reference a source of more extensive introductory materials. Section three introduces the data source and relevant institutional information pertinent to our specific data analysis. In Section four we describe details of four different forecasting distributions for the historical data that we study here, along with the computational arrangements we have required. Section five describes the properties of "proper" scoring rules we use to assess the quality of the sequential mixture forecasting distributions. Numerical and graphical results are presented in Section six, and are discussed in the concluding Section seven. Appendices 1 and 2 display the relevant sequence of Dow-Jones daily return rates and an array of EP density functions. Appendix 3 presents four Tables summarising features of terminal mixing distributions.

2. A Primer on Operational Subjective Data Analysis

The foundational work in probability of Bruno de Finetti (1906-1986) is widely regarded as innovative and insightful. Nonetheless, the procedures of statistical practice that his work supports are not commonly followed or even known. During the century in which the objectivist understanding of probability won widespread support and inspired the common practice of hypothesis testing and parameter estimation, de Finetti's work followed a distinctly different path. He characterised probabilities not as unobservable objective properties of nature, but rather as numerical representations of individuals' uncertain knowledge about historical occurrences which are measurable by numerically defined categories. If probabilities "do not exist" (de Finetti, 1974, p. x) as unobservable entities to be estimated, the entire objectivist statistical practice of hypothesis testing regarding alternative "generating distributions" and the estimation of parameters of favoured distributions is a programme that follows an illusion. In the place of these procedures, a complete operational subjective statistical programme has by now been developed.

About any sequence of measured historical quantities (whether the result of designed experiments or mere historical observations of happenstance) the initial

step in the programme is to formulate probability structures that adequately represent considered yet uncertain opinions about the historical occurrence of these measurements. When a joint distribution can be formulated regarding a sequence of observations, each actual observation value is incorporated into a sequence of conditional forecasting distributions that specify probabilities for various possible values of the next observation conditioned on the sequence that has actually occurred. The conditioning is based on standard laws of probability which are motivated by the principle of coherency of probabilities. Coherency is a generalisation of the principle of non-contradiction in the two-valued logic of certainty, and applies to the array of operational claims that define uncertainty probabilities.

The central unifying feature of joint distributions over a sequence of quantities that specifies forms of learning by conditioning is *exchangeability* and its extensions to various forms of partial exchangeability. This is a judgement that formalises the symmetries that often characterise considered attitudes towards observations that occur in a sequence, either temporal or spatial. A complete introduction to the concept and its application must be deferred to a work such as Lad (1996, 3.8-3.12). In this article we shall not dwell on fundamental characterisations regarding symmetry over permutations. Rather we shall use directly a very important result of de Finetti's work, that infinitely extendible exchangeable distributions in whatever form of exchangeability can be represented by conditionally independent *mixture distributions*. A joint density $f(x_1, x_2)$ is a conditionally independent mixture with respect to a parameter, θ, if it is representable in the form $f(x_1, x_2) = \int f(x_1|\theta) \, f(x_2|\theta) \, f(\theta) \, d\theta$. In such a situation $f(x_2|x_1)$ is quite different from $f(x_2)$. This brief review is meant to make our allusions to exchangeability in the article clear.

Probability *distributions* over an array of possible measurement values are not *required* for the specification of professional uncertainty. In many practical problems the elicitation of expectations of the measured quantity would suffice. Such expectations are called "previsions" in the technical construction of de Finetti's mathematics. However in other problems, the assessment of a distribution is fairly essential to the operational specification of uncertainty. Unknown rates of return on investment portfolios held by banks are a specific case in point. Specifically, early in the 1990's it became required that banks hold specific reserves among their assets to offset the extreme "tail probabilities" for extreme losses that they might conceivably occur, specific to the types of risk they allow in their investment portfolio choice. For this reason it is important to be able to compare the performance of different opinion *distributions* in terms of the entire array of probabilities assessed rather than merely in terms of some specific quantity expectations.

When several possible families of mixture distributions are presented as contenders to adequately represent considered scientific opinion regarding an uncertain historical process, operational statistical procedures have been developed that are designed to compute comparable *scores* for each of the tendered families in order to assess which of them provides a more realistic assessment of the uncertainties involved. It is widely agreed among proponents of the subjectivist statistical position that the *scoring rules* to be used for this programme should be *proper*. A scoring rule for a probability distribution is a function of both the observation that comes to be observed and of the probabilities assessed that the observation will occur in various regions of its possible realm. The rule is said to be proper if anyone who assesses a personal probability distribution cannot "expect" (in the technical sense, according to the tendered distribution) to achieve a greater score by professing a probability distribution that is different from the actual distribution as assessed. That is, no gain can be expected from systematic distortion of professed opinions.

An important proper scoring rule for distributions, which is widely honored and which shall be used in the application of the present article, is the *logarithmic scoring rule*. This awards the score to an uncertainty density that equals the logarithm of the density function evaluated at the observed value of the quantity that eventuates. If you assert a density for a quantity X as $f(x)$, and X is observed to equal the value x^o, say, then the score you are awarded is $\log[f(x^o)]$. As you observe a sequence of values of successive X's, these logarithmic values will be accumulated and compared to the scores that accumulate for alternative tendered uncertainty distributions. In this context, the logarithmic score that you "expect" to achieve when you tender a density $f(x)$ is computed via the expectation $\int \log[f(x)] \, f(x) \, dx$. This integral can be recognised as the negative of the entropy in your density (entropy being defined exactly as this integral value but preceded by a negative sign). Thus, its value is commonly termed the "negentropy" measure of your distribution. Now the negentropy is not a probability distribution but an expectation, or prevision. The theory of proper scoring rules also extends to proper scores for previsions. When the time to score negentropies for competing distributions arises in this article, they shall be scored according to the quadratic scoring rule, which is also a proper rule.

It is hoped that this brief introduction will familiarise the reader with the context in which the data analysis we present in this article will proceed. We shall make brief technical notes for matters that may be unfamiliar when they arise. However in the main, we are hoping that this article will exemplify the perhaps unfamiliar format of an operational subjective statistical analysis of a real contentious applied scientific question for readers who might wonder what might be the alternative to "hypothesis testing" procedures which subjectivist

statisticians completely reject. The reader who is interested by these ideas and who finds them new is directed to the sophisticated yet introductory level text of Lad (1996) that covers these and many more matters in great detail. It dwells on foundational issues of the meaning of statistical activities while providing complete technical developments with practical applications.

3. Data: Operational Definitions and Their Implications

Suppose that a sequence of observed daily closing prices for a financial instrument or index is denoted by the variables $P_0, P_1, P_2, ..., P_t, ..., P_T$. The daily rates of return that are accrued by an owner of the instrument are then commonly described by the transformed variables $X_t = \log(P_t/P_{t-1})$. Since the increase in value of the instrument over the course of a single day is expected to be small, the interpretation of X_t as the daily rate of return derives from the mathematical fact that $\log(1+r) \approx r$ for small values of r. In this article we study the sequence of daily closing values of the Dow-Jones Industrial Index, covering the period of 25 October, 1984 through 25 October, 2006. This series is readily available on the web site "www.djindexes.com". A plot of the daily rates of return over the period of our data series is displayed in Appendix 1 to the present article.

It is worthwhile to make some comments of common sense about the theoretical and computational understanding of procedures for forecasting this data. In the arena of economic "high theory" that is largely conducted within an objectivist understanding of probability, it has become common to speak forthrightly about the "generation" of a price series in a rational market in terms of a continuous stochastic process of independent increments governed by some fat-tailed distribution. It needs to be recognised at this outset that the observed price variables cannot possibly be generated in such a way. Firstly, of course, the possible observable measurements of a specific price data series or an index series are necessarily discrete. Thus, the data series cannot really be generated by a continuous-valued process. Moreover, there is no real relevance of "fat tail" properties of a supposed generating distribution in the way these properties are mathematically defined. If one thinks about a series as commonplace as the Dow-Jones average, one should be aware that if historical events would occur that could stimulate either a doubling or a halving of this price index on a single day, for example, the directors of the New York Stock Exchange would doubtless suspend trading well before such a cataclysm could occur, merely to allow traders' nerves to settle. After a controversial decision to interrupt trading for such a reason in October, 1987, the NYSE publicly announced new standard procedures for suspending trading for limited periods in the case that the Dow-Jones index drops by 10, 20, and 30% during a day's trading. The directors of the exchange may decide to close the exchange to trading for other reasons as well. For example,

the exchange was closed during the week following 9/11.

This information is presented here only to highlight an explicit awareness that the application of mixture EP distributions to opinions about the measured rates of return is only an approximation to distributions that are actually discrete. Our use of EP mixtures instead of Normal mixtures is motivated simply by the fact that members of the EP family are more centrally peaked and display relatively larger tail probabilities than do members of the Normal family. The usual theoretical interpretations of continuous distributions are amusing further because many computations, such as used in this article for the determination of entropies, necessarily resort once again to discreteness in order to perform the required numerical integrations!

Another approximative aspect of our analysis here is that pure exchangeability is not precisely appropriate to informed attitudes about the rate of return series. Analysis such as the work of Lim et al (2006) already cited highlights the recognition that market prices do go through periods of varying volatility, though it is hard to identify precisely when they will occur. Partial exchangeability via another level of mixing distributions is required to portray this. The approximation of exchangeability is specified in this application so that we can focus on the construction of an appropriate computational forecasting procedure that can be scored by proper rules.

4. Mixture EP Distributions Used in Forecasting

Our forecasting scenario is designed to represent the learning process of someone who accepts that the expectation of a day's closing price, conditional on a sequence of closing price values on previous days, equals the closing price value on the preceding day. It is presumed that previous data observations do not provide useful information to motivate an expectation of a price increase or decline. This is to say that the price sequence is understood as a martingale. The expected rate of return over a day, conditioned on an observed sequence of returns over previous days, always equals zero for all the distributions we compare. In more detail however, accumulating data observations may allow one to learn about the variability that is to be expected in the price sequence. We utilise the awareness of empirical tail properties of return histograms by representing opinions via exchangeable mixtures over a linear combination of EP distributions. Starting with an initial value of the price sequence, called time $t = 0$, we compute sequential forecasting densities for the transformed sequence of daily returns in the form of $f(x_1), f(x_2|x_1), f(x_3|x_1, x_2), ..., f(x_{t+1}|\mathbf{x}_t), ...$, where the bold symbol $\mathbf{x}_t$ denotes the vector of observations of x_1 through x_t.

The $\text{EP}(p, \mu, \sigma_p)$ family is a three-parameter family of distributions, composed

by member densities of the form

$$f_X(x|p, \mu, \sigma_p) \;=\; [2p^{1/p}\sigma_p\Gamma(1 + 1/p)]^{-1}\exp[-(p\sigma_p{}^p)^{-1}|x - \mu|^p] \qquad,$$

defined over all real values of x. The term μ is a location parameter, the conditional expectation of X, and may be any real number. The term σ_p is a scale parameter that relates to the standard deviation of X given (p, μ, σ_p) through the equation

$$SD(X|\mu, p, \sigma_p) \;=\; \left[\; p^{2/p}\,\frac{\Gamma(3/p)}{\Gamma(1/p)}\;\right]^{1/2}\;\sigma_p \qquad. \qquad (4.1)$$

Finally, the variable value of $p \in (0, \infty)$ is a shape parameter, identifying the tail and curvature properties of a family member. Member densities with $p < 2$ have tails that are fatter than a Normal density, while those with $p > 2$ have thinner tails than does a Normal. The Normal density itself is the family member corresponding to $p = 2$. A graphical display of several members of the EP family of density functions appears in Appendix 2 of this article.

    We should remark on the technical meaning of our allusion to "fat tails". Two distribution functions, $F(\cdot)$ and $G(\cdot)$ are said to be "equivalent in their right tails" if there exists a constant $\lambda_+ \in (0, \infty)$ such that $lim_{x \to +\infty}[1 - G(x)]/[1 - F(x)] = \lambda_+$. If the distribution functions support densities, then this limiting ratio also equals the limit of the density ratio, according to L'Hospital's rule. A similar definition governs equivalence in the left tail. A distribution $F(\cdot)$ is said to have "fat tails" with respect to a Normal (Gaussian) $G(\cdot)$ if the limit of this same ratio equals 0. It should be evident that a mixture distribution combining fat tailed distributions also has fat tails. Specifically relevant to our analysis, the $EP(p, \mu, \sigma_p)$ family of distributions displays fat tails with respect to a Normal distribution whenever $p < 2$. Moreover, any convex mixture of such distributions also has fat tails relative to a Normal.

    Exchangeable distributions that are infinitely extendible as exchangeable distributions can be represented by mixtures of conditionally independent distributions. The first forecasting distribution we describe for the sequence of daily rates of return is a simple, two-parameter mixture of EP distributions, mixed over family members with $p < 2$. The mixing function is degenerate on $\mu = 0$, and is rather mild with respect to p and σ_p over a wide but sensible grid. Once the details of this structure are explained, we shall embellish the forecasting distributions to a mixture over a linear combination of three EP distributions, all degenerate on $\mu = 0$. These will be denoted by $EP(\boldsymbol{\alpha}_3, \mathbf{p}_3, \boldsymbol{\sigma}_{p3})$, meaning a mixture over a linear combination of distributions with convex coefficients $\boldsymbol{\alpha}_3 \equiv (\alpha_1, \alpha_2, \alpha_3)^T$ and a common tri-part parameter structure over (p, σ_p). Details will appear when the construction is proposed. As exemplified in the second preceding sentence,

throughout this article we use bold notation for variables that are vectors, subscripted by the number of their components.

A final clarification should be highlighted, regarding the difference between a distribution for a combination of variables and a combination of distributions for a variable. This well-known difference can be recalled most easily via some examples. The distribution for a linear combination of quantities, each distributed Normally, is of course distributed Normally. However a linear combination of Normal distributions for a variable is not a Normal distribution. A linear combination of Normal distributions can be considered as a mixture distribution. A Normal mixture distribution of Normals (via the location parameter) is Normal. Any other mixture of Normals is not. Our description in the next two subsections will clarify what we mean by an important family of distributions for the analysis in this article, the MixLC3EP family. By this we mean

$$f(x) = \int [\alpha_1 \, f_1(x|p_1, \sigma_{p_1}) + \alpha_2 \, f_2(x|p_2, \sigma_{p_2}) + \alpha_3 \, f_3(x|p_3, \sigma_{p_3})] \, dF(\boldsymbol{\alpha}_3, \mathbf{p}_3, \boldsymbol{\sigma}_{p_3})$$
,

which is aptly called a mixture of a Linear (convex) Combination of 3 Exponential Power distributions. Details follow.

4.1 Mixing functions and mixture distributions

We shall begin our forecasting analysis by describing the details of a simple mixture of $EP(p, \sigma_p)$ distributions for which the location is fixed at $\mu = 0$ for every p. This will simplify the subsequent description of the extension to mixture linear combinations of EP members. Theoretically speaking, any form of initial mixing density $f(p, \sigma_p)$ can be used to represent a mixture EP forecast distribution for the sequence of X_t observations. For computational reasons, we have found it convenient to use a mixing function that is defined by masses over a grid of discrete values of p and σ_p, and to compute the mixture probabilities via summations. The grid is chosen to be fine enough to represent reasonable opinions. The computations will eventually involve numerical integrations for both the Negentropy and the expected density value of the distributions as well. While MCMC methods might be used for the computations of any single step of this forecasting analysis, they are impractical for the computation of the entire sequential analysis of this data series, including more than 5000 observations.

We express the initial mixing function over p and σ_p using the conditional-marginal factorisation $f(p, \sigma_p) = f(p) \, f(\sigma_p|p)$. The "prior" mixing distribution on p is meant to represent an opinion that recognises the importance of the feature of fat tails for the uncertainty distribution, but is not very precise in understanding just "how fat" the tails should be. Specifically, the mixing probabilities on p place positive weights only on the discrete digits from .4 through 2.1, in steps

of .02. This grid is described using Matlab notation by $p = [0.4 : 0.02 : 2.1]$, a notation which is common in many other programming languages as well. We shall use this notation through the rest of this article when it is helpful. The weights we assessed for our initial mixing function at the beginning of the data series are best described by viewing the top half of Figure 1.
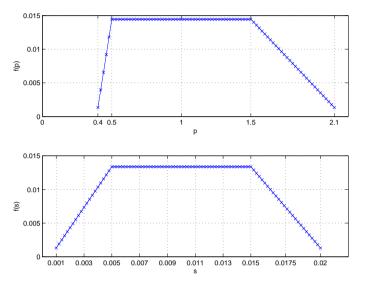

Figure 1: Factored components of the initial mixing function $f(p, s) = f(p)f(s)$ over grids of their arguments.

As to the conditional mixing functions $f(\sigma_p|p)$, they are designed to represent an opinion that is not very precise in identifying the variability to be expected in the return sequence, but is fairly sure that the standard deviation of the uncertainty about the rates of return should be less than .02, and most likely much less. Recall from equation (4.1) the proportional relation between $SD(X|p, \sigma_p)$ and the mixing parameter σ_p. For each p we allow the corresponding value of $SD(X|p, \sigma_p)$ to cover the grid range $s = [0.001 : 0.0002 : 0.02]$. For each value of p, this grid over s is transformed to the corresponding grid range for σ_p. As is apparent, this initial mixing function specifies the mixture over "s" independently from p, implying a specific dependence between p and σ_p. However, the form of this dependence will surely change as the conditional forecasting sequence progresses, and thus a dependence between p and s will emerge in the posterior mixing functions as well.

Again, the best way to understand the weights on our initial mixing function and the grid over values of $SD(X|p, \sigma_p)$ is to view now the *bottom half* of Figure 1. This mixing function expresses explicitly the idea that initial opinions are fairly precise about limits on the values of s but they are rather uninformed about how

s, and thus σ_p, ought to vary with p in any way more complex than is known through equation (4.1). As the mixing function is subsequently updated with each observation of x_t, a more informed dependence will become specified in a dependent joint posterior mixing function for p and σ_p.

Having implicitly specified the grid (p, σ_p), we can now simply express that the initial forecast density for X_1 is computed for each possibility of x_1 via the mixture density

$$f(x_1) \;=\; \sum_p \sum_{\sigma_p} \; f(x_1|p, \sigma_p) \; f(\sigma_p|p) \; f(p) \qquad . \qquad (4.2)$$

Furthermore, for any step t in the sequence, the conditional forecast density is computed via

$$f(x_t|\mathbf{x}_{t-1}) \;=\; \sum_p \sum_{\sigma_p} \; f(x_t|p, \sigma_p) \; f(p, \sigma_p|\mathbf{x}_{t-1}) \qquad (4.3)$$

for each possibility of x_t, but only for the conditioning observed values of $\mathbf{x}_{t-1}$. The sequential mixing functions $f(p, \sigma_p|\mathbf{x}_{t-1})$ appearing in (4.3) are computed iteratively according to Bayes' rule via

$$f(p, \sigma_p|\mathbf{x}_{t-1}) \;\propto\; f(x_{t-1}|p, \sigma_p) \; f(p, \sigma_p|\mathbf{x}_{t-2}) \; . \qquad (4.4)$$

Computationally, the factorisation of $f(p, \sigma_p)$ into $f(\sigma_p|p) \; f(p)$ is made only in equation (4.2) to construct the mixing mass function values at the appropriate values of (p, σ_p) in a matrix form. Subsequently, the mixing function matrix is updated according to equation (4.4) merely by multiplying each component of that matrix by the appropriate likelihood value and normalising the components of the matrix so they sum to 1.

4.2 Extension to mixture linear combinations of the EP family

The fact that some EP distributions are fat-tailed relative to Normal distributions does not *necessarily* motivate the representation of informed uncertainty by a simple mixture EP distribution. Convex combinations of such EP distributions can also be fat-tailed, as are many other distributions, both parametric and non-parametric. We have extended the array of forecasting distributions under study to mixtures of linear combinations of three EP distributions, parameterised by the vectors $\boldsymbol{\alpha}_3, \mathbf{p}_3$, and $\boldsymbol{\sigma}_{p3}$. Essentially, the LC3 $\boldsymbol{\alpha}_3$ parameters are spread across the unit-simplex $\mathbf{S}^2$: the domain of α_1 values is $[0.1 : 0.1 : 0.8]$; for each value of α_1, the domain of α_2 is $[0.1 : 0.1 : 0.9 - \alpha_1]$; and finally, for each pair of (α_1, α_2), the value of $\alpha_3 = 1 - \alpha_1 - \alpha_2$. The ranges of values for p_1, p_2, and p_3 are staggered about $p_1 = [0.4 : 0.3 : 1.9]$, with the range for $p_2 = p_1 + 0.1$, and $p_3 = p_1 + 0.2$.

This computational strategy allows us to cover the appropriate range of values for p (specified in the simple mixture EP distribution) in the linear combinations without expanding the sizes of matrices required for the computations beyond what is practical. The associated domains for the three σ_p parameters are staggered as well, to cover the same range for their associated standard deviations as for the MixEP, 0.001 through 0.023. Details of the three associated initial mixing functions for the three p's and s_i's can be viewed in Figure 2. Sequential forecasting according to the associated distribution is labeled MixLC3EP in the displays of scores for distributions appearing in Section 5. They are computed similarly to the strategy outlined in equations (4.2-4.4), but those equations are embellished in any instance to replace $f(x_t|p, \sigma_p)$ by $f(x_t|\boldsymbol{\alpha}_3, \mathbf{p}_3, \boldsymbol{\sigma}_{p3})$, and to replace the initial $f(p, s)$ by $f(\boldsymbol{\alpha}_3, \mathbf{p}_3, \mathbf{s}_3) = \prod_{i=1}^{3} f(\alpha_i)f(p_i)f(s_i)$ with appropriate modifications to the embellished mixing functions.
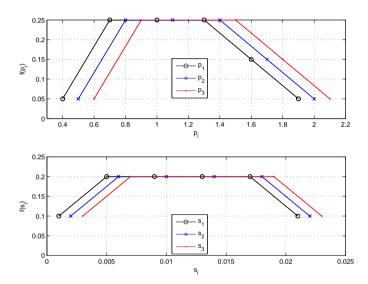

Figure 2: Marginal components $f(\mathbf{p}_3)$ and $f(\mathbf{s}_3)$ of the initial mixing function for MixLC3EP.

4.3 Mixture Nnormal forecasts, and mixture linear combinations of normals

For completeness of comparisons, we also computed a mixture Normal forecasting procedure, and we extended this to a mixture of three linear combinations of Normal distributions as well. These forecasting procedures will be labeled MixN and MixLC3N in the report of the scored forecasts. These forecasting distributions can be considered as special cases of the MixLC3EP distributions, but with mixing functions that are degenerate in the p-dimension that specifies the

Normal family as a subfamily of EP distributions. The initial mixing functions in the SD-dimension were specified identically to the mixing functions constructed for the MixEP and MixLC3EP distributions.

5. Assessment of Forecast Strategies via Proper Scoring Rules

We shall now describe the sequential computational procedure we follow for computing scores of these forecasting distributions when the unknown quantity comes to be observed in the various periods t as values denoted by $X_t = x_t^o$. Two scoring functions are relevant here: the first is a score of the forecasting distributions themselves, namely the *logarithmic scoring rule*; the second is a score of the expected score to be achieved, according to the assessment of the distribution itself. As will be explained, for the latter score we shall report on the *quadratic score* of the expected score for each distribution.

Developed within the context of the subjective theory of probability, the two scoring functions we employ are both said to be "proper" scoring rules. Someone asserting a probability distribution for a quantity with the awareness that the logarithmic function will be used to evaluate the quality of the distribution will expect to achieve a maximum score only by asserting honestly the actual distribution that represents his/her uncertainty. No improvement can be expected to be achieved via false posturing. The relevant theory of proper scoring rules for distributions has a long history. It was central to de Finetti's ideas about assessing the relative values of different subjective forecasting distributions. See de Finetti (1962). The application to scoring continuous densities was described in the article of Matheson and Winkler (1976), while the general theory and application has been reviewed in the text of Lad (1996, Chapter 6) and in the recent article by Gneiting and Raftery (2007). Specified in terms of a generic density for a quantity X on the basis of observing the value of X to equal x^o, the logarithmic scoring function is defined by $S_{\log}(X = x^o, f_X(x)) \equiv \log[f_X(x^o)]$. While there are many functions that qualify as proper scoring rules for distributions, the logarithmic scoring rule is the unique scoring rule that is a function of the observed datum *only* through the actual value of the observation. The score does not depend on the probability density assessed at values of X that might have been, but were not observed. In enjoying this property, it mimics a feature of the likelihood principle for inference. See Bernardo (1979). It is also a proper rule for which the score of a joint distribution equals the cumulative summed score for the sequential conditional distributions. For these reasons, we shall compare the quality of the four sequential forecasting distributions under consideration according to how they fare with respect to the cumulating sum of the logarithmic scores they achieve.

It should be evident that the expectation of the logscore to be achieved by a

distribution, according to the distribution itself, equals the negative entropy of the distribution: $E[S_{\log} f_X(\cdot)] = \int \log[f_X(x)] f_X(x) dx = -Entropy[f_X(\cdot)]$. Since the Negentropy in X is an assessment that varies according to the uncertainty distribution of the assessor, it is appropriate to track the distributions according to a proper score of their Negentropy assessments as well. We shall use a quadratic score of the Negentropy relative to the achieved logarithmic score do to this. The quadratic scoring function is also a proper scoring rule for expected values. It is conventional that scoring functions are scaled so that a larger score is a better score. Thus, the quadratic rule we employ to score the forecasts' assessed negentropies is defined by $S_{quad}(X = x^o, Negentropy[f_X(\cdot)]) \equiv -\{\log[f_X(x^o)] - Negentropy[f_X(\cdot)]\}^2$. As we shall see, the scoring of the Negentropies inhering in the various distributions shall provide us with an understanding of the differences in the scores achieved according to the logarithmic scoring rules of the distributions. The Negentropy in a distribution is a measure of the amount of information inherent in the distribution. Of course a forecaster might make a reasonable evaluation of the amount of information his/her uncertainty provides, or might be mistaken. The quadratic score of the Negentropy can assess this.

One proviso must be made regarding proper scoring rules: that they are meaningful only up to a linear transformation. That is, their enviable properties define them uniquely only up to a linear transformation, similar to utility functions which they can be taken to represent. See DeGroot (1984). As a result, comparisons of forecasts according to any particular scaling factor are useful only for *ordering* the quality of the forecasts. Assessing the scale of the differences between the forecasts is achieved by viewing the forecasting distributions themselves, to observe the extent of the differences in uncertainty assessment they imply. This is the tack we shall follow in presenting our computational results.

To review the entire computational strategy that we have described to this point, let us list a row of objects to be computed for each time period, $t = 1, 2, ..., T$ in the time series. The row for time t corresponding to any forecast density $f(x_t|\mathbf{x}_{t-1})$ and the observation $X_t = x_t^o$ will consist of the triple $S_{\log}[x_t^o, \ f(x_t|\mathbf{x}_{t-1})]$, $E\{S_{\log}[x_t^o, \ f(x_t|\mathbf{x}_{t-1})]\}$, and $S_{quad}\{S_{\log}[x_t^o, \ f(x_t|\mathbf{x}_{t-1})] \ E\{S_{\log}[x_t^o, \ f(x_t|\mathbf{x}_{t-1})]\}\}$. These three computations are, respectively, the log score of the forecast density, the expectation of that score which is embedded in that density (its negative entropy), and a quadratic score of this expectation against the actual log score that obtains. The computation of this vector for each observation can be used to evaluate the relative merits of the various types of forecasting distributions we have described for consideration. As we compute these functions through each period, we shall cumulate the scores of the forecast densities and the associated

expected values of the scores that are embedded within them. The results of these computations are presented graphically in the next Section.

6. Statistical Results

We begin with a reminder to pay attention to the units of all displayed graphs. These are marked on the top left corner of the ordinate scale (the "y-coordinate axis").

6.1 Cumulative logarithmic scores of the distributions

Figure 3(a) introduces the scores we report with a global comparison of the cumulating logarithmic scores of the four forecasting distributions MixLC3EP, MixEP, MixLC3N, and MixN. These names are stated in the order from the best scoring to the worst scoring among these four distributions over the period. This ordering among the first three distributions is not completely apparent, because on the scale of the cumulative scores over twenty years, the first three of them appear to achieve so similarly. Moreover, the four distributions achieve virtually indistinguishable scores during the first 750 trading days of the period we tracked. Figure 3(b) will be required to display the differences, as explained below.
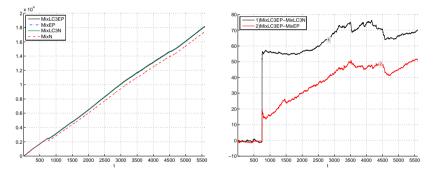

Figure 3: Gross and detailed comparisons of the logarithmic scoring rule assessments of the distributions. Left panel: Gross cumulative logarithmic scores of the forecasting distributions over twenty years. Right panel: Differences between the cumulative logarithmic scores of MixLC3EP and the scores of MixEP and MixLC3N.

Trading day 752 was the infamous day of 19 October, 1987, when the Dow-Jones index dropped 554 points, which was one-quarter of its total value at the time. On this day, the NYSE abruptly announced a "circuit-breaker" procedure and temporarily closed the market. Subsequently, the NYSE established formal procedures for implementing automatic temporary closures. The relevance of day 752 to our scoring analysis is that the MixN distribution abruptly dropped off the

rank of a creditably contending form of forecasting distribution for daily returns, relative to the other three contenders. Moreover, it is evident even on the gross scale of Figure 3(a) that the quality of its forecasting performance deteriorated fairly steadily relative to the others throughout the remaining forecast period.

Details of the differences in the scores for the forecasting performance of the three best distributions appear in Figure 3(b). This displays the *differences* between the cumulating score of MixLC3EP and the cumulating scores of the other two distributions. On day 752, the MixLC3EP distribution begins a fairly steady, though sometimes erratic, improvement in its cumulating log score relative to MixEP and MixLC3N. Implied by the fact that the difference MixLCEP-MixLC3N is the greater of these two difference functions is that MixEP scores better than MixLC3N. The scale of these differences is small relative to the scale of the cumulative scores used in Figure 3(a). What is surprising relative to current widespread interest in "fat-tailed" distributions is that the MixLC3N distribution, though admittedly inferior to the other two as cumulated over the entire period, does perform rather well. It even scores better during a few time intervals, as indicated by the periods when the difference value of cumulative MixLC3EP-MixLC3N scores declines in Figure 3(b). Nonetheless, the MixLC3N is *not* considered to be a "fat tailed" distribution in the technical sense we have discussed. Even more provocative may be a visual inspection of Figures 4(b), which shows that to the crude eye, the tail areas of the MixLC3N distribution appear fatter over the sub-domain [.035, .06] than either of the mixture EP distributions shown there. In practical discussion, we need be precise in our technical use of "fat-tail" to describe distributions, which is a property of densities that only bites at the "infinite" end of the density domain!

6.2 Distinctive features of the predictive densities

A striking feature evident from Figure 3 is that while the scores of all four distributions are virtually indistinguishable through trading day 751, the score of MixN drops away quite quickly while the other three distributions appear to be sensitive enough to be not so drastically affected in their accumulating scores.

The pairs of Figures 4 and 5 exhibit enough differences among the four sequential forecasting densities to summarise how this has occurred. Each of these Figures displays all four forecasting densities we are considering, but at different points in time. Figure 4 shows the forecasting distributions appropriate to trading days 3 and 5584, the beginning and end days of our sequence, while Figure 5 shows the distributions appropriate to days surrounding one of the major price adjustments of the century. The ordinate scales displayed run only over the interval from −.04 through +.04, though the density domains are actually unbounded. Comparing the left-hand panels of Figures 4 and 5, you can see what was learned

about return variation over the two years prior to the big drop on 19 October, 1987; then comparing the two sides of Figure 5 you see what immediate influence the observation of the four days' tumult would have had on assessors' attitudes about the distribution of daily returns. Figure 4(b) displays how the forecasters' opinions would have settled by the end of the 20 years of trading experience. We have done our best to make the graphical distinctions evident in a black and white printing, but they will always be clearer in colour on the electronic version of this *Journal*.
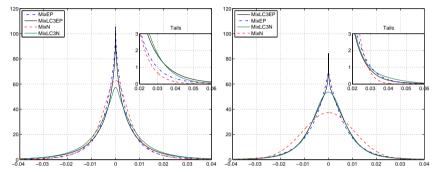

Figure 4: The four predictive densities at the beginning and the end of the sequential forecasting analysis. Left panel: Four predictive densities for the rate of return on the third day of trading in the series. Right panel: Four predictive densities for the rate of return after learning from more than 20 years' trading experience.
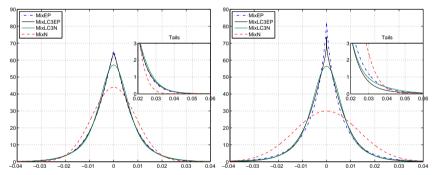

Figure 5: The four predictive densities shortly before and shortly after the striking drop and partial recovery of stock prices in October, 1987. Left panel: Four predictive densities for the rate of return on trade day 750, which was 19 October, 1987. Right panel: Four predictive densities for the rate of return only four trading days later.

Viewing the entire range of predictive distributions, it is noticeable that the mixture EP distributions both assert higher forecast probabilities for rates of return in the vicinity of zero, and for rates of return more extreme than 3.5 percent, than does the mixture Normal distribution. It is evident to the eye that

the mixture of a linear combination of three Normals does have some capacity to represent a similar form of distribution too, much more so than a simple mixture Normal. It is also evident that the MixLC3EP makes use of its flexibility to include more sharpness around zero and in the tails than does a simple mixture EP. The relevance of these seemingly minor differences to practical matters revolves around issues of "value at risk", or VAR characteristics of portfolios held by banks. This concerns the risk exposure of banks to extreme changes in stock prices, which is regulated in the USA by the Federal Deposit Insurance Corporation. See, for example, the report of Lopez (1998). A specialised study to score the relevant tail areas of the distributions we are tracking here is being conducted by the authors.

6.3 Signals about the relative importance of the LC3 component of the mixture distribution ... from changes in the mixing function

In our introduction to this article we had raised the question as to whether it is useful to think in terms of a single "fat-tailed" distribution from the EP family when assessing a sequence of daily price returns, or whether the perspective that a mixture of such family members would be more appropriate. Attention to the development of the mixing functions for the sequential forecasting distributions provides a useful way to get a handle on this question.

Under the initial grid specification we described in Section 3.3, the MixLC3EP distribution involved a mixture over more than 2 million parametric LC3EP distributions. Remember that the parameter vector specifying each such distribution is $(\boldsymbol{\alpha}_3, \mathbf{p}_3, \boldsymbol{\sigma}_{p3})$. The initial grid over this 8-dimensional parameter space (recall that the convex coefficients $\boldsymbol{\alpha}_3$ sum to 1) covers $45 \times 216 \times 216$ points. However, by the time 500 days of trading were observed, the mixing function was updated to the extent that only 2550 of these parameter vectors were accorded weights exceeding .0001, with only three exceeding .001 while not exceeding .01. The mixing function does coalesce further to some extent, with time. By the final observation in our sequential forecasting study the mixing function for MIXLC3EP assesses 56 parameter vectors with weights exceeding .001, of which fourteen exceed .01, while six exceed .03, and only two exceed .1. The largest weight is .36, associated with the vector $(\boldsymbol{\alpha}_3, \mathbf{p}_3, \boldsymbol{\sigma}_{p3}) = (0.1, 0.3, 0.6, 0.4, 1.4, 1.8, 0.013, 0.014, 0.017)$. Summary Tables reporting the sequential development of the mixing functions for MixLC3N and MixLC3EP appear in Appendix 3.

The conclusion from this review of the mixing function's development is that the spread of the mixture weights over the EP parameter space remains quite diffuse, even after more than five thousand observations in the data sequence. By contrast, the mixture LC3N distribution settled virtually uniquely with all its weight on the single parameter vector $(\boldsymbol{\alpha}_3, \boldsymbol{\sigma}_{p3}) = (0.4, 0.5, 0.1, 0.005, 0.010, 0.023)$

after only 2500 observations (weight = .96, which increases to .9999 by the end of
the observation sequence). In still another contrast, the mixture Normal forecast-
ing distribution MixN would have degenerated into a simple Normal distribution
quite early on in the time sequence. This distribution performs quite poorly
relative to the others we have assessed.

6.4 Self-assessed negentropy of the forecasts: their expected scores

We have described how the negative entropy of a forecasting distribution
represents its own assessment of expectation ("prevision" in de Finetti's ter-
minology) for the logarithmic score the distribution would achieve. It is also
a well-known measure of the amount of information inherent in a distribution,
and thus represents the assessor's understanding of the amount of information
his/her own uncertainty proclaims. Figure 6(a) shows that, like the scores of the
distributions themselves, the negentropies of the forecast distributions were fairly
indistinguishable prior to the sharp decline in prices on 19 October, 1987 (period
752 of our sequence). However, after that time, the information measure embed-
ded in MixLC3EP is slightly though noticeably larger than that in MixEP and
MixLC3N (which diminish in the order mentioned). However, the Negentropy
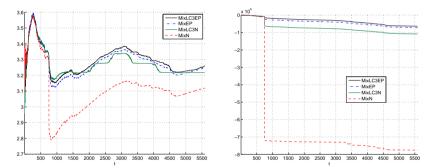measure for MixN is much smaller.

Figure 6: Sequence of the Negentropies and their quadratic scores. Left panel:
Sequential Negentropy information measures of the four forecasting distribu-
tions. Right panel: Cumulative quadratic scores of the daily Negentropies for
the four forecasting distributions.

It may be surprising to readers unaccustomed to the applied computation
of entropy measures that the "Negentropies" of the sequential forecasting distri-
butions, displayed in Figure 6(a), are all positive-valued! This is not an error.
Discrete probability distributions, described by a vector of probabilities $\mathbf{p}_N$, nec-
essarily entail a negative value to the function $\Sigma p_i \log(p_i)$. That is why "entropy"
was defined by Shannon (1948) with a negative sign on this function value, fol-
lowing Boltzmann, and ensuring that "entropy" is a positive-valued measure.

Negentropy, defined without the negative sign, is a negative-valued function of a probability mass function. In a continuous context, however, the entropy of a density, defined by an integral, is no longer assured of being positive-valued, nor is Negentropy necessarily negative-valued. In fact, in the context of the densities studied in this article, which are supported almost entirely over a very short domain, all display positive Negentropies.

6.5 Scoring the negative entropies: a clue to the success of MixLC3EP

For reasons of computational simplicity, we scored the assessed expectations of the log scores (their negative entropies) by a quadratic score. Figure 6(b) shows that once again, MixLC3EP achieves the best score of the four, followed in order by the scores of MixEP and MixLC3N, just as the entropy information measures themselves. Again, the score of the MixN distribution is in another league.

The purpose of conducting this kind of analysis is to exhibit that not only does the MixLC3EP distribution proclaim that it expects to achieve a better logarithmic score than do the others proclaim (i.e. asserts a greater expected score), it actually achieves a better score on a regular basis, as assessed by the quadratic score of this expectation. It turns out that this is a common feature in our experience: when finding that one distribution achieves a better logarithmic score than does another, it is common also to find that the preferred distribution both claims to entail more information, as measured by its Negentropy, and indeed it does achieve a better score in its self-assessed Negentropy.

7. Concluding Discussion

The computational procedures we followed for this article are based on mixing functions that are evaluated over what we deemed as appropriate discrete grids in a parameter space. While attention to the size of the grid in this multidimensional problem was required to make the computations feasible, we have performed robustness checks over finer grids with smaller portions of the data. These results suggest that extending the fineness of the grid would not generate computational differences of any practical significance. We are aware that developments in sequential Monte Carlo programming would allow us in principle to perform the computations in this mode as well. However the running time of Monte Carlo procedures for 5500 sequential forecasting distributions involving mixture distributions over 9 dimensions would be impractical. Recall that numerical integrations are required at each step to assess the information in the distributions via negentropy. We are pleased that the grid mixture strategy of computation we have followed allows practical insights into a substantive problem.

The largest methodological contribution of our programming strategy is to display the richness and the practicality of the method of sequential scoring of alternative probability distributions, for purposes of resolving scientific questions. This method was proposed by Bruno de Finetti in various writings. Computational details and practice have been developed in a tradition of researchers who subscribe to his basic point of view regarding probability as a representation of uncertainty.

We have computed much more information of interest than can be included in this article. Further information is available in a research report of the authors[1]. The most important extensions to the results discussed here are those that compare the assessments of the forecast distributions using the logarithmic scoring rule with assessments based on other proper rules. This matter is important, since the choice of scoring rules is based on utilitarian assessments of the consequences of forecast errors. However, the results from other scoring rules are similar. Also included in our computations were comparisons of the quality of maximum likelihood based forecasting relative to the fully integrated mixture distributions that we have reported on here. In brief, the mixture distributions do perform better than the maximum likelihood forecast procedure, although the differences do reduce as the data accumulates. They do not eliminate completely, because of the diffuseness that remains in the mixing distribution for MixLC3EP displayed in Appendix 3.

Acknowledgments

---

[1] Agrò, G., Sanfilippo, G. and Lad, F. (2007). www.unipa.it/ sanfilippo/seqforecast/ , click on CompanionFile.pdf .

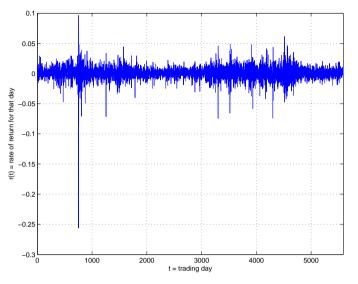## Appendix 1. Dow-Jones index daily rates of return

Figure 7: Data series of daily returns covering the period of 25 October, 1984 through 25 October, 2006

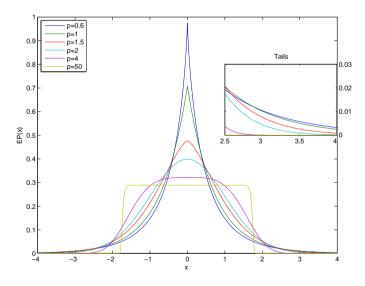## Appendix 2. Selected exponential power densities

Figure 8: EP densities for some values of p and SD=1.

Appendix 3. Sequential mixing function summary tables

Table 1: MixLC3EP. Point of maximum for the Posterior Mixing Function at time t.

t	p_1	p_2	p_3	s_1	s_2	s_3	α_1	α_2	α_3	**Mixing P Mass**
500	1.0	1.1	1.5	0.009	0.007	0.007	0.1	0.3	0.6	0.00104
1000	1.3	0.5	1.5	0.013	0.007	0.007	0.4	0.1	0.5	0.00890
1500	0.7	1.7	1.5	0.021	0.011	0.011	0.1	0.4	0.5	0.02586
2000	0.4	1.7	1.5	0.021	0.011	0.011	0.1	0.4	0.5	0.02689
2500	0.4	1.7	1.5	0.021	0.011	0.011	0.1	0.5	0.4	0.08480
3000	0.4	1.7	1.2	0.017	0.011	0.011	0.1	0.5	0.4	0.08396
3500	0.4	1.7	1.5	0.017	0.011	0.011	0.1	0.5	0.4	0.33464
4000	0.4	1.7	1.5	0.017	0.011	0.011	0.1	0.4	0.5	0.35594
4500	0.4	1.7	1.5	0.017	0.011	0.011	0.1	0.3	0.6	0.13044
5000	0.4	1.7	1.8	0.017	0.007	0.007	0.1	0.3	0.6	0.25091
5500	0.4	1.4	1.8	0.013	0.007	0.007	0.1	0.3	0.6	0.29455

Table 2: **MixLC3EP.** Number of points of the Posterior Mixing Function greater than a fixed value, P.

t	$P > 0.0001$	$P > 0.001$	$P > 0.01$	$P > 0.03$	$P > 0.1$
500	2550	3	0	0	0
1000	1635	231	0	0	0
1500	914	218	7	0	0
2000	597	190	17	0	0
2500	477	146	19	3	0
3000	326	104	22	7	0
3500	170	59	12	8	2
4000	147	61	15	7	2
4500	111	52	21	13	1
5000	107	30	11	6	3
5500	93	31	16	10	2

Table 3: MixLC3N. Point of maximum for the posterior mixing function at time t.

t	p_1	p_2	p_3	s_1	s_2	s_3	α_1	α_2	α_3	**Mixing P Mass**
500	2.0	2.0	2.0	0.005	0.015	0.015	0.6	0.3	0.1	0.10473
1000	2.0	2.0	2.0	0.005	0.023	0.023	0.4	0.5	0.1	0.54008
1500	2.0	2.0	2.0	0.005	0.023	0.023	0.4	0.5	0.1	0.96274
2000	2.0	2.0	2.0	0.005	0.023	0.023	0.4	0.5	0.1	0.85799
2500	2.0	2.0	2.0	0.005	0.023	0.023	0.5	0.4	0.1	0.96427
3000	2.0	2.0	2.0	0.005	0.023	0.023	0.6	0.3	0.1	0.97053
3500	2.0	2.0	2.0	0.005	0.023	0.023	0.5	0.4	0.1	0.98354
4000	2.0	2.0	2.0	0.005	0.023	0.023	0.5	0.4	0.1	0.86350
4500	2.0	2.0	2.0	0.005	0.023	0.023	0.4	0.5	0.1	0.99977
5000	2.0	2.0	2.0	0.005	0.023	0.023	0.4	0.5	0.1	0.99911
5500	2.0	2.0	2.0	0.005	0.023	0.023	0.4	0.5	0.1	0.99996

Table 4: MixLC3N. Number of points of the Posterior Mixing Function greater than a fixed value, P.

t	$P > 0.0001$	$P > 0.001$	$P > 0.01$	$P > 0.03$	$P > 0.1$
500	211	109	27	5	1
1000	23	12	4	3	3
1500	8	5	2	1	1
2000	6	2	2	2	2
2500	4	2	2	2	1
3000	6	2	2	1	1
3500	4	3	2	1	1
4000	4	2	2	2	2
4500	2	1	1	1	1
5000	2	1	1	1	1
5500	1	1	1	1	1

References

Agrò, G. (1995). Maximum likelihood estimation for the exponential power function parameters, *Communications in Statistics: simulation and computation*, **24**, 523-536.

Agrò, G. (1999). Parameter orthogonality and conditional profile likelihood: the exponential power function case, *Communications in Statistics: theory and methods*, **28**, 1759-1768.

Agrò, G. (2002). Portfolio value at risk using the exponential power functions, *University of Palermo, Department of Statistics research report*.

Bernardo, J. M. (1979). Expected information as expected utility, *Annals of Statistics*, **7**, 686-690.

Box, G. E. P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley.

Choy, S. T. B. and Walker, S. (2003). The extended exponential power distribution and Bayesian robustness, *Statistics and Probability Letters*, **65**, 227-232.

DeGroot, M. (1984). Changes in utility as information. *Theory and Decision*, **17**, 287-303. See errata noted in **18** (1985) 319.

de Finetti, B. (1937). La prévision, ses lois logiques, ses sources subjectives, *Annales de L'Institute Henri Poincaré*, **7**, pp. 1-68. H. Kyburg (tr.) Foresight, its logical laws, its subjective sources, in H. Kyburg and H. Smokler (eds.) *Studies in Subective Probability*, second edition, 1980, Krieger.

de Finetti, B. (1962). Does it make sense to speak of good probability appraisers?, in I. J. Good (ed.) *The Scientist Speculates: an anthology of half-baked ideas*, Basic Books.

de Finetti, B. *Theory of Probability: a critical introductory treatment* (1974,1975). 2 volumes, A.F.M. Smith and A. Machi (trs.), John Wiley. Translation of *Teoria della probabilita: sintesi introduttiva con appendice critica* (1970) Einaudi.

Gneiting, T. and Raftery, A. (1970). Strictly proper scoring rules, prediction and estimation, *Journal of the American Statistical Association*, **102**, 359-378.

Heath, D. and Sudderth, W. (1976). De Finetti's theorem for exchangeable random variables, *The American Statistician*, **30**, 188-189.

Lad, F. (1996). *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, John Wiley.

Lim, G. C., Martin, G. M. and Martin, V. L. (2006). Pricing currency options in the presence of time-varying volatility and non-normalities, *Journal of Multinational Financial Management*, **16**(3), 291-314.

Lim, G. C., Lye, J. N., Martin, G. M. and Martin, V. L. (1998). The distribution of exchange rate returns and the pricing of currency options, *Journal of International Economics*, **45**, 351-368.

Lopez, J. A. (1998). Methods for evaluating value-at-risk estimates, *Federal Reserve Bank of New York Research Paper* # 9802.

Mantegna, R. N. and Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index, *Nature*, **376**, 46-49.

Matheson, J. and Winkler, R. (1976). Scoring rules for continuous probability distributions, *Management Science*, **22**, 1087-1096.

Shannon, C. E. (1948) A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379-423, 623-653; reprint, *The Mathematical Theory of Communication*, C. E. Shannon and W. Weaver (eds.). University of Illinois Press, 1949.

Subbotin, M. T. (1923) On the law of frequency of error, *Mathematicheskii Sbornik*, 296-330.

Timmerman, A. (1995). Scales and stock markets, *Nature*, **376**, 18-19.

Vianelli, S. (1963). La misura della variabilità condizionata in uno schema generale della Curve Normali di frequenza, *Statistica*, **23**, 447-473.

Gianna Agrò
Dipartimento di Scienze Statistiche e Matematiche "S.Vianelli"
University of Palermo
Viale delle Scienze, ed. 13
90128 Palermo, Italy
agro@unipa.it

Frank Lad
Department of Mathematics and Statistics
University of Canterbury
Private Bag 4800
Christchurch, NEW ZEALAND, 8020
F.Lad@math.canterbury.ac.nz

Giuseppe Sanfilippo
Dipartimento di Scienze Statistiche e Matematiche "S.Vianelli"
University of Palermo
Viale delle Scienze, ed. 13
90128 Palermo, Italy
sanfilippo@unipa.it