1-2005

Smooth test for density

Aurobindo GHOSH
Singapore Management University, AUROBINDO@SMU.EDU.SG

Anil K BERA

DOI: https://doi.org/10.2139/ssrn.658861

Smooth Test for Density Forecast Evaluation

Aurobindo Ghosh
School of Economics and Social Sciences,
Singapore Management University,
469 Bukit Timah Road,
Singapore-259756. Singapore.
e-mail: aurobindo@smu.edu.sg. Phone: +65-6822-0863

Anil K. Bera
Department of Economics, University of Illinois ,
487 Wohlers Hall,
1206 S. Sixth Street,
Champaign, IL -61820. USA.
e-mail: abera@uiuc.edu. Phone: 217-333-4596

January 31, 2005

Keywords: Score test, probability integral transform, model selection, GARCH model, simulation based method, sample size selection

JEL Classification: C12, C52, C53

Abstract

Recently econometricians have shifted their attention from point and interval forecasts to density forecasts because at the heart of market risk measurement is the forecast of the probability density functions of various financial variables. In this paper, we propose a *formal test* for density forecast evaluation based on Neyman's smooth test procedure. Apart from accepting or rejecting the tested model, this approach provides specific sources (such as the location, scale and shape of the distribution) of rejection, thereby helping in deciding possible modifications of the assumed model. Our applications to S&P 500 returns indicate capturing time-varying volatility and non-gaussianity significantly improve the performance of the model.

1 Introduction and Motivation

In the estimation literature in statistics there was a natural progression of point estimation to interval estimation, and then to the full (non-parametric) density estimation. In the context of time series forecasting, we also observe similar pattern of advancement from point forecast to interval forecast, and then finally to density forecast, though construction of density forecast in empirical work is a recent phenomenon. It is, therefore, not surprising that evaluating density forecast techniques is in its infancy. There has been only a few papers, we are aware of, that directly address the question of evaluation of density forecasts; such as Diebold, Gunther and Tay (1998), Berkowitz (2001), Hong (2001), Wallis (2003) and Sarno and Valente (2004). The importance of density forecast evaluation cannot be overemphasized. Recent developments in risk evaluation clearly indicate that we can no longer rely on a few moments or certain regions of the distribution; very often we will need to forecast the *entire* distribution. Also, as demonstrated by Diebold et al. (1998) and Granger and Pesaran (2000), only when a forecast density coincides with the true data generating process, then that forecast density will be preferred by all forecast users regardless of their attitude to risk (loss function). The importance of density forecast evaluation in economics has been aptly depicted by Crnkovic and Drachman (1997, p. 47) as follows: "At the heart of market risk measurement is the forecast of the probability density functions (PDFs) of the relevant market variables ... a forecast of a PDF is the central input into any decision model for asset allocation and/or hedging ... therefore, the quality of risk management will be considered synonymous with the quality of PDF forecasts."

Anderson (1994) showed that a modified version of Pearson χ^2 statistic could be decomposed into components directed at different moments of the original data (Boero, Smith and Wallis 2004). He proposed a class of "Pearson analog" χ^2 tests that can be used both against a general alternative hypothesis (omnibus test) as well as for more specific alternatives (directional test) using some very simple (though ad hoc) set of orthogonal polynomials. He proposed that a necessary condition for a locally optimal test where the distributions under H_0 and H_1 differ in the k^{th} moment is that there should be k intersections between the two distributions, essentially indicating that the number of class intervals should be at least one more than the moments to be tested. He addressed the issue of size distortion of traditional score types tests like Pearson χ^2, and also showed simulation results to illustrate that the size corrected power of the modified Pearson type test is better than the traditional

Jarque-Bera test of normality. Boero et al. (2004) questioned the generality of the modified or "Pearson analog" tests proposed by Anderson (1994). The main reason being the independence assumption of the different components of the test directed towards location, scale, skewness or shape parameters of the distribution is violated when the classes are non-equiprobable.

Sarno and Valente (2004) suggested a test based on integrated squared difference of the kernel density functions of the competing predictive density forecast models, using a norm similar to Li (1996) as discussed in Pagan and Ullah (1999, pp. 68-69). The asymptotically normal test statistic thus obtained is a natural analog of Diebold-Mariano test (Diebold and Mariano 1995) for forecast accuracy in the domain of point forecasts. The simulation results reported shows attractive size and power properties with very little, if any, size distortion. The test statistic requires bootstrap replications in order to calculate its standard error, and would be unsuitable for either applications in a time series context with time-dependent parameters, or for adaptive model selection for finding the "best" model. Giacomini (2002) explored weighted likelihood ratio tests proposed originally by Vuong (1989) for non-nested hypotheses to compare competing, possibly misspecified, models of density forecast using decision theory based methods or "scoring rules" (Granger and Pesaran 1996, 2000).

From a pure statistical perspective, density forecast evaluation is essentially a goodness-of-fit test problem. In a seminal paper, though never used directly in econometrics, Neyman (1937) demonstrated how "all" goodness-of-fit testing problems can be converted into testing only *one kind of hypothesis*. Specifically Neyman considered the probability integral transform (PIT) of the density $f(x)$. Under the null hypothesis of correct specification of $f(x)$, PIT is distributed as $U(0,1)$ irrespective of the form of $f(x)$. As an alternative to the $U(0,1)$ density, Neyman specified a *smooth* density using normalized Legendre polynomials. A major benefit of Neyman's formulation is that in addition to a formal test procedure we can identify the specific sources of rejection when the data is not compatible with the tested density function. Therefore, Neyman's smooth test provides natural guidance to specific directions to revise a model. The purpose of the paper is to use Neyman's idea to devise a *formal test* for density forecast evaluation.

The plan of the paper is as follows. In Section 2 we review Neyman (1937) smooth test approach. For a fuller account see Rayner and Best (1989) and Bera and Ghosh (2001). Section 3 uses the framework of Diebold et al. (1998) and proposes a smooth test for density forecast evaluation. An application to S&P 500 returns data is given in Section 4. Section 5 provides some Monte Carlo results to examine size properties of the proposed test. Section 6 concludes.

2 Neyman Smooth Test

We want to test the null hypothesis (H_0) that our assumed density $f(x)$ is the true density function for the random variable X, based on n independent observations $x_1, x_2, ..., x_n$. The specification of $f(x)$ will be *different* depending on the problem at hand. Neyman (1937, pp. 160-161) first transformed *any* hypothesis testing problem of this type to testing only *one kind of hypothesis* using the probability integral transform (PIT). Neyman suggested this test to rectify some of the drawbacks of Pearson's (1900) goodness-of-fit statistic [see Bera and Ghosh (2001) for more on this issue, and for a historical perspective], and called it a *smooth* test since the alternative density is close to the null density and has few intersections with the null density.

We construct a new random variable Y by defining $Y_i = F(X_i)$, $i = 1, 2, ..., n$, that is, the probability integral transform (PIT)

$$y_i = \int_{-\infty}^{x_i} f(u|H_0)\, du \equiv F(x_i). \tag{1}$$

Suppose under the alternative hypothesis, the density and the distribution functions of X is given by $g(.)$ and $G(.)$, respectively. Then, in general, the distribution function of Y is given by

$$\begin{aligned} H(y) &= \Pr(Y \leq y) = \Pr(F(X) \leq y) \\ &= \Pr\left(X \leq F^{-1}(y)\right) = G\left(F^{-1}(y)\right) \\ &= G(Q(y)), \end{aligned} \tag{2}$$

where $Q(y) = F^{-1}(y)$ is the quantile function of Y. Therefore, the density of Y can be written as [see Bera and Ghosh (2001, p. 185)]

$$h(y) = \frac{d}{dy} H(y) = g(Q(y)) \frac{d}{dy} F^{-1}(y) = \frac{g(Q(y))}{f(Q(y))}, \ 0 < y < 1. \tag{3}$$

Although this is the ratio of two densities, $h(y)$ is a proper density function when F and G are strictly increasing functions. We will call $h(.)$ the *ratio density function* (RDF) since it is both a ratio of two densities and a density function itself. When

$f(.)$ is the true density we have $Y \sim U(0,1)$. And, under the alternative hypothesis $h(y)$ will differ from 1 and that provides a basis for the Neyman smooth test.

Neyman (1937, p. 164) considered the following smooth alternative to the uniform density:

$$h(y) = c(\theta) \exp\left[\sum_{j=1}^{k} \theta_j \pi_j(y)\right], \tag{4}$$

where $\theta = (\theta_1, \theta_2, ..., \theta_k)'$, $c(\theta)$ is the constant of integration and $\pi_j(y)$ are orthonormal polynomials of order j satisfying

$$\int_0^1 \pi_i(y)\, \pi_j(y)\, dy = \delta_{ij}, \quad \text{where } \delta_{ij} \begin{array}{l} = 1 \quad if \quad i = j \\ = 0 \quad if \quad i \neq j. \end{array} \tag{5}$$

Under $H_0 : \theta_1 = \theta_2 = ... = \theta_k = 0$, since $c(\theta) = 1$, $h(y)$ in (4) reduces to the uniform density.

Under the alternative, we take $h(y)$ as given in (4) and test $\theta_1 = \theta_2 = ... = \theta_k = 0$. Therefore, the test utilizes (3) which looks more like a "likelihood ratio". To get an idea of the the exact nature of $h(y)$, let us consider some particular cases. When the two distributions differ only in location; for example, $f(.) \equiv \mathcal{N}(0,1)$ and $g(.) \equiv N(\mu, 1)$, $\ln(h(y)) = \mu y - \frac{1}{2}\mu^2$, which is *linear* in y. Similarly, if the distributions differ in scale parameter, such as, $f(.) \equiv \mathcal{N}(0,1)$ and $g(.) \equiv \mathcal{N}(0, \sigma^2)$, $\sigma^2 \neq 1$, $\ln(h(y)) = \frac{y^2}{2}\left[1 - \frac{1}{\sigma^2}\right] - \frac{1}{2}\ln\sigma^2$, a quadratic function of y. Considering some commonly used non-normal densities as alternatives, we note that $f(.) \equiv \mathcal{N}(0,1)$ and $g(.) \equiv \chi_4^2$ yield $\ln(h(z)) = \frac{1}{2}z^2 - \frac{1}{2}z + \ln z + \ln\left(\frac{\sqrt{2\pi}}{4}\right)$. If we have $f(.) \equiv \mathcal{N}(0,1)$ and $g(.) \equiv t_4$, then we have $\ln(h(z)) = \frac{z^2}{2} + \frac{5}{2}\ln\left[1 + \frac{z^2}{4}\right] + \ln\left(\frac{\sqrt{2\pi}}{2}\right)$. These illustrative examples suggest that departures from the null hypothesis can be tested using an appropriate function (or functions) estimating the RDF, $h(y)$. From observing the plots of the different ordered normalized Legendre polynomials, we believe that the test will not only be powerful but also informative on identifying particular source(s) of departure(s) from H_0 (Ghosh 2003).

Using the multiparameter version of the generalized Neyman-Pearson lemma, Neyman (1937) derived the locally most powerful unbiased (LMPU) symmetric test for $H_0 : \theta_1 = \theta_2 = ... = \theta_k = 0$ against the alternative H_1 : at least one $\theta_i \neq 0$, for small values of $\theta_i' s$. The test is symmetric in the sense that the asymptotic power of the test depends only on the Euclidean distance,

$$\lambda = \left(\theta_1^2 + ... + \theta_k^2\right)^{\frac{1}{2}}, \tag{6}$$

between H_0 and H_1. The test statistic is

$$\Psi_k^2 = \sum_{j=1}^{k} \frac{1}{n} \left[\sum_{i=1}^{n} \pi_j (y_i) \right]^2 , \tag{7}$$

which under H_0 asymptotically follows a χ_k^2, and under H_1 follows a non-central χ_k^2 with non-centrality parameter λ^2.

We now show that the test statistic Ψ_k^2 can be simply obtained using Rao's (1948) score (RS) test principle. Taking (4) as the PDF under the alternative hypothesis, the log-likelihood function $l(\theta)$ can be written as

$$l(\theta) = n \ln c(\theta) + \sum_{j=1}^{k} \theta_j \sum_{i=1}^{n} \pi_j (y_i) . \tag{8}$$

The RS test for testing the null $H_0 : \theta = \theta_0$ is given by

$$RS = s(\theta_0)' \mathcal{I}(\theta_0)^{-1} s(\theta_0) , \tag{9}$$

where $s(\theta)$ is the score vector $\partial l(\theta)/\partial\theta$, $\mathcal{I}(\theta)$ is the information matrix $E\left[-\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right]$ and in our case, $\theta_0 = \mathbf{0}$. It is easy to see that

$$\begin{aligned} s(\theta_j) &= \frac{\partial l(\theta)}{\partial \theta_j} \\ &= n \frac{\partial \ln c(\theta)}{\partial \theta_j} + \sqrt{n} u_j, \;\; j = 1, 2, ..., k, \\ \text{with } u_j &= \sum_{i=1}^{n} \pi_j (y_i) / \sqrt{n}. \end{aligned} \tag{10}$$

Differentiating the identity $\int_0^1 h(z)\, dz = 1$ with respect to θ_j, we have

$$\frac{\partial c(\theta)}{\partial \theta_j} \int_0^1 \exp\left[\sum_{j=1}^{k} \theta_j \pi_j (y)\right] dy + c(\theta) \int_0^1 \exp\left[\sum_{j=1}^{k} \theta_j \pi_j (y)\right] \pi_j (y)\, dy = 0. \tag{11}$$

Evaluating (11) under $\theta = 0$, we have $\left.\frac{\partial \ln c(\theta)}{\partial \theta_j}\right|_{\theta=\mathbf{0}} = \left.\frac{\partial c(\theta)}{\partial \theta_j} \times \frac{1}{c(\theta)}\right|_{\theta=\mathbf{0}} = 0$, and therefore, under the null hypothesis

$$s(\theta_j) = \sqrt{n} u_j. \tag{12}$$

To get the information matrix, let us first note from (10) that

$$\frac{\partial^2 l(\theta)}{\partial\theta_j\partial\theta_l} = n\frac{\partial^2 \ln c(\theta)}{\partial\theta_j\partial\theta_l}, \tag{13}$$

which is a constant. Therefore, under H_0 the $(j,l)^{th}$ element of the information matrix $\mathcal{I}(\theta)$ is simply $-n\partial^2 \ln c(\theta)/\partial\theta_j\partial\theta_l$ evaluated at $\theta = \mathbf{0}$. Differentiating (11) with respect to θ_l and evaluating it at $\theta = \mathbf{0}$, after some simplification, we have

$$\left.\frac{\partial^2 c(\theta)}{\partial\theta_j\partial\theta_l}\right|_{\theta=0} + \int_0^1 \pi_j(y)\,\pi_l(y)\,dy = 0. \tag{14}$$

Using the orthonormal property in (5)

$$\left.\frac{\partial^2 c(\theta)}{\partial\theta_j\partial\theta_l}\right|_{\theta=0} = -\delta_{jl}. \tag{15}$$

Further, using (11), $c(\theta) = 1$, $\frac{\partial c(\theta)}{\partial\theta_j}$ and $\frac{\partial c(\theta)}{\partial\theta_j} = 0$ for any j, we have

$$\frac{\partial^2 \ln c(\theta)}{\partial\theta_j\partial\theta_l} = \frac{\partial}{\partial\theta_l}\left(\frac{\partial c(\theta)}{\partial\theta_j}\frac{1}{c(\theta)}\right) = \frac{\frac{\partial^2 c(\theta)}{\partial\theta_j\partial\theta_l}c(\theta) - \frac{\partial c(\theta)}{\partial\theta_j}\frac{\partial c(\theta)}{\partial\theta_l}}{(c(\theta))^2},$$

and, hence

$$\mathcal{I}(\theta_0) = nI_k, \tag{16}$$

where I_k is a $k \times k$ identity matrix. Combining (9), (12) and (16) the RS test statistic has the simple form

$$RS = \sum_{j=1}^{k} u_j^2. \tag{17}$$

Neyman's approach was to compute the smooth test statistic in terms of the probability integral transform Y defined in (1). It is, however, easy to recast the testing problem in terms of the original observations on X and PDF, say, $f(x;\gamma)$. Writing (1) as $y = F(x;\gamma)$ and defining $\pi_i(y) = \pi_i(F(x;\gamma)) = q_i(x;\gamma)$, we can express the orthogonality condition (5) as

$$\int_0^1 \{\pi_i(F(x;\gamma))\}\{\pi_j(F(x;\gamma))\}\,dF(x;\gamma) = \int_0^1 \{q_i(x;\gamma)\}\{q_j(x;\gamma)\}\,f(x;\gamma)\,dx = \delta_{ij}. \tag{18}$$

Then, from (4) the density under the alternative hypothesis takes the form

$$\begin{aligned} g(x;\gamma,\theta) &= h(F(x;\gamma))\frac{dy}{dx} \\ &= c(\theta;\gamma)\exp\left[\sum_{j=1}^{k}\theta_j q_j(x;\gamma)\right] f(x;\gamma). \end{aligned} \tag{19}$$

Under this formulation we have the same test statistic Ψ_k^2, but now written in terms of the original observations, $x_1, x_2, ..., x_n$:

$$\Psi_k^2 = \sum_{j=1}^{k}\frac{1}{n}\left[\sum_{i=1}^{n} q_j(x_i;\gamma)\right]^2. \tag{20}$$

In order to implement this we need to replace the nuisance parameter γ by an efficient estimate $\hat{\gamma}$, and that will not change the asymptotic distribution of the the test statistic (Thomas and Pierce 1979), although there could be some possible change in the variance of the test statistic [see, for example, Boulerice and Ducharme (1995)].

3 Smooth Test for Density Forecast Evaluation

Suppose that we have time series data (say, the daily returns to the S&P 500 Composite Index) given by $\{x_t\}_{t=1}^m$. One of the most important questions that we would like to answer is, what is the sequence of the true density functions $\{g_t(x_t)\}_{t=1}^m$ that generated this particular realization of the data? At time t we know all the past values of x_t, i.e., the *information set* at time t is $\Omega_t = \{x_{t-1}, x_{t-2}, ...\}$. Let us denote the one-step-ahead forecast of the sequence of densities as $\{f_t(x_t)\}$ conditional on Ω_t. Our objective is to determine to what extent the forecast density $\{f_t\}$ depicts the true density $\{g_t\}$. The main problem in performing such a test is that both the actual density $g_t(.)$ and the one-step-ahead predicted density $f_t(.)$ could depend on the time t and, thus, on the information set Ω_t. This problem is unique, since, on one hand, it is a classical goodness-of-fit problem but, on the other, it is also a combination of several different, possibly dependent, goodness-of-fit tests.

One approach to handling this particular problem would be to reduce it to a more tractable one in which we have the same, or similar, hypotheses to test, rather than a host of different hypotheses. Following Neyman (1937) this is achieved using the probability integral transform

$$y_t = \int_{-\infty}^{x_t} f_t(u)\, du. \tag{21}$$

which has the density function

$$h_t(y_t) = 1,\ 0 < y_t < 1, \tag{22}$$

under the null hypothesis $H_0 : g_t(.) = f_t(.)$, i.e., our forecasted density is the true density.

If we are only interested in performing a goodness-of-fit test that the variable y_t follows a uniform distribution, we can use a parametric test like Pearson's χ^2 on grouped data or non-parametric tests like the Kolmogorov-Smirnov (KS) or the Cramér-von Mises (CvM) or a test using the Kuiper statistics (see Crnkovic and Drachman 1997, p. 48). Any of these suggested tests would work as a good *omnibus* test of goodness-of-fit. If we fail to reject the null hypothesis we can conclude that there is not enough evidence that the data is *not* generated from the forecasted density $f_t(.)$; however, a rejection would not throw any light on the possible form of the true density function.

The fundamental basis of Neyman's smooth test is the result that when $x_1, x_2, ..., x_n$ are independent and identically distributed (*IID*) with a common density $f(.)$, then the probability integral transforms $y_1, y_2, ..., y_n$ defined in equation (21) are *IID*, $U(0,1)$ random variables. In econometrics, however, we very often have cases in which $x_1, x_2, ..., x_n$ are not *IID.* In that case we can use Rosenblatt's (1952) generalization of the above result.

Theorem 1 (Rosenblatt) *Let* $(X_1, X_2, ..., X_n)$ *be a random vector with absolutely continuous density function* $f(x_1, x_2, ..., x_n)$. *Then, the* n *random variables defined by*

$$Y_1 = P(X_1 \leq x_1), Y_2 = P(X_2 \leq x_2 | X_1 = x_1),$$
$$..., Y_n = P(X_n \leq x_n | X_1 = x_1, X_2 = x_2, ..., X_{n-1} = x_{n-1})$$

are IID $U(0,1)$.

The above result can immediately be seen using the Change of Variable theorem that gives

$$
\begin{aligned}
P\left(Y_i \leq y_i, i=1,2,...,n\right) &= \int_0^{y_1}\int_0^{y_2}...\int_0^{y_n} f\left(x_1\right)dx_1 f\left(x_2|x_1\right)dx_2 ... f\left(x_n|x_1,...,x_{n-1}\right)dx_n \\
&= \int_0^{y_1}\int_0^{y_2}...\int_0^{y_n} dt_1 dt_2 ... dt_n \qquad (23)\\
&= y_1 y_2 ... y_n.
\end{aligned}
$$

Hence, $Y_1, Y_2, ..., Y_n$ are *IID* $U\left(0,1\right)$ random variables.

Diebold et al. (1998) used Theorem 1, to test $H_0 : g_t\left(.\right) = f_t\left(.\right)$ by checking whether the probability integral transform y_t in (21) follows *IID* $U\left(0,1\right)$. They employed a graphical (visual inspection) approach to decide on the structure of the alternative density function by a two-step procedure. First, they visually inspected the histogram of y_t to see if it comes from $U\left(0,1\right)$ distribution. Then, they looked at the individual correlograms of each of the first four powers of the variable $z_t = y_t - 0.5$ in order to check for any residual effects of bias, variance or higher-order moments. In the absence of a more analytical test of goodness-of-fit, this graphical method has also been used in Diebold, Tay and Wallis (1999) and Diebold, Hahn and Tay (1999). For more on interval and density forecasting along forecast evaluation methods, see Diebold and Lopez (1996), Christofferson (1998) and Tay and Wallis (2000). The procedure suggested is very attractive due to its simplicity of execution and intuitive justification; however, the resulting size and power of this informal procedure is unknown.

Neyman's smooth test provides an *analytic* tool to determine the structure of the density under the alternative hypothesis using orthonormal polynomials (normalized Legendre polynomials). Specifically, Neyman used $\pi_j\left(y\right)$ as the orthogonal polynomials that can be obtained by using the following conditions,

$$
\pi_j\left(y\right) = a_{j0} + a_{j1}y + ... + a_{jj}y^j, a_{jj} \neq 0,
$$

given the restrictions of orthogonality given in (5). Solving these the first five $\pi_j\left(y\right)$ are (Neyman 1937, pp. 163-164) $\pi_0\left(y\right) = 1$, $\pi_1\left(y\right) = \sqrt{12}\left(y - \frac{1}{2}\right)$, $\pi_2\left(y\right) = \sqrt{5}\left(6\left(y-\frac{1}{2}\right)^2 - \frac{1}{2}\right)$, $\pi_3\left(y\right) = \sqrt{7}\left(20\left(y-\frac{1}{2}\right)^3 - 3\left(y-\frac{1}{2}\right)\right)$, $\pi_4\left(y\right) = 210\left(y-\frac{1}{2}\right)^4 - 45\left(y-\frac{1}{2}\right)^2 + \frac{9}{8}$.

While, on one hand, the smooth test provides a basis for a classical goodness-of-fit test, on the other hand, it can also be used to determine the sensitivity of the power of the test to departures from the null hypothesis in different directions, for example, deviations in scale (variance) and the shape of the distribution (skewness and kurtosis). We can see that the Ψ_k^2 statistic for Neyman's smooth test defined in equation (7) is comprised of k components of the form $\frac{1}{n}\left(\sum_{i=1}^{n} \pi_j(y_i)\right)^2, j = 1, ..., k,$ which are nothing but the squares of the efficient score functions. Using Rao(1948) and Neyman (1959) one can risk the "educated speculation" that an *optimal test* should be based on the *score function* [for more on this, see Bera and Billias (2001a, b)]. From that point of view, we achieve *optimality* using the smooth test.

There is one more issue that is central to any test applied to real data when the density function $f(.)$ under the null hypothesis is completely unknown. Hence, we have to estimate the PDF generating the data using an estimation sample. Let us assume that we know a general functional form of the density function $f(.;\beta)$ generating the data but have to estimate the parameter β based on the estimation sample of size m. As we mentioned earlier our test is based on a sample of size n. The "true" test statistic is given in (7), with

$$y_i = F(x_i;\beta) = \int_0^{x_i} f(u;\beta)\, du,\ \ i = 1, 2, ..., n. \tag{24}$$

However, since we do not know the true value of β, we estimate it using $\hat{\beta}$ to get

$$\hat{\Psi}_k^2 = \sum_{j=1}^{k} \hat{u}_j^2 = \sum_{j=1}^{k} \frac{1}{n}\left(\sum_{i=1}^{n} \pi_j(\hat{y}_i)\right)^2, \tag{25}$$

where $\hat{y}_i = F\left(x_i;\hat{\beta}\right) = \int_0^{x_i} f\left(u;\hat{\beta}\right) du$, $i = 1, 2, ..., n$, are the estimated PITs and $\hat{\beta}$ is any $\sqrt{m}-$consistent estimator of β. We have the following theorem which shows that for certain values of m and n, we can ignore the effect of parameter estimation on our results.

Theorem 2 *Let m and n be the estimation and test sample sizes, respectively, $\hat{\beta}$ be a $\sqrt{m}-$consistent estimator of the parameter β and $E\left[\frac{d\pi_j(F(x_i;\beta))}{d\beta}\right] < \infty$. Then, if $n = O\left(m^{\frac{1}{2}}\right)$, under the null hypothesis H_0, $\hat{\Psi}_k^2 - \Psi_k^2 = o_p(1)$.*

Proof. See Appendix A. ∎

4 Application to Asset Returns on S&P 500

We consider the daily returns on the value-weighted S&P 500 Composite Index from July 3, 1962 to December 31, 2003. The sample is split into in-sample and out-of-sample periods for model estimation and density forecast evaluation. There are 8431 in-sample observations (07/03/62-12/29/95) and 2016 out-of-sample observations (01/02/96-12/31/2003). In order to obtain a test with desirable actual size using the smooth test principle, we chose a significantly smaller sample size for the evaluation sample compared to the estimation sample. Diebold et al. (1998) also used daily data on the value-weighted S&P 500 returns with dividends, from 02/03/62 through 12/29/95 in order to demonstrate the effectiveness of a graphical procedure based on the probability integral transform, however in their case the sample split was at the middle of the data range. Figure 1 compares the density estimates between the in-sample and the out-of-sample data.

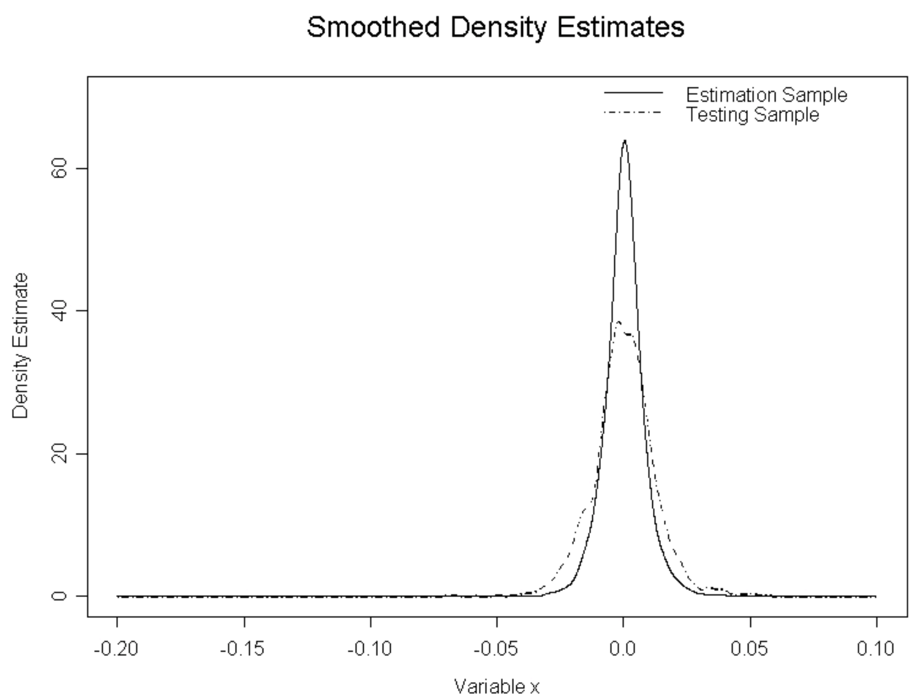

Figure 1: Kernel Density Estimates of S&P 500 Returns

Following Diebold et al. (1998), we used progressively richer models to find the best model to fit the estimation sample and then freeze it to do forecasting of the evaluation data. Using the empirical distribution function (EDF) of the estimation sample, we generate the PIT of the evaluation data and present an estimate of its density (histogram) in Figure 2 From a visual analysis of the histogram it is clear that the PITs do not seem to follow an $U(0,1)$ distribution, the conclusion is more apparent if we compare the PDF of $U(0,1)$ distribution with the *ratio density function* (RDF) of the PIT (Bera, Ghosh and Xiao 2004). In order to better fit the model for forecasting future observations, we use a naive MA(1), MA(1)- normal-GARCH(1,1) and finally, a MA(1)-t-GARCH(1,1) model to the estimation sample where the degrees of freedom of the t-distribution is obtained through maximum likelihood method. From visual analysis of the histograms we can infer that introducing a time varying conditional heteroskedasticty term clearly improves the forecast and it also causes the histograms of the PITs to be closer to that of an $U(0,1)$ PDF. However, the improvement is not very apparent with the introduction of a non-Gaussian error term (Figure 4 and Figure 5).

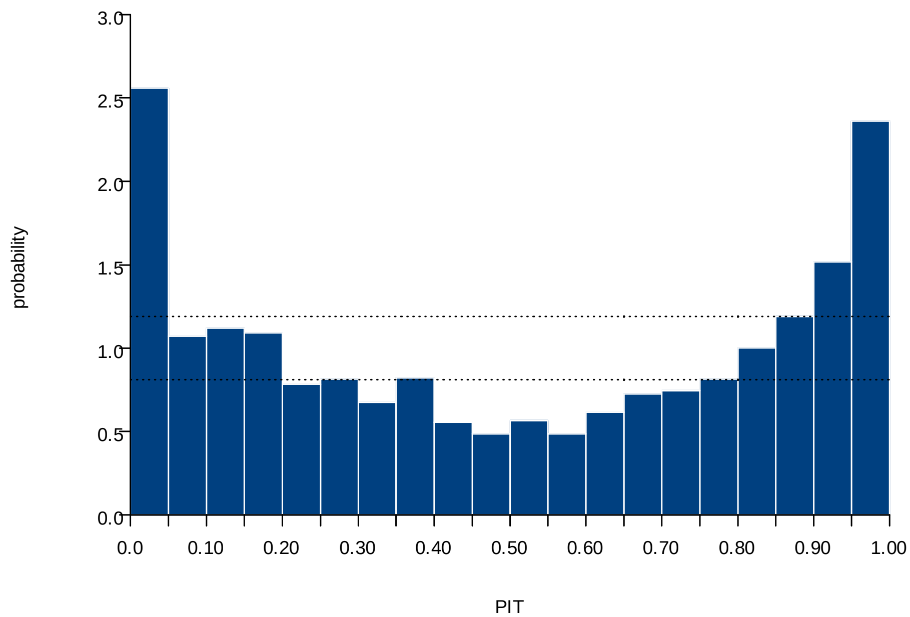

Figure 2: Histogram for the probability integral transforms using EDF

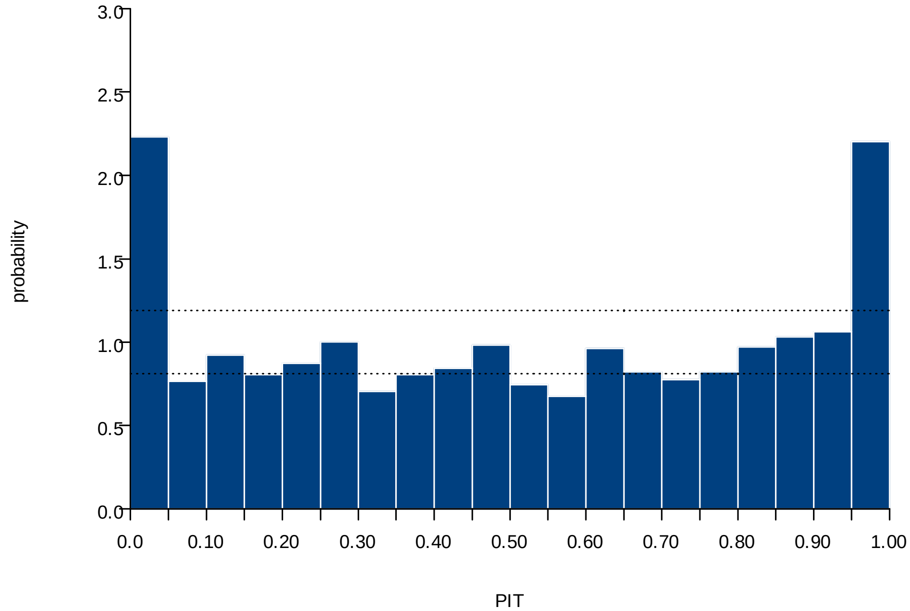

Fig. 3: Histogram for the PIT with MA(1)-normal model

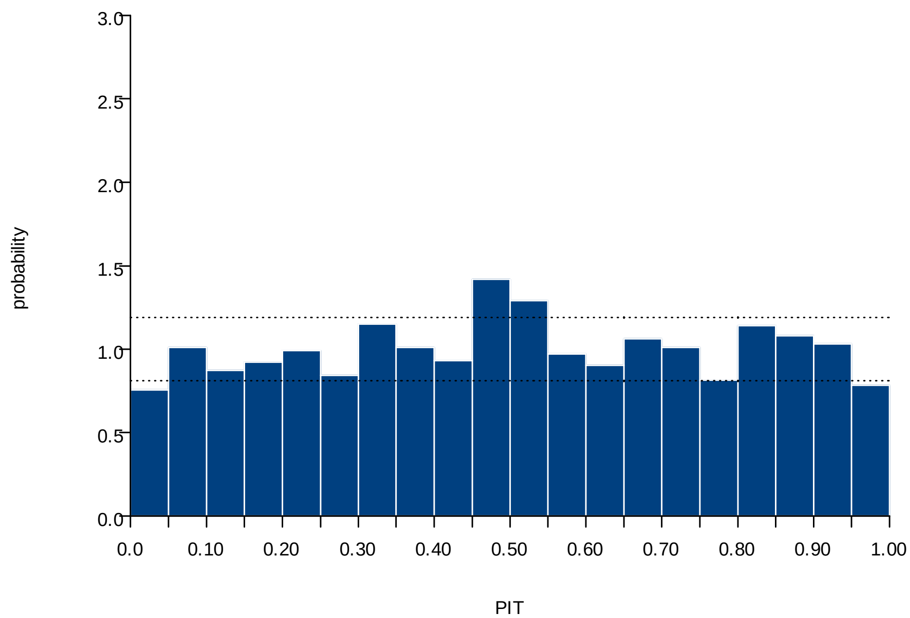

Figure 4: Histogram for PIT with MA(1)-normal GARCH(1,1) model

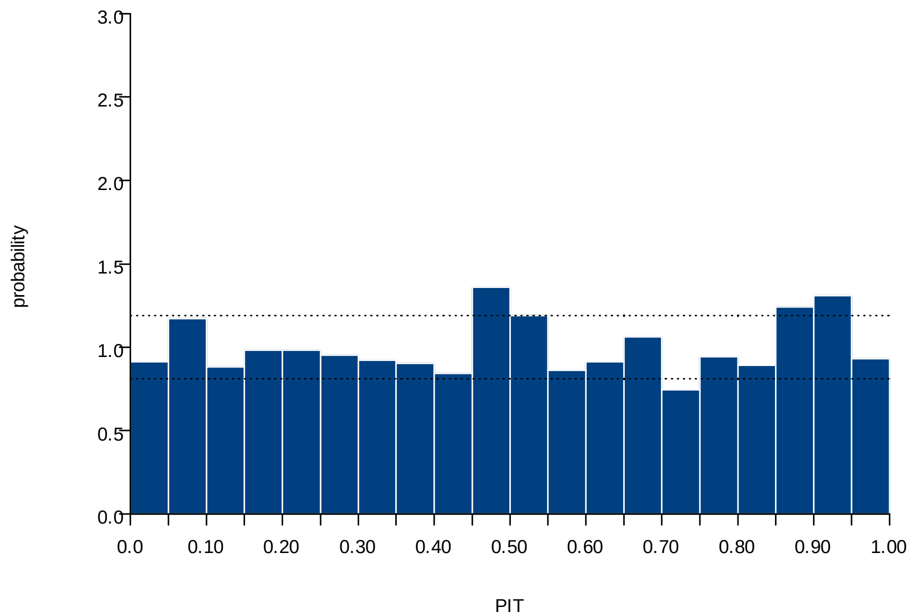

Figure 5: Histogram for the PIT with MA(1)-t-GARCH (1,1)

Data	Estimation	Test
Observations	8431	2016
Mean	0.00032	0.00037
Standard Deviation	0.00858	0.01246
Skewness Coefficient	-1.5624	-0.0089
Excess Kurtosis	43.7935	2.3472
Minimum	-0.20467	-0.06867
1^{st} Quartile	-0.00394	-0.00649
Median	0.00036	0.00039
3^{rd} Quartile	0.00457	0.00744
Maximum	0.09099	0.05731

Table 3. Return distribution for estimation and test samples

	Test Statistic	Critical Values Upper .1%
D^+	4.19843	1.859
D^-	4.89182	1.859
KS	4.89182	1.95
Kuiper	9.0979	2.303
CvM	10.62024	1.167
A-D	94.37819	6.0
W	10.60013	0.385

Table 4. Goodness-of-Fit statistics based on EDF with $m = 8431$ and $n = 2016$, Critical values are from D'Agostino and Stephens (1986).

As attractive as it may seem, this graphical procedure is a subjective method of identifying the problems of a forecasted PDF after comparison with the true distribution (See Figure 1). This also implies that we cannot evaluate the performance of such an informal test of hypothesis with other existing tests of goodness-of-fit like the Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM) or Anderson-Darling (A-D) reported in Table 4 in terms of size and power characteristics. Although, to do full justice to the precursor of the current paper we should also mention that Berkowitz (2001, p. 466) commented on the Diebold et al.(1998) procedure: "Because their interest centers on developing tools for diagnosing *how* models fail, they do not pursue formal testing."

Our aim is to use a formal test using Neyman's smooth test principle. We use order $k = 4$ which we believe is sufficient to capture most of the global characteristics of distribution of value-weighted S&P 500 returns. In Table 5, we report the results of the smooth test.

Hypothesis	$\hat{\Psi}_4^2$	$\hat{u}_1^2$	$\hat{u}_2^2$	$\hat{u}_3^2$	$\hat{u}_4^2$
EDF	608.2575***	0.2304	522.0063***	0.0197	86.0012***
	(0.00000)	(0.63123)	(0.00000)	(0.88843)	(0.00000)
MA(1)	390.3732***	1.6088	203.3362***	0.2192	185.209***
with Normal error	(0.00000)	(0.20466)	(0.00000)	(0.63966)	(0.00000)
MA(1)-	13.4074***	1.0692	12.137***	0.1364	0.0648
Normal GARCH (1,1)	(0.00945)	(0.30112)	(0.00049)	(0.71188)	(0.79905)
MA(1)-	1.8544	1.0572	0.3445	0.3837	0.0691
t_8 GARCH (1,1)	(0.76252)	(0.30386)	(0.55722)	(0.53566)	(0.79272)

*** *significant at* 1% *level.*

Table 5.Neyman's smooth statistics and components (p-values are in parenthesis).

Initially, we used the empirical distribution function of the estimation sample to calculate the PIT of each observation of the test sample and computed the smooth test statistic. We should mention that this is a non-parametric procedure since we do not assume any structure of the underlying PDF generating the model. However, this does not take account of the dependent structure of the data. Using an order $k = 4$, we get a score test statistic of 608.2575 which is statistically highly significant. We also can identify that the main sources of this deviation in the overall $\hat{\Psi}_4^2$ statistic are the second ($\hat{u}_2^2$) and fourth ($\hat{u}_4^2$) components. From analyzing this we can infer that, there are departures, mainly, in the directions of the second and the fourth order polynomials, which in turn would indicate the sources of departure are most likely in the second and fourth moments. Therefore, through pure non-parametric estimation of the EDF with no assumption of time varying conditional heteroskedasticty, we can conclude that there are deviations in the directions of the second and fourth order polynomials that can be related to second and fourth moments. We should also point out that the nature of the normalized Legendre polynomials indicate that the second order term is present in the fourth order polynomial, hence it would be difficult to pin point whether the main direction of departure is in the second or the fourth moments of the distribution.

At the next stage to start with a simple parametric model, we estimate an MA(1) model with Gaussian error terms, and we obtain a highly significant $\hat{\Psi}_4^2$ statistic of 390.3732. The discrepancy from the null hypothesis seems to be again in the directions of the second ($\hat{u}_2^2$ =203.33619) and fourth ($\hat{u}_4^2$ =185.20897) orders polynomials.

However, in this case the discrepancy in the fourth order term seems to be more pronounced than the purely non-parametric case. We still do not find the third order term to be statistically significant. Keeping this result in mind, we proceed to incorporate a time varying volatility model through a GARCH(1,1) model for conditional heteroskedasticty keeping the MA(1) component for the conditional mean (or level) equation with Gaussian errors. This more general framework nests the previously used naive MA(1) model with normal errors. The $\hat{\Psi}_4^2$ statistic is now reduced substantially (390.3732 to 13.4074), although it is still highly significant at the 1% level. A cursory inspection of the components revealed that only the second component is still significant although by a much lesser degree ($\hat{u}_2^2$ is now 12.137 compared to the earlier value of 203.3362). Therefore, introduction of conditional heteroskedasticty into the forecast density model substantially improves its performance. Finally, we introduce a non-Gaussian error term in the form of Student's t distribution along with the MA(1)-GARCH(1,1) formulation. With this general model, we find that $\hat{\Psi}_k^2$ =1.8544, which is not in the rejection region of χ_4^2, and so are all its 4 components. This implies that a time varying conditional heteroskedasticty component together with the MA(1) conditional mean model with Student's t density for the error term provides an acceptable model.

We also tried higher orders beyond $k = 4$ but the marginal impact was negligible in the final model. Therefore, we believe $k = 4$ is sufficient for the data on hand.(we also applied data-driven smooth test methods proposed by Ledwina (1994), and in most cases k was between 2 and 4). We chose t distribution with 8 degrees of freedom, since that was the closest integer value that maximizes the likelihood functions. We should mention that, although we have chosen to divide our sample into 8431 and 2016 observations, this is not necessarily an optimal split. We used a 4:1 split as a rule of thumb as this was an acceptable choice using cross-validation type methods. In fact, we have seen that the actual size of the test goes up on average as we increase the size of the test sample keeping the estimation sample fixed. Diebold et al. (1998) used 4133 and 4298 split, and we suspect that this sample splitting will have very large implied size. In a previous version of this paper we kept the estimation sample 4133 (with a test sample size of 1000) so as to compare the results obtained by Diebold et al. and our formal test procedure. Our current results turned out to be quite similar to those of the previous ones, with some differences, particularly in the significance of the fourth order Legendre polynomial.

From Table 5, overall, we can conclude that there is no evidence to suggest that the forecasted model MA(1)-t-GARCH(1,1) fails to predict the density of the future realizations of S&P 500 returns. We can also see from the results based on the EDF that there is more of unaccounted volatility than other departures. Looking at the $\hat{u}_2^2$ and $\hat{u}_4^2$ components we can say that, introduction of conditional heteroskedasticty improved the model by reducing the "butterfly" pattern in the PIT histogram (or the ratio density function). It is not clear from pure visual inspection of Figures 4 and 5 that a non-Gaussian error term should be incorporated in the model [see Diebold et al. (1998)]. However, application of the smooth test indicated a better fit for the model with the errors following a Student's t distribution where $\hat{u}_2^2$ component reduced from highly significant 12.137 to statistically insignificant 0.3445 (see Table 5). Although the smooth test did not directly address whether there was dependence in the data, it did pick up the effect of this unaccounted dependence in the data incorporating conditional heteroskedasticity.

One possible interpretation of the apparent failure of the normal GARCH(1,1) could be the possibility of a hidden Markov type model that Weigend and Shi (2000) discussed in evaluating the density of daily returns of S&P 500 index. They assumed one of several "states" or "experts" generates the true observation in certain financial time series data, like S&P 500 returns, where the signal to noise ratio is pretty small and the discrete number of states jump from one to the other with a time-varying or time invariant transition probability matrix. They reported that their model performed slightly better than normal GARCH(1,1) model. In fact, they worked under a more restrictive Gaussian framework although a more general exponential family distribution would have been more appropriate.

Our results from the smooth test indicate that part of the reason for the strong significance of the fourth order orthogonal polynomial in our naive models, a term connected to the kurtosis of the distribution of the PIT, is a deviation in the second and fourth moments. This also indicates leptokurtic nature of the original data. We should, however, note that since both the second and the fourth order terms are present in the normalized Legendre polynomial $\pi_4(y)$, it is not possible to exactly separate out these two effects.

5 Monte Carlo Evidence

Figure 6 shows the distribution of the $\hat{\Psi}_4^2$ statistic under the null hypothesis of correct specification of the model, t-GARCH(1,1), with the χ_4^2 distributions for samples of size 1000. We also inspect the plots (presented in Figure 7) of the components to check whether the individual u_i^2 asymptotically follow the χ_1^2 distribution.

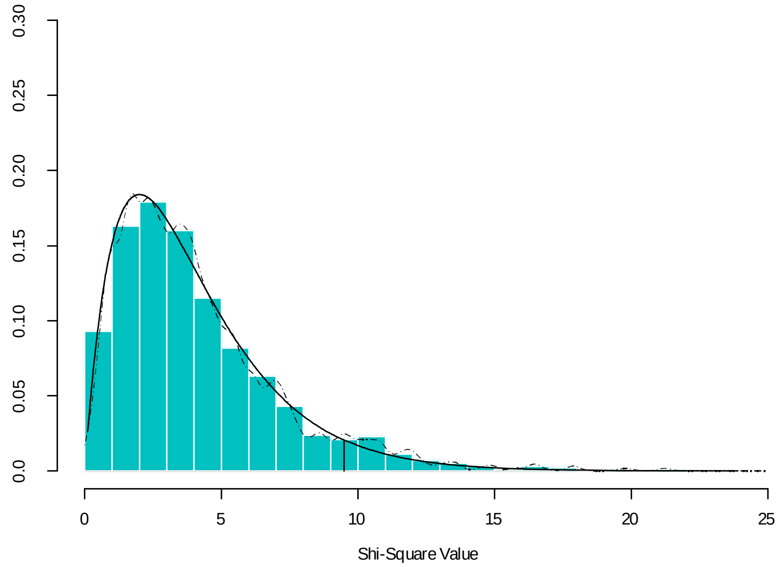

Figure 6: Histogram and distribution of Ψ_4^2 under the null hypothesis

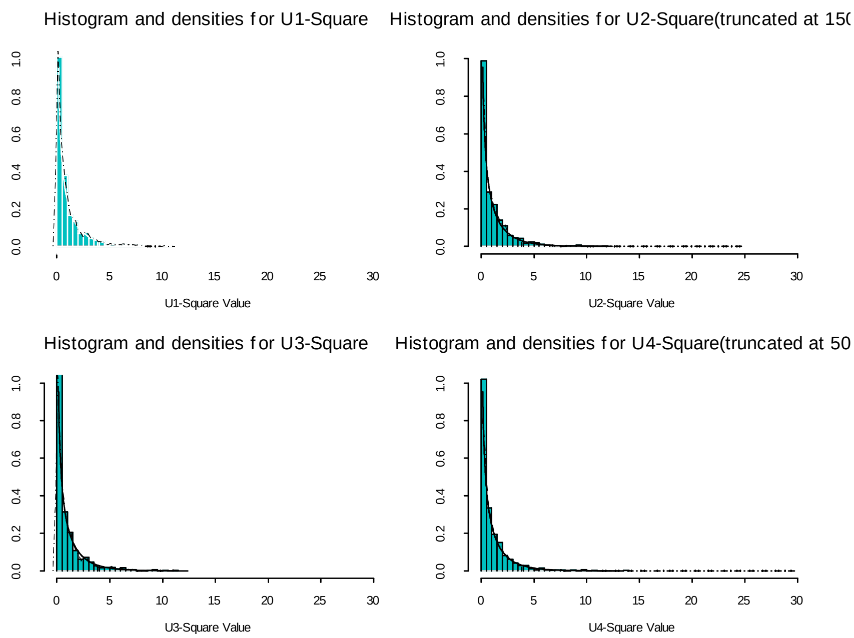

Figure 7: Distribution of individual u^2

However, since we are using estimated parameters in place of the true parameters of the distribution, we must estimate the distribution with sufficient accuracy in order to do evaluate the performance of forecasts. We generated a sample of size 2500 from a $t_7 - GARCH(1,1)$ distribution:

$$y_t = \sqrt{\frac{5h_t}{7}} t_7$$
$$h_t = 0.2 + 0.15 y_{t-1}^2 + 0.65 h_{t-1}. \quad (26)$$

After estimating the parameters of the sample with the first 2000 observations ($m = 2000$) we freeze it and generate the density forecast for the last 500 observations ($n = 500$). Hence we obtain the probability integral transform of the latter 500 observations using the estimated PDF. We performed the modified smooth test on the forecasted sample and replicated it to get the size properties of this test. Our results, though not reported here but available upon request, show that even with estimated parameters the Ψ_4^2 statistic seem to follow a central χ^2 distribution with 4 degrees of freedom, and also, the individual component u_i^2 seem to follow the χ_1^2 distribution under the correct specification of the model.

One of the very important questions that left to be answered is what should be the sample split in order to estimate the parameters to a fair degree of accuracy so that the modified smooth test is consistent and an empirical level of significance close to the nominal size. We kept the initial estimation sample size $m = 2000$ fixed and considered several testing sample sizes (n). The actual sizes for different values of n with 200 replications are plotted in Figure 8 when the nominal level is 5%. We note that with n, the empirical size tends to go up, and after the value of $n = 500$, the size goes up considerably (with m being fixed at 2000). Therefore, for our smooth test on S&P 500 returns with $m = 8431$, we chose the maximum 4:1 split of the sample size, i.e., selected the test sample size $n = 2016$, close to $m/4$.

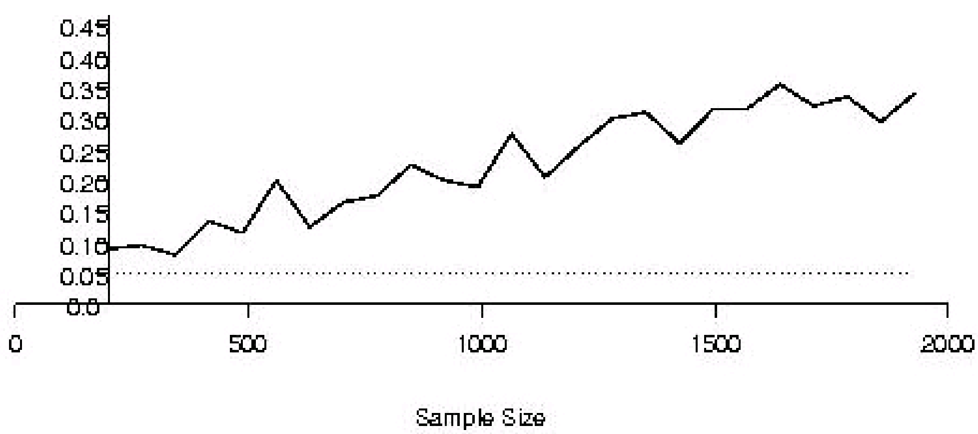

Figure 8. Plot of the size of the test as a function of n ($m = 2000$)

For small sample sizes we can use cross validation based method to decide on the sample split. Since, our main objective is to minimize size distortion in finite or small samples we can select the sample size that minimizes the distance from the distribution under H_0 or in other words, minimizes distance between the density of PIT and the uniform distribution. We should admit that where the exact sample split should occur is not a easy problem to solve analytically and this investigation is part of our ongoing research.

6 Conclusion and Further Research

One of the main problems in the area of market risk management has been the evaluation of the probability density forecasts. Using Neyman's (1937) smooth test procedure we suggest an easily implementable formal test to achieve that. When a forecast probability density is rejected, this procedure can identify the specific source(s) of rejection. Our approach is illustrated with an application to S&P 500 returns. Our test can also be used in areas of macroeconomics such as evaluating the density forecasts of realized inflation rates. Diebold, Tay and Wallis (1999) used a graphical technique for the density forecasts of inflation from the *Survey of Professional Forecasters.*

Neyman's smooth test can also be extended to a multivariate setup of dimension N for m time periods, by taking a combination of Nm sequences of univariate densities as discussed by Diebold, Hahn and Tay (1999). This could be particularly useful in fields like financial risk management to evaluate densities for high-frequency financial data like stock or derivative (options) prices and foreign exchange rates. While our smooth test using estimated parameters provides specific directions for the alternative models based on the data on S&P 500 returns, it should be bourne in mind that originally the smooth test was not designed for dependent data. In our empirical applications to stock returns, we have tried to capture dependence through conditional heteroskedasticty. It will be more interesting to incorporate the dependence structure directly into the density function. Currently, we have work-in-progress along that direction. Since the Smooth test is essentially a score test, it enjoys certain optimal properties, and also, we do not need to estimate the parameters under the alternative hypothesis. The latter benefit makes it conducive to models with a large number of parameters, particularly when we want to incorporate complicated dependence structures.

Acknowledgement 1 *We would like to thank the participants at various confer-*

ences, particularly the High Frequency Data and Forecasting Conference jointly hosted by Singapore Management University and IMS-NUS, Singapore, 2004, for helpful comments. We are also thankful to Yongmiao Hong, Yiu Kuen Tse, Ken Wallis, and Zhijie Xiao for helpful discussions and suggestions. The first author would like to thank Singapore Management University-Wharton School Office of Research, Research Grant Number. 03-C208-SMU-025 for the financial support and the Centre for Academic Computing at SMU for research assistance.

Appendices

Appendix A (Proof of Theorem 2)

Proof. From equations (7), (24) and (25)

$$\hat{\Psi}_k^2 - \Psi_k^2 = \sum_{j=1}^{k} \frac{1}{n} \left[\left(\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \hat{\beta} \right) \right) \right)^2 - \left(\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \beta \right) \right) \right)^2 \right]$$

$$= \sum_{j=1}^{k} \left[\hat{u}_j^2 - u_j^2 \right]. \tag{27}$$

Now applying the Mean Value Theorem, we get

$$\hat{u}_j^2 = \frac{1}{n} \left[\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \hat{\beta} \right) \right) \right]^2$$

$$= \frac{1}{n} \left[\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \beta \right) \right) \right]^2 + \frac{1}{n} \left(\hat{\beta} - \beta \right) \frac{d}{d\beta} \left[\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \beta \right) \right) \right]^2 \Bigg|_{\beta=\beta^*}$$

where β^* is such that $\left| \hat{\beta} - \beta \right| \geq \left| \beta^* - \beta \right|$.

$$\text{Hence, } \hat{u}_j^2 - u_j^2 = \frac{2}{n} \left(\hat{\beta} - \beta \right) \left[\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \beta^* \right) \right) \right] \left[\sum_{i=1}^{n} \frac{d\pi_j \left(F \left(x_i; \beta^* \right) \right)}{d\beta} \right]$$

$$= 2 \left(\frac{n}{\sqrt{m}} \right) \left[\sqrt{m} \left(\hat{\beta} - \beta \right) \right] \left[\frac{1}{n} \sum_{i=1}^{n} \pi_j \left(F \left(x_i; \beta^* \right) \right) \right] \times \left[\frac{1}{n} \sum_{i=1}^{n} \frac{d\pi_j \left(F \left(x_i; \beta^* \right) \right)}{d\beta} \right]. \tag{28}$$

Furthermore, we know that under $H_0 : y_i = F(x_i; \beta)$ is distributed as $U(0, 1)$ for

$i = 1, 2, ..., n$. Hence, using orthogonality of $\pi_j(.)$ under H_0 for $j = 1, 2, ..., k$,

$$E(\pi_j(y_i)) = \int_0^1 \pi_j(u)\,du = 0. \tag{29}$$

Applying the WLLN (Khinchine's theorem, Rao(1973 p. 112) we have as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n} \pi_j(F(x_i;\beta)) \xrightarrow{p} E(\pi_j(y_i)) = 0. \tag{30}$$

For arbitrary but fixed m, β^* is fixed. For $i = 1, 2, ..., n$, $F(x_i;\beta^*)$ is a (an absolutely) continuous function of x_i. Hence, if $X_1, X_2, ..., X_n$ are *IID* random variables having a CDF $F(x;\beta)$ then, $y_i^* = F(x_i;\beta^*)$, $i = 1, 2, ..., n$ are also *IID* with a density function (called the ratio density function or RDF)

$$h(y) = \frac{f(x;\beta)}{f(x;\beta^*)} = \frac{f(F^{-1}(y;\beta);\beta)}{f(F^{-1}(y;\beta^*);\beta^*)}.$$

Hence, $y_1, y_2, .., y_n$ are *IID* random variables with a density function $h(y)$ and has a finite first moment. Using the WLLN, for $j = 1, 2, ..., k$,

$$\frac{1}{n}\sum_{i=1}^{n} \pi_j(F(x_i;\beta^*)) \xrightarrow{p} E[\pi_j(F(x_i;\beta^*))]. \tag{31}$$

Now, we have $\hat{\beta} \xrightarrow{p} \beta$ as $\hat{\beta}$ is a $\sqrt{m}-$consistent estimator of β. Since, $\left|\hat{\beta} - \beta\right| \geq |\beta^* - \beta|$, β^* is also converges to β in probability. If $\pi_j(F(x;\beta))$ is a continuous function of β at $\beta = \beta^*$, we have

$$E[\pi_j(F(x;\beta^*))] \xrightarrow{p} E[\pi_j(F(x;\beta))],\ j = 1, 2, ..., k. \tag{32}$$

Hence, as m and n go to infinity, using results in (29), (30), (31) and (32), we have

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{i=1}^{n} \pi_j(F(x_i;\beta^*)) \xrightarrow{p} E[\pi_j(F(x_i;\beta^*))] \xrightarrow{p} E[\pi_j(F(x;\beta))] = 0, \\ &\text{i.e.,} \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \pi_j(F(x_i;\beta^*)) = a_1 = o_p(1). \end{aligned} \tag{33}$$

We should note that this result holds only under H_0, otherwise we will only have

$\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right)=O_p\left(1\right)$. Applying the WLLN again, for sufficiently large m,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta} &\xrightarrow{p} E\left[\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \\ &\xrightarrow{p} E\left[\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right]<\infty \\ \Rightarrow \frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta} &= a_2 = O_p\left(1\right). \end{aligned} \tag{34}$$

By assumption $E\left[\frac{d\pi_j(F(x_i;\beta))}{d\beta}\right]<\infty$, hence $\frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j(F(x_i;\beta^*))}{d\beta}=O_p\left(1\right)$. Since, $\hat{\beta}$ is a $\sqrt{m}-$consistent estimator,

$$\sqrt{m}\left(\hat{\beta}-\beta\right)=a_3=O_p\left(1\right). \tag{35}$$

Hence from equation (28) using the results in (33), (34) and (35), we obtain

$$\begin{aligned}\hat{u}_j^2-u_j^2 &= 2\left(\frac{n}{\sqrt{m}}\right)\left[\sqrt{m}\left(\hat{\beta}-\beta\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \\ &\times\left[\frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \\ &= 2\frac{n}{\sqrt{m}}a_1a_2a_3 \\ &= \frac{n}{\sqrt{m}}o_p\left(1\right). \end{aligned} \tag{36}$$

From (27) using (36) for fixed k,

$$\hat{\Psi}_k^2-\Psi_k^2=\frac{n}{\sqrt{m}}o_p\left(1\right). \tag{37}$$

which proves Theorem 2. ■

Appendix B (Data and Computational Methods)

The data used for the empirical analysis was the returns on value-weighted S&P 500 Composite Index with dividends. The data for the time period July 3, 1962 till December 31, 2003 was extracted from CRSP Daily Returns database from WRDS. This is the most current data available at the time of completion of this paper.

We have estimated the following MA(1)-GARCH (1,1) model

$$\begin{aligned} y_t &= \beta_0 + \beta_1 \varepsilon_{t-1} + \varepsilon_t \\ V\left(\varepsilon_t | \Omega_t\right) &= h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 h_{t-1}, \end{aligned} \tag{38}$$

with conditionally Gaussian or Student's t error distribution. For our Monte Carlo experiments, we set the initial values at $\varepsilon_0 = 0$ and $h_0 = 1$ as defaults. From (38)

$$\begin{aligned} &\varepsilon_1 = y_1 - \beta_0 - \beta_1 \varepsilon_0 = \tilde{y}_1 \\ &\quad \Rightarrow \varepsilon_2 = y_2 - \beta_0 - \beta_1 \varepsilon_1 = y_2 - \beta_0 - \beta_1 \tilde{y}_1 \\ &\quad \Rightarrow \varepsilon_3 = y_3 - \beta_0 - \beta_1 \left(y_2 - \beta_0 - \beta_1 \tilde{y}_1\right) = y_3 - \beta_0 \left(1 - \beta_1\right) - \beta_1 y_2 + \beta_1^2 \tilde{y}_1. \end{aligned}$$

For a general t,

$$\begin{aligned} \varepsilon_t &= y_t - \beta_0 \left(1 - \beta_1 + \beta_1^2 + ... + (-1)^{t-2} \beta_1^{t-2}\right) - \beta_1 y_{t-1} + \beta_1^2 y_{t-2} + ... + (-1)^{t-1} \beta_1^{t-1} \tilde{y}_1 \\ &= y_t - \beta_0 \frac{1 - (-\beta_1)^{t-1}}{1 - (-\beta_1)} - \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + (-1)^{t-1} \beta_1^{t-1} \tilde{y}_1. \end{aligned} \tag{39}$$

In matrix notation, this can be written as

$$\epsilon = \mathbf{B}\mathbf{y} + \mathbf{a}$$

where $\epsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_t)^\top$, $\mathbf{y} = (\tilde{y}_1, y_2, y_3, ..., y_t)^\top$,

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ (-\beta_1) & 1 & 0 & \cdots & 0 \\ (-\beta_1)^2 & (-\beta_1) & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (-\beta_1)^{t-1} & (-\beta_1)^{t-2} & (-\beta_1)^{t-3} & \cdots & 1 \end{bmatrix} = (-\beta_1)^{\mathbf{P}},$$

$$\text{with } \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ \\ t & t-1 & t-2 & 1 \end{bmatrix}, \text{ and } \mathbf{a} = -\frac{\beta_0}{1+\beta_1} \begin{bmatrix} 1 \\ 1+\beta_1 \\ (1-\beta_1^2) \\ \\ \left(1 - (-\beta_1)^{t-1}\right) \end{bmatrix}. \tag{40}$$

The elements of $\mathbf{P}$ are obtained as p_{ij}, with $p_{ij} = \max(i-j, 0) - \min(\max(i-j, 0), 1)$.

To simplify h_t, we write

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 h_{t-1} = k_{t-1} + \alpha_2 h_{t-1}, \text{ where } k_t = \alpha_0 + \alpha_1 y_{t-1}^2.$$

with default initial value of $h_0 = 1$ or $k_0 = \alpha_0 + \alpha_1 h_0 = \tilde{k}_0$, we can write

$$\begin{aligned} &h_1 = k_0 + \alpha_2 h_0 \Rightarrow h_2 = k_1 + \alpha_2 k_0 + \alpha_2^2 h_0 = k_1 + \alpha_2 \tilde{k}_0, \\ &\text{i.e., in general, } h_t = k_{t-1} + \alpha_2 k_{t-2} + ... + \alpha_2^{t-1} \tilde{k}_0. \end{aligned}$$

In matrix notation

$$\mathbf{h} = \mathbf{A}\mathbf{k},$$

where $\mathbf{h} = (h_1, h_2, ..., h_t)^\top$, $\mathbf{k} = \left(\tilde{k}_0, k_1, ..., k_{t-1}\right)^\top$ and the lower triangular matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ (\alpha_2) & 1 & 0 & \cdots & 0 \\ (\alpha_2)^2 & (\alpha_2) & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (\alpha_2)^{t-1} & (\alpha_2)^{t-2} & (\alpha_2)^{t-3} & \cdots & 1 \end{bmatrix} = \alpha_2^{\mathbf{P}},$$

where $\alpha_2^{\mathbf{P}}$ is a matrix where the components are powers, i.e., $\alpha_2^{p_{ij}}$. We use this to evaluate the series of conditional means and standard deviations using BHHH type Algorithm. We used S+Finmetrics to obtain the parameter estimates for the GARCH model with conditional Gaussian and Student's t distributions. The above matrix notations were used to vectorize the calculations. Given the estimates $\hat{\beta}_1, \hat{\alpha}_0, \hat{\alpha}_1$ and $\hat{\alpha}_2$, we calculated the estimated probability integral transforms as

$$\sqrt{\frac{df}{(df-2)\,h_t}} \left(y_t - \hat{\beta}_1 \varepsilon_{t-1}\right) \sim t_{df},$$

where df is the degrees of freedom for the conditional Student's t distribution. We use a similar algorithm for the conditionally Gaussian distribution $\left(y_t - \hat{\beta}_1 \varepsilon_{t-1}\right)/h_t$ having a standard normal distribution. We can also estimate the MA(1) model with the above procedure by selecting $\alpha_1 = \alpha_2 = 0$ in equation (38).

References

[1] Anderson, G. (1994), "Simple tests of distributional form," *Journal of Econometrics*, 62, 265-276.

[2] Bera, A. K., and Bilias, Y. (2001a), "Rao's score, Neyman's C(α) and Silvey's LM test: An essay on historical developments and some new results," *Journal of Statistical Planning and Inference,* 97, 9-44.

[3] ——— (2001b), "On some optimality properties of Fisher-Rao score function in testing and estimation," *Communications in Statistics, Theory and Methods*, 30, 1533-1559.

[4] Bera, A. K., and Ghosh, A. (2001), "Neyman's smooth test and its applications in Econometrics," in: *Handbook of Applied Econometrics and Statistical Inference*, Eds. A. Ullah, A. Wan and A. Chaturvedi, New York: Marcel Dekker, pp. 177-230.

[5] Bera, A. K., Ghosh, A., and Xiao, Z. (2004), "A Smooth Test of Comparing Equality of Two Densities," Working Paper, Singapore Management University.

[6] Berkowitz, J. (2001), "Testing Density Forecasts, With Applications to Risk Management," *Journal of Business & Economic Statistics*, 19: 4, 465-474.

[7] Boero, G., Smith, J., and Wallis, K. F. (2004), "Decomposition of Pearson's chi-squared test," *Journal of Econometrics*, forthcoming.

[8] Boulerice, B., and Ducharme, G. R. (1995), "A note on smooth tests of goodness of fit for location-scale families," *Biometrika*, 82, 437-438.

[9] Christoffersen, P. F. (1998), "Evaluating interval forecasts," *International Economic Review*, 39, 841-862.

[10] Crnkovic, C., and Drachman, J. (1997), "Quality control," in: *VAR: Understanding and Applying Value-at-Risk*, London: Risk Publication.

[11] D'Agostino, R. B., and Stephens, M. A. (1986), *Goodness-of-Fit Techniques.* New York: Marcel Dekker.

[12] Diebold, F. X., Gunther, T. A., and Tay, A. S.(1998), "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, 39, 863-883.

[13] Diebold, F. X., Hahn, J., and Tay, A. S. (1999), "Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns in foreign exchange," *Review of Economics and Statistics*, 81, 661-673..

[14] Diebold, F. X., and Lopez, J. A..(1996), "Forecast evaluation and combination," in: *Handbook of Statistics*, Vol. 14, eds. G.S. Maddala and C.R. Rao, Amsterdam: North-Holland, pp. 241-268.

[15] Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

[16] Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999), "Evaluating density forecasts of inflation: the survey of professional forecasters," in *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive W. Granger*, eds. R.F Engle and H. White, New York: Oxford University Press, pp. 76-90.

[17] Garratt, A., Lee, K., Pesaran, M. H., and Shin, Y. (2003), "Forecast Uncertainties in Macroeconomic Modeling: An Application to the UK Economy," *Journal of the American Statistical Association*, 98: 464, 829-837.

[18] Ghosh, A. (2003), "*Smooth Test and Its Applications in Economics and Finance*," Unpublished Ph.D. Dissertation, Department of Economics, University of Illinois.

[19] Giacomini, R. (2002), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests: Asymptotic and Boostrap Methods," *mimeo*, University of California, San Diego.

[20] Granger, C. W. J., and Pesaran, M.H. (1996), "A Decision Theoretic Approach to Forecast Evaluation," in *Statistics and Finance: An Interface*, eds. W. S. Chan, W. K. Li and H. Tong, Imperial College Press, London, Chapter 15, pp. 261-278.

[21] ——— (2000), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19, 537-560.

[22] Hong, Y. (2001), "Evaluation of out-of-sample probability density forecasts with applications to stock prices," *mimeo*, Department of Economics, Cornell University.

[23] Kopecky, K. J., and Pierce, D. A.(1979), "Efficiency of smooth goodness-of-fit tests," *Journal of the American Statistical Association*, 74, 393-397.

[24] Li, Q. (1996), "Nonparametric testing of closeness between two unknown distribution functions," *Econometric Reviews*, 15, 261-274.

[25] Ledwina, T. (1994), "Data-driven version of Neyman's smooth test of fit," *Journal of the American Statistical Association*, 89, 1000-1005.

[26] Neyman, J. (1937), "“Smooth test” for goodness of fit," *Skandinaviske Aktuarietidskrift*, 20, 150-199.

[27] ——— (1959), "Optimal asymptotic test of composite statistical hypothesis," in *Probability and Statistics: the Harold Cramér Volume*, ed. U. Grenander, Uppsala: Almqvist and Wiksell, pp. 213-234.

[28] Pagan, A., and Ullah, A. (1999), *Nonparametric Econometrics*. Cambridge, UK: Cambridge University Press..

[29] 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 reasonably be supposed to have arisen from random sampling," *Philosophical Magazine* 5^{th} Series, 50, 157-175.

[30] Rao, C.R. (1948), "Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation," *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.

[31] ——— (1973), *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons.

[32] Rayner, J. C. W., and Best, D. J. (1989), *Smooth Tests of Goodness of Fit*. New York: Oxford University Press.

[33] Rosenblatt, M..(1952), "Remarks on a multivariate transformation," *Annals of Mathematical Statistics*, 23, 470-472.

[34] Sarno, L., and Valente, G. (2004), "Comparing the Accuracy of Density Forecasts from Competing Models," *Journal of Forecasting*, 23, 541-557.

[35] Tay, A. S., and Wallis, K. F. (2000), "Density forecasting: A Survey," *Journal of Forecasting*, 19, 235-254..

[36] Thomas, D. R., and Pierce, D. A. (1979), "Neyman's smooth goodness-of-fit test when the hypothesis is composite," *Journal of the American Statistical Association*, 74, 441-445.

[37] Vuong, Q. H. (1989), "Likelihood Ratio Tests for Model Selection and Non-nested Hypothesis," *Econometrica*, 57, 307-333.

[38] Wallis, K. F. (2003), "Chi-Squared tests of interval and density forecasts, and the Bank of England's fan charts," *International Journal of Forecasting*, 19, 165-175.

[39] ——— (2004), "An Assessment of Bank of England and National Institute Inflation Forecast Uncertainties," *mimeo*, University of Warwick.

[40] Weigend, A., and Shi., S. (2000), "Predicting Daily Probability Distributions of S&P500 Returns," *Journal of Forecasting*, 19, 375-392.