A Formal test of Density Forecast Evaluation.[1]

Anil K. Bera
Department of Economics,
University of Illinois
at Urbana-Champaign.
e-mail: abera@uiuc.edu
Phone: 217-333-4596

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

October 1, 2003.

[1]This is a preliminary version of the paper, please quote results of teh paper with the authors' permission.

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 (PDF) of various market variables. One of the main problems in this area has been evaluation of the density forecasts. In this paper, we propose a formal test for density forecast evaluation using Neyman (1937) smooth test procedure. Apart from giving indication of acceptance or rejection of the tested model, this approach provides specific sources (such as the mean, variance, skewness and kurtosis or the location, scale and shape of the distribution) of rejection, thereby helping in deciding possible modifications of the assumed model. Our applications to value weighted S&P 500 returns indicated that introduction of a conditional heteroscedasticity model significantly improved the model over a model with constant conditional variance.

1 Intoduction

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 (Christoffersen, 1998), and then finally to density forecast, though construction of density forecast in empirical work is only a recent phenomenon. Therefore, it is not surprising that evaluating density forecast techniques is very much in its infancy. The only two published papers, we are aware of, that directly address the question of evaluation of density forecasts are Diebold, Gunther and Tay (1998) and Berkowitz (2001). 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."

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 and showed that PIT is distributed as $U(0,1)$ irrespective of the specification 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 the next section we review Neyman (1937) approach, though a fuller account is given in Bera and Ghosh (2001). Section 3 uses the framework of Diebold et al. (1998) and considers a density forecast evaluation. An application to S. & P. 500 return data is given in Section 4. Section 5 gives some limited Monte Carlo results on simulated data to see 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. 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).[1] Neyman suggested this test to rectify some of the drawbacks of Pearson (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

[1] In the context of testing several different hypotheses, Neyman (1937 [27], p. 160) argued this quite eloquently as follows :

> "If we treat all these hypotheses separately, we should define the set of alternatives for each of them and this would in practice lead to a dissection of a unique problem of a test for goodness of fit into a series of more or less disconnected problems.
>
> However, this difficulty can be easily avoided by substituting for any particular form of the hypothesis H_0, that may be presented for test, another hypothesis, say h_0, which is equivalent to H_0 and which has always the same analytical form. The word equivalent, as used here, means that whenever H_0 is true, h_0 must be true also and inversely, if H_0 is not correct then h_0 must be false."

as [see Neyman (1937, p. 161), Pearson (1938, p. 138) and 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 in that sense that $h(y) \geq 0$, $y \in (0,1)$ and $\int_0^1 h(y)\, dy = 1$, if we assume that F and G are also strictly increasing functions. We will call it the *ratio density function* (RDF) since it is both a ratio of two densities and a proper 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 $c(\theta)$ is the constant of integration depending only on $(\theta_1, ..., \theta_k)$, $\pi_j(y)$s are orthonormal polynomials of order j satisfying

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

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 see the exact form 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. Looking at some commonly used non-normal densities as alternatives we can observe that $f(.) \equiv \mathcal{N}(0,1)$ against $g(.) \equiv$Central χ^2_4 yields $\ln(h(z)) = \frac{1}{2}z^2 - \frac{1}{2}z + \ln z + \ln\left(\frac{\sqrt{2\pi}}{4}\right)$. If we have $f(.) \equiv \mathcal{N}(0,1)$ and $g(.) \equiv$Central t_4, the RDF is given from $\ln(h(z)) = \frac{z^2}{2} + \frac{5}{2}\ln\left[1 + \frac{z^2}{4}\right] + \ln\left(\frac{\sqrt{2\pi}}{2}\right)$. The above examples suggests that departures from the null hypothesis could be tested using an appropriate function (or functions) estimating the RDF. From looking at the plots of the different ordered normalized Legendre polynomials, we believe that the test will not only be powerful but also will be informative on identifying particular source(s) of departure(s) from H_0 (see Ghosh, 2001)

Using the multiparameter version of the generalized Neyman-Pearson (N-P) 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 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

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

which under H_0 asymptotically follows a central 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 simply be obtained using Rao (1948) score (RS) test principle. Taking (4) as the PDF under the alternative hypothesis, the log-likelihood function $l(\theta)$ can be written as

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

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

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

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

$$\begin{aligned} s\left(\theta_j\right) &= \frac{\partial l(\theta)}{\partial \theta_j} \\ &= n \frac{\partial \ln c(\theta)}{\partial \theta_j} + \sqrt{n} u_j, \;\; j = 1, 2, ..., k. \end{aligned} \tag{10}$$

Using (4), $\int_0^1 h(z)\, dz = 1$. Differentiating this identity with respect to θ_j

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

Evaluating this under $\theta = 0$, we have $\left.\frac{\partial \ln c(\theta)}{\partial \theta_j}\right|_{\theta=0} = 0$ and under the null

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

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 $-n\partial^2 c(\theta) / \partial \theta_j \partial_l$ evaluated at $\theta = 0$. Differentiating (11) with respect to θ_l and evaluating it at $\theta = 0$, 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 orthogonality conditions in (5)

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

and

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

where I_k is a $k \times k$ identity matrix. Combining (9), (13) and (16) the RS test statistic has the simple form

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

Neyman's approach requires the computation of the probability integral transform (1) in terms of Y. 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 alternative density in terms of X 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 the test statistic Ψ_k^2 reduces to

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

which has the same asymptotic distribution as before. 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 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? Since this is time series data, at time t we know all the past values of x_t up to time t or the *information set* at time t, namely, $\Omega_t = \{x_{t-1}, x_{t-2}, ...\}$. 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 how much 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

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

Using equations (21), the density function of the transformed variable y_t is given by

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

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

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 KS or the CvM or a test using the Kuiper statistics (see Crnkovic and Drachman 1997, p. 48). Any of those 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*

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

are IID $U(0,1)$.

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

$$\begin{aligned} P\left(Y_i \leq y_i, i=1,2,...,n\right) &= \int_0^{y_1}\int_0^{y_2}...\int_0^{y_n} f\left(x_1\right)dx_1 f\left(x_2|x_1\right)dx_2...f\left(x_n|x_1,...,x_{n-1}\right)dx_n \\ &= \int_0^{y_1}\int_0^{y_2}...\int_0^{y_n} dt_1dt_2...dt_n \\ &= y_1y_2...y_n. \end{aligned} \tag{23}$$

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

Diebold, Gunther and Tay (1998) used Theorem 1, and tested $H_0 : g_t(.) = f_t(.)$ by checking whether the probability integral transform y_t in (21) follows *IID* $U(0,1)$. They employed a graphical (visual) 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 $U(0,1)$ 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 reviews on density forecasting and forecast evaluation methods, see Tay and Wallis (2000) and Diebold and Lopez (1996). 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 (1937) provides an *analytic* tool to determine the structure of the density under the alternative hypothesis using orthonormal polynomials (normalized Legendre polynomials) $\pi_j(y)$ defined in (5).[2] While, on one hand, the smooth test provides a basis for a classical goodness-of-fit test (based on the generalized N-P lemma), on the other hand, it can also be used to determine the sensitivity of the power of the test to departures from the null

[2]Neyman (1937) used $\pi_j(y)'s$ as the orthogonal polynomials which can be obtained by using the following conditions,

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

given the restrictions of orthogonality given in (5). Solving these the first five $\pi_j(y)$ are (Neyman 1937, pp. 163-164)

$\pi_0(y) = 1,$
$\pi_1(y) = \sqrt{12}\left(y - \frac{1}{2}\right),$
$\pi_2(y) = \sqrt{5}\left(6\left(y - \frac{1}{2}\right)^2 - \frac{1}{2}\right),$
$\pi_3(y) = \sqrt{7}\left(20\left(y - \frac{1}{2}\right)^3 - 3\left(y - \frac{1}{2}\right)\right),$
$\pi_4(y) = 210\left(y - \frac{1}{2}\right)^4 - 45\left(y - \frac{1}{2}\right)^2 + \frac{9}{8}$

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 $\frac{1}{n}\left(\sum_{i=1}^{n}\pi_j\left(y_i\right)\right)^2, j=1,...,k,$ which are nothing but the squares of the efficient score functions. Using Rao and Poti (1946), 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, 2001b)]. 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 where we cannot assume that the true data generating process is known under the null hypothesis $(H_0 : g_t(.) = f_t(.))$. Hence, we have to estimate the PDF generating the data using an estimation sample. Then, we use those estimated CDFs to evaluate the PIT for constructing the test statistic for the smooth test. Let us first assume that we know a general functional form of the distribution function $F(.;\beta)$ generating the data but have to estimate the parameter β based on the estimation sample of size m. As we mentioned before our test is done on a sample of size n. The true test statistic for the smooth test under true model is given in (7) where

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

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

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

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 $\sqrt{n}-$consistent estimator of β and $\hat{u}_j^2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_j\left(\hat{y}_i\right)$. 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 *If the sample size for the estimation sample is m and that for the testing sample is n in a split sample, $\hat{\beta}$ is a $\sqrt{m}-$consistent estimator of the parameter β and given*

$$E\left[\frac{d\pi_j\left(F\left(x_i;\beta\right)\right)}{d\beta}\right] < \infty \tag{26}$$

using equations (7) and (25) under the null hypothesis H_0

$$\hat{\Psi}_k^2 - \Psi_k^2 = o_p(1)$$

if $n = O\left(m^{\frac{1}{2}}\right)$.

Proof. See Appendix A. ■

4 Application to asset returns on S&P 500

Diebold, Gunther and Tay (1998) used the daily data on 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. The sample is split into in-sample and out-of-sample periods for model estimation and density forecast evaluation. There are 4133 in-sample observations (07/03/62-12/29/78) and 4298 out-of-sample observations (01/02/79-12/29/95). However, in order to obtain a test with desirable size using the the smooth test principle, we have to modify the sample size of the evaluation sample, keeping the sample size of the estimation sample the same. Our density evaluation sample is of size 1000, while the estimation sample size is 4133. The following figure (Fig I) compares the density estimates between the in-sample and the out-of-sample data.

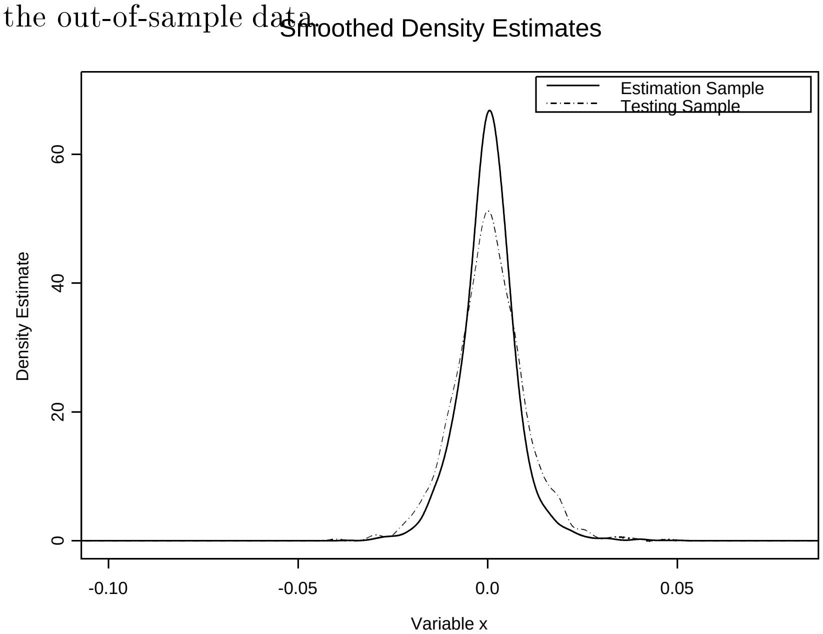

Figure 1. Kernel density estimators for estimation and testing samples

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. In the following figure (Figure 2), we used a non-parametrically smoothed EDF to generate the PIT of the evaluation data. From a visual analysis of the plot 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 [see Bera, Ghosh and Xiao (2003)]. In order to better fit the model for forecasting future observations, we use a MA(1), MA(1)- normal GARCH(1,1) and finally, a MA(1)-t-GARCH(1,1) model to the estimation sample obtained through AIC and BIC criteria. The histograms of the PITs and the autocorrelation functions (correlograms) thus generated are given in the following figures. From visual analysis of these plots we can infer that introducing a time varying conditional heteroscedasticity term clearly improves the forecast as reflected in the ACF plots and it also causes the histogram of the PITs to be closer to a $U(0,1)$ PDF. Introducing a non-Gaussian error term marginally improves the histogram of the PIT by reducing the "butterfly" pattern to some extent.

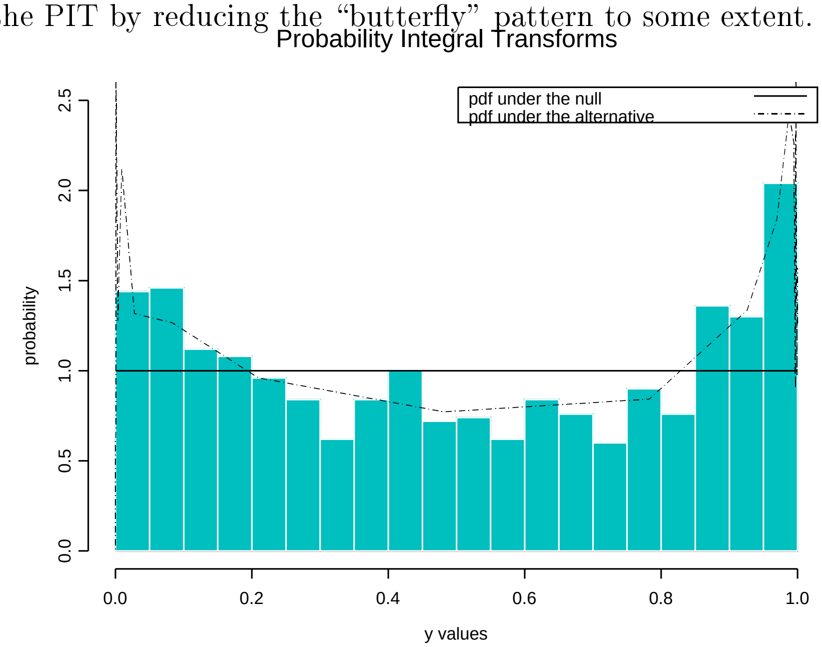

Figure 2. PIT histogram with density ratio using EDF

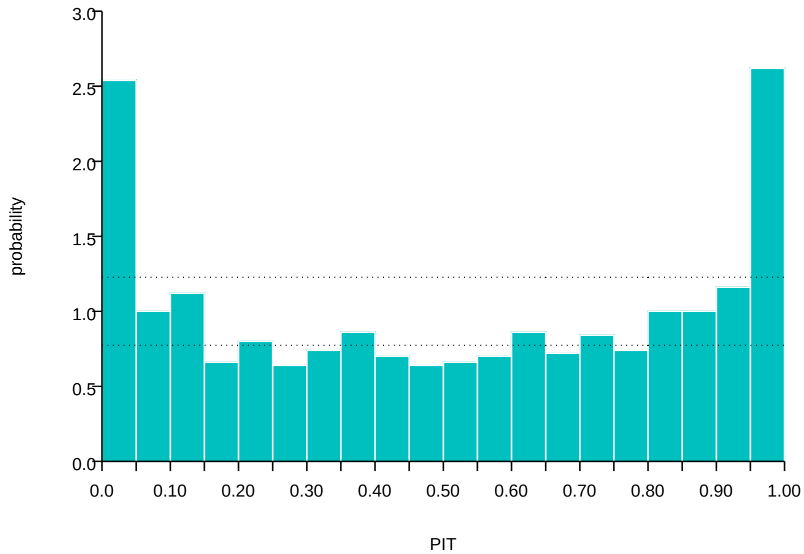

Figure 3. PIT histogram MA(1) with normal errors

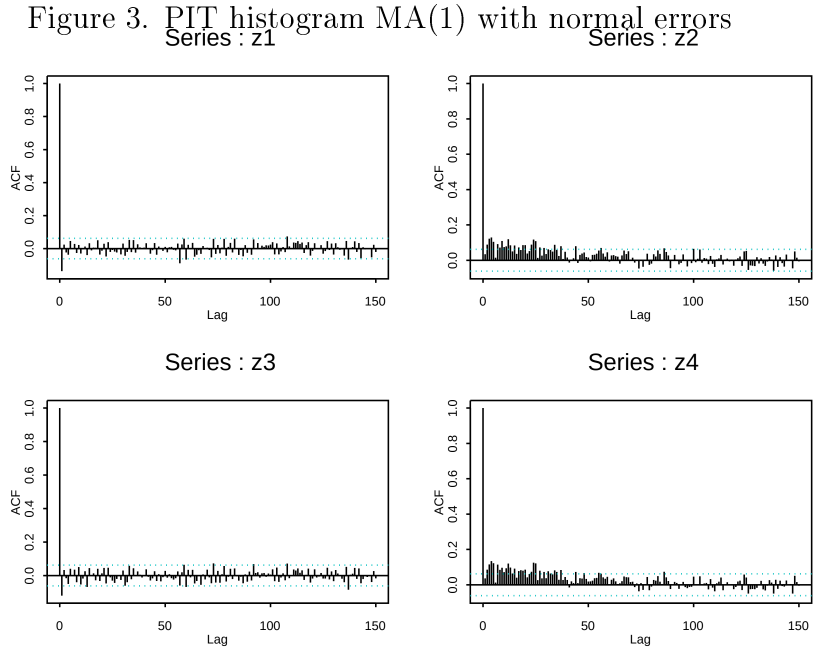

Figure 4. ACF of powers of PIT from MA(1)

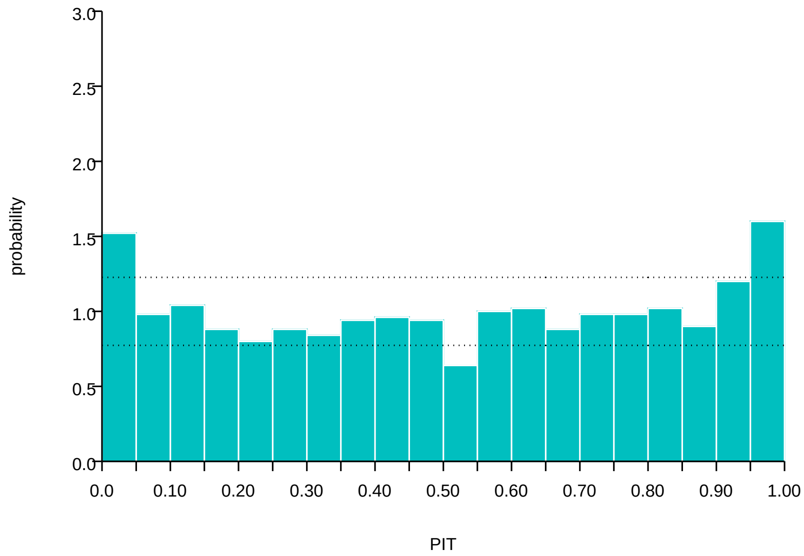

Figure 5. PIT histogram with MA(1)- normal GARCH(1,1)

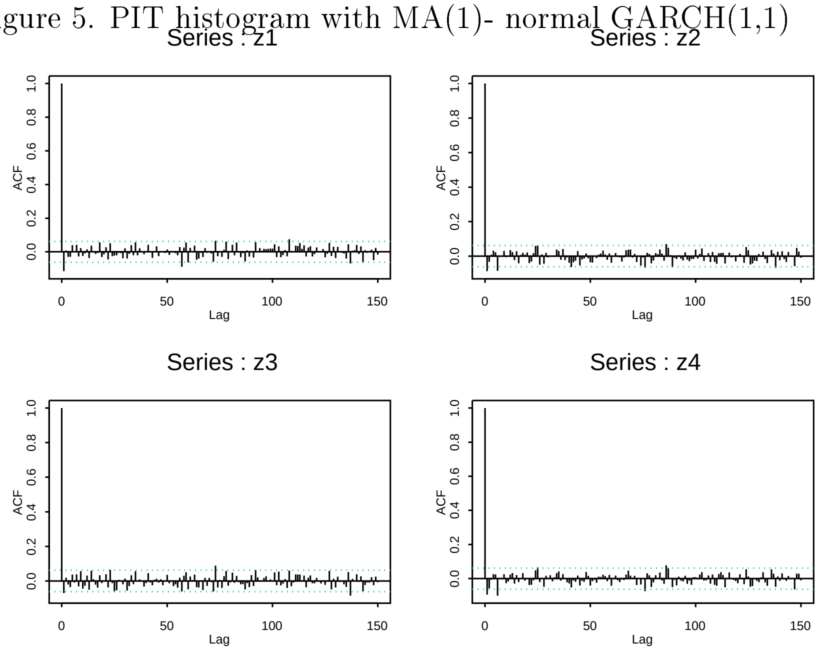

Figure 6. ACF of powers of PIT, MA(1)-normal GARCH(1,1)

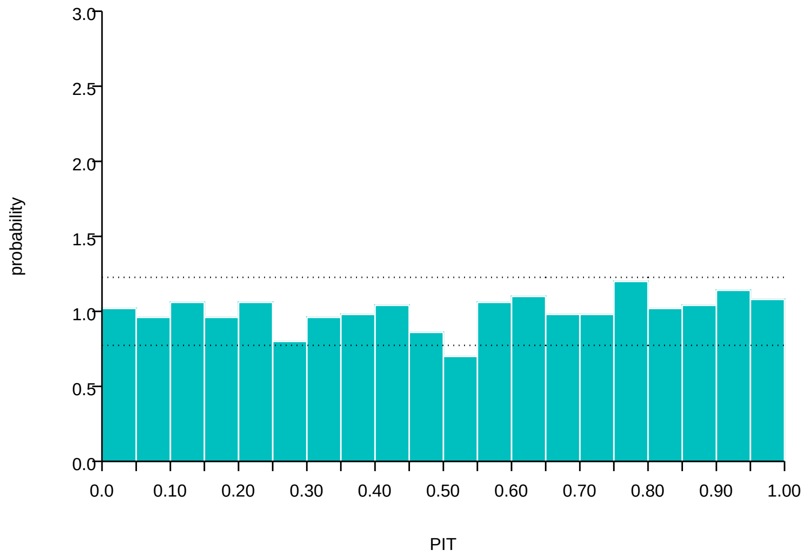

Fig 7. PIT histogram with MA(1)-tGARCH(1,1)

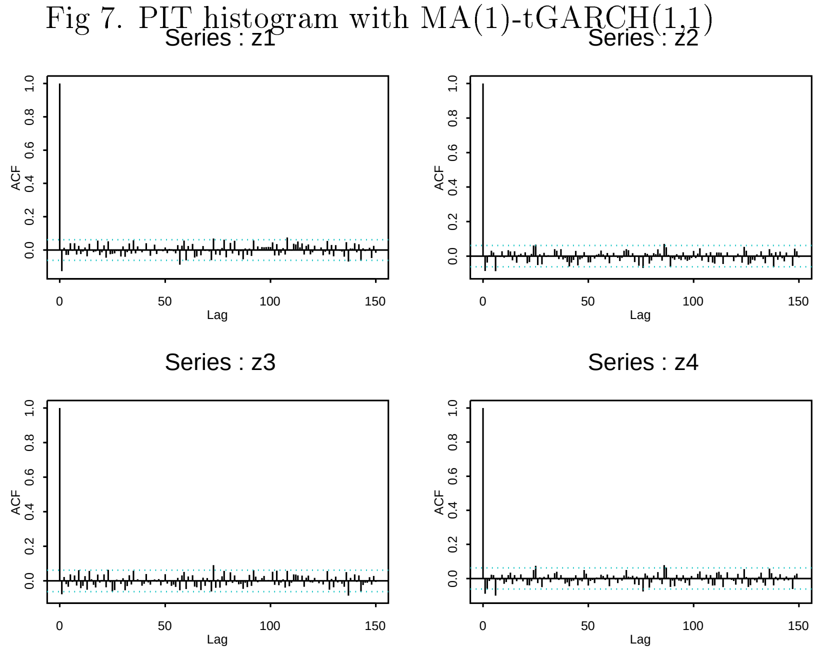

Figure 8. ACF of powers of PIT;MA(1)-tGARCH(1,1)

Data	Estimation	Testing
Observations	4133	1000
Mean	0.00016	4.02×10^{-4}
Standard Deviation	0.00756	0.0094
Skewness Coefficient	0.2828	0.2192
Excess Kurtosis	3.171	1.658
Minimum	-0.0367	-0.04
1^{st} Quartile	-0.0038	-0.0052
Median	3×10^{-4}	2.73×10^{-4}
3^{rd} Quartile	0.0042	0.0058
Maximum	0.0502	0.0476

Table 3. Return distribution for estimation and test sample

	Test Statistic	Critical Values Upper .1%
D^+	1.92466955	1.859
D^-	2.72979457	1.859
KS	2.72979457	1.95
Kuiper	4.6601984	2.303
CvM	2.090640	1.167
A-D	17.18036	6.0
W	2.059542	0.385

Table 4. Goodness-of-Fit Statistics based on EDF with $m = 4133$ and $n = 1000$.

As attractive as it may seem, the above procedure is a subjective method of identifying the problems of a forecasted PDF after comparing 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, Cramér- von Mises or Anderson-Darling 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 all the global characteristics of distribution of value-weighted S&P 500 returns. In Table 5, we report the results of the smooth test.

Hypothesis	$\hat{\Psi}_4^2$	u_1^2	u_2^2	u_3^2	u_4^2
EDF	110.342***	0.3715	95.847***	4.6007**	9.5229***
	(6.1525E-23)	(0.5422)	(1.2412E-22)	(0.032)	(0.002)
MA(1) with Normal error	373.0118***	1.18103	213.7938***	.0196	158.0175***
	(1.8816E-79)	(0.2771)	(2.0427E-48)	(0.8888)	(3.0678E-36)
MA(1)- Normal GARCH (1,1)	53.2501***	1.2664	29.9906***	.0582	21.935***
	(7.5541E-11)	(0.2604)	(4.3415E-08)	(0.8094)	(2.8205E-06)
MA(1)- t_7 GARCH (1,1)	3.2386	1.1003	1.6677	.2916	.179
	(0.5187)	(0.2942)	(0.1966)	(0.5892)	(0.6723)

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

*** *significant at* 1%, ** *significant at* 5%, * *significant at* 10%.

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 110.342 which is highly significant. We also can identify that the main sources of this deviation in the overall $\hat{\Psi}_4^2$ statistic are the second (u_2^2) and fourth (u_4^2) components, although the third (u_3^2) component is also slightly significant. From analyzing this we can infer that, there must be some unaccounted for variability, mainly, in 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 heteroscedasticity, we can conclude that there are deviations in the directions of the second and fourth order polynomials which can be related to second and fourth moments.

In our next stage, we estimate a parametric MA(1) model with Gaussian error terms, we end up getting an overall $\hat{\Psi}_4^2$ statistic at 373.01 which is even more significant than the case when we used EDF's. This might be partly because of the imposition of the "Gaussianity". The discrepancy from the null hypothesis is in the direction of the second (u_2^2 =213.79) and fourth (u_4^2 =158.02) order normalized Legendre polynomials, however in this case the discrepancy seem to be more pronounced than the pure non-parametric case. However, we find that the third order term does not seem to have any significance in that direction whatsoever. Keeping this result in mind, we proceed to incorporate a more involved time varying volatility model through a GARCH(1,1) model for conditional heteroscedasticity (Bollerslev, 1986) keeping the MA(1) component for the mean equation and normal 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 (373.01 to 53.25), although it is still highly significant at 1% level. Once again a cursory inspection of the components revealed that the second and fourth components are still significant although by a much lesser degree ($u_2^2 = 29.99$ and $u_4^2 = 21.93$). Finally, we introduce a non-Gaussian error term in the form of Students' t distribution along with the MA(1)-GARCH(1) formulation. With this general model, we find that $\hat{\Psi}_k^2 = 3.24$ which is in the acceptance region, and so are all its 4 components. This implies that a time varying conditional heteroscedasticity component together with the MA(1) mean model with t error, we arrive at an acceptable model.

We also tried higher orders beyond $k = 4$ but the marginal impact was negligible in the final model. For example, we had $\hat{\Psi}_6^2 = 3.35192$ (p-value=0.76356)with $u_5^2 = 0.03874$, $u_6^2 = 0.07461$. Therefore, we had no reason to believe that any higher order will have a significant impact. We chose t distribution with 7 degrees of freedom, since that was the closest to maximize the likelihood functions over degrees of freedom 3 through 25. We should mention that, although we have chosen to divide our sample by 4133 and 1000, this is not necessarily an optimal split. In fact, we have seen that the actual size of the test goes up on an average as we increase the size of the test sample keeping the estimation sample the same. Diebold et al. (1998) used 4133 and 4298 split, and we suspect that this sample splitting will have very large implied size. We kept the estimation sample 4133 so as to compare the results obtained by Diebold et al. and our formal testing procedure. The other factor which also motivated the selection to some extent was the presence of the 1987 crash data being in the testing sample in Diebold et al. This also could have caused a regime change or significant other changes in model estimation and the calculation of PIT at that point (Weigend and Shi, 2000). Since for constructing the smooth test statistic we have to calculate powers of the probability integral transform for the normalized Legendre polynomials, the test results could be influenced by outlying observations (although PITs always take values between 0 and 1). However, the latter problem was of secondary importance for the choice of the split, from our point of view we wanted to take a test sample size that was big enough.

From Table 5. 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 u_2^2 and u_4^2 components we can say that, introduction of conditional heterscedasticity improved the model by reducing the "butterfly" pattern in the PIT histogram (or the ratio density function). This was further reduced and practically eliminated when we introduced a non-Gaussian error term. So although, the smooth test did not directly address whether there was dependence in the data, it did however pick up the effect of this unaccounted for dependence in the data through incorporation of conditional heteroscedasticity.

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 where 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. Furthermore, their PDF version of the informal test of forecasting accuracy following Diebold et al. (1998) is similar to the approach of the smooth test. The only difference is that the smooth test is based on the probability integral transform rather than the raw data. Therefore, the smooth test is more widely applicable. Their main purpose was to find a suitable measure of predicting accuracy using out-of-sample performances over different conditional and unconditional models of mixture distributions.

Our results from the modified smooth test indicate that part of the reason for the strong significance of the fourth order orthogonal polynomial, a term highly similar to the kurtosis of the distribution of the PIT, is a deviation in the second and fourth moments. This also indicates the 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 π_4, it is not possible to exactly separate out these two effects. Having said that, we believe that if we include an explicit *dependent* structure in the model and the presence of the second Legendre polynomial π_2, we can identify specific departures.

5 Monte Carlo Evidence

The following Figure 9 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 look at the plots of the components to see that the individual u_i^2 follow χ_1^2 asymptotically for sufficiently large test sample of size n (Figure 10).

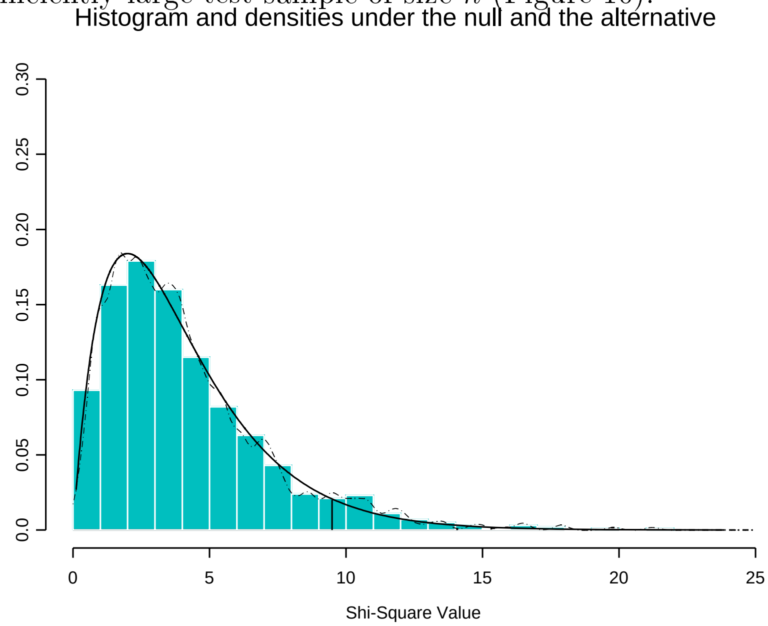

Figure 9: Distribution of Ψ_4^2

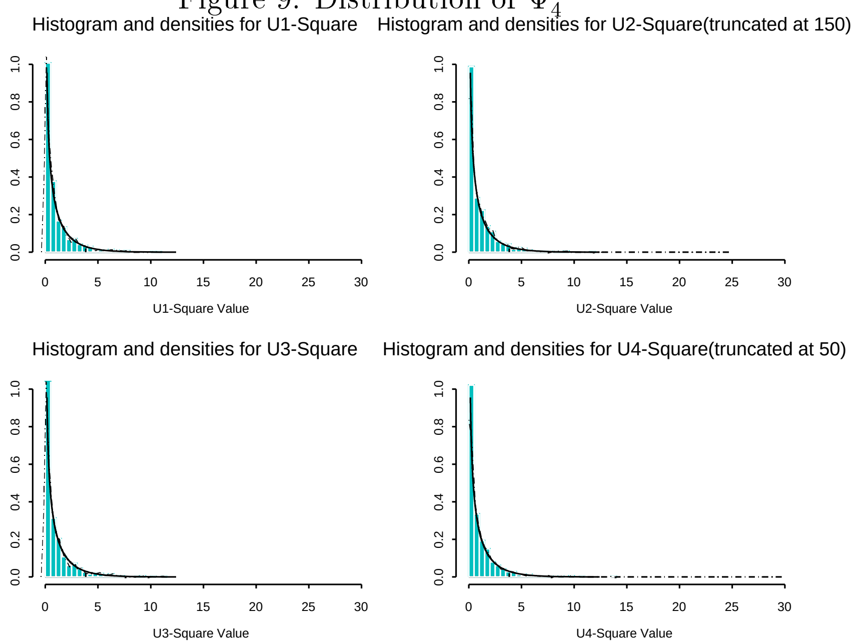

Figure 10. Size of individual u^2

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 (Bollerslev, 1986)

$$\begin{aligned} y_t &= \sqrt{\frac{5h_t}{7}} t_7 \\ h_t &= 0.2 + 0.15 y_{t-1}^2 + 0.65 h_{t-1}. \end{aligned} \tag{27}$$

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. The results are reported in Figures 11 and 12. These illustrate that even for the estimated parameters the Ψ_4^2 statistic seem to follow a central χ^2 distribution with 4 degrees of freedom. Not only that, the individual component u_i^2 also seem to follow χ_1^2 distributions to a large extent under the correct specification of the model.

One of the very important questions that left to be answered is the fact that what should be the sample split used in order to estimate the parameters to a fair degree of accuracy so that the modified smooth test would be consistent and have a level of significance close to the nominal size. We kept the initial estimation sample size $m = 2000$ the same and looked at several different sample sizes for the testing sample (n) and obtained the actual size of these tests with 25 different sample sizes with 200 replications to obtain the size of the test (Figure 11). We observe that although the actual size does not reach the nominal 5% level of the test, it is very close at approximately $n = 400$ to 500, so for our smooth test on S&P 500 returns with $m = 4133$, we chose the maximum which is a 4:1 split of the sample size.

Figure 11. Plot of sample size with the test size ($m = 2000$)

We must mention that the where the exact split should occur is still not obtained analytically and is a one of the main objectives of our ongoing research. The split that we have chosen was in part to compare our results with those of Diebold et al. (1998).

6 Conclusion

Neyman's smooth-type test can also be used in other areas of macroeconomics such as evaluating the density forecasts of realized inflation rates. Diebold, Tay and Wallis (1999) used a graphical technique as did Diebold et al. (1998) on the density forecasts of inflation from the *Survey of Professional Forecasters*. Neyman's smooth test in its original form was intended mainly to provide an *asymptotic test* of significance for testing goodness-of-fit for "smooth" alternatives. So, one can argue that although we have large enough data in the daily returns of the S&P 500 Index, we would be hard pressed to find similar size data for macroeconomic series such as GNP, inflation. This might make the test susceptible to significant small-sample fluctuations, and the results of the test might not be strictly valid. In order to correct for size or power problems due to small sample size, we can do a size correction [similar to other score tests, see Harris (1985), Harris (1987), Cordeiro and Ferrari (1991), Cribari-Neto and Ferrari (1995) and Bera and Ullah (1991) for applications in econometrics]. This promises to be an interesting direction for future research.

We can easily extend Neyman's smooth test 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. For example, if we have a sequence of the joint density forecasts of more than one, say 3, daily foreign exchange rates over a period of $1,000$ days, we can evaluate its accuracy using the smooth test for $3,000$ univariate densities. One thing that must be mentioned here, there could be both temporal and contemporaneous dependencies in these observations, we are assuming that taking conditional distribution both with respect to time and across-variables is feasible (see, for example, Diebold, Hahn and Tay 1999, p. 662).

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. We have indicated some ways to capture dependence and we need to pursue it further. Also, since the Smooth test is really a score test it enjoys the optimal properties of tests based on the score function and we don't 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 dependent structures.

7 Appendix A (Theorem 2)

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

$$\begin{aligned}\hat{\Psi}_k^2 - \Psi_k^2 &= \sum_{j=1}^{k} \frac{1}{n} \left[\left(\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \hat{\beta} \right) \right) \right)^2 - \left(\sum_{i=1}^{n} \pi_j \left(F \left(x_i; \beta \right) \right) \right)^2 \right] \\ &= \sum_{j=1}^{k} \left[\hat{u}_j^2 - u_j^2 \right]. \end{aligned} \tag{28}$$

Now applying the Mean Value Theorem we get

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

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

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

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

$$= 2\left(\frac{n}{\sqrt{m}}\right)\left[\sqrt{m}\left(\hat{\beta}-\beta\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \times \left[\frac{1}{n}\sum_{i=1}^{n} \frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \tag{29}$$

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 $j = 1,2,...,k$,

$$E\left(\pi_j\left(y_i\right)\right) = \int_0^1 \pi_j(u)\,du = 0. \tag{30}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta\right)\right) \xrightarrow{p} E\left(\pi_j\left(y_i\right)\right) = 0. \tag{31}$$

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)

$$h(y) = \frac{f(x;\beta)}{f(x;\beta^*)} = \frac{f\left(F^{-1}(y;\beta);\beta\right)}{f\left(F^{-1}(y;\beta^*);\beta^*\right)}. \tag{32}$$

which is a valid density function if F is absolutely continuous. Hence, $y_1, y_2, .., y_n$ are *IID* random variables with a density function $h(y)$ and has a finite first moment as y is defined on a compact support $[0,1]$ by dominated convergence theorem

$$0 \leq E(y) = \int_0^1 y \frac{f\left(F^{-1}(y;\beta);\beta\right)}{f\left(F^{-1}(y;\beta^*);\beta^*\right)} dy \leq \int_0^1 \frac{f\left(F^{-1}(y;\beta);\beta\right)}{f\left(F^{-1}(y;\beta^*);\beta^*\right)} dy = 1.$$

Hence, using the WLLN (Khinchine's theorem) for $j = 1, 2, ..., k$,

$$\frac{1}{n}\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta^*\right)\right) \xrightarrow{p} E\left[\pi_j\left(F\left(x_i;\beta^*\right)\right)\right]. \tag{33}$$

Now, we have $\hat{\beta} \xrightarrow{p} \beta$ as $\hat{\beta}$ is a $\sqrt{m}-$consistent estimator of β. Since, $\left|\hat{\beta}-\beta\right| \geq |\beta^*-\beta|$, for any $\varepsilon > 0$

$$0 \leq P\left\{|\beta^*-\beta| > \varepsilon\right\} \leq P\left\{\left|\hat{\beta}-\beta\right| > \varepsilon\right\},$$

hence, β^* is also converges to β in probability. If $\pi_j\left(F\left(x;\beta\right)\right)$ is a continuous function of β at $\beta = \beta^*$ we have

$$E\left[\pi_j\left(F\left(x;\beta^*\right)\right)\right] \xrightarrow{p} E\left[\pi_j\left(F\left(x;\beta\right)\right)\right],\ j = 1, 2, ..., k. \tag{34}$$

Hence, as m and n goes to infinity, using results in (30), (31), (33) and (34) we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta^*\right)\right) \xrightarrow{p} E\left[\pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \xrightarrow{p} E\left[\pi_j\left(F\left(x;\beta\right)\right)\right] = 0 \qquad (35)\\ &\Rightarrow \frac{1}{n}\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta^*\right)\right) = a_1 = o_p(1). \end{aligned}$$

Please note, that it is essential for this result that H_0 is true, otherwise we will only have that $\frac{1}{n}\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta^*\right)\right) = O_p(1)$ (bounded in probability). Furthermore, applying the WLLN yet again we get that for sufficiently large m,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta} &\xrightarrow{p} E\left[\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \\ &\xrightarrow{p} E\left[\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] < \infty \\ \Rightarrow \frac{1}{n}\sum_{i=1}^{n} \frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta} &= a_2 = O_p(1). \qquad (36) \end{aligned}$$

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

$$\sqrt{m}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{\beta}^{-1}\right) \tag{37}$$

where $\mathcal{I}_{\beta}$ is the Fisher information of $\hat{\beta}$. Hence, if $\mathcal{I}_{\beta} < \infty$,

$$\sqrt{m}\left(\hat{\beta}-\beta\right) = a_3 = O_p\left(1\right). \tag{38}$$

Hence from equation (29) using the results in (35), (36) and (38) we immediately get

$$\begin{aligned}
\hat{u}_j^2 - u_j^2 &= 2\left(\frac{n}{\sqrt{m}}\right)\left[\sqrt{m}\left(\hat{\beta}-\beta\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \\
&\times\left[\frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \\
&= 2\frac{n}{\sqrt{m}}a_1 a_2 a_3 \\
&= \frac{n}{\sqrt{m}}o_p\left(1\right).
\end{aligned} \tag{39}$$

From (28) using (39) we get for fixed k

$$\hat{\Psi}_k^2 - \Psi_k^2 = \frac{n}{\sqrt{m}}o_p\left(1\right). \tag{40}$$

which proves Theorem 2. ■

References

[1] A.K. Bera, Y. Bilias. 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, v. 97, pp.9-44 (2001a).

[2] A.K. Bera, Y. Bilias. On some optimality properties of Fisher-Rao score function in testing and estimation. Communications in Statistics, Theory and Methods, v. 30, 1533-1559 (2001b).

[3] A. K. Bera, A. Ghosh, 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, Marcel Dekker, pp. 177-230, 2001.

[4] A. Bera, A. Ghosh and Z. Xiao. Testing Equality of Two Densities Using Neyman's Smooth Test. Working Paper, 2003.

[5] A.K. Bera, A. Ullah. Rao's score test in econometrics. Journal of Quantitative Economics 7:189-220, 1991.

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

[7] P.J. Bickel, K.A. Doksum. Mathematical Statistics: Basic Ideas and Selected Topics. Oakland, California: Holden-Day, 1977. (not referenced)

[8] T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31:307-327, 1986.

[9] B. Boulerice, G.R. Ducharme. A note on smooth tests of goodness of fit for location-scale families. Biometrika 82:437-438, 1995.

[10] P.F. Christoffersen. Evaluating interval forecasts. International Economic Review 39:841-862, 1998.

[11] G.M. Cordeiro, S.L.P. Ferrari. A modified score test statistic having chi-squared distribution to order n^{-1}. Biometrika 78:573-582, 1991.

[12] H. Cramér . Mathematical methods of Statistics. New Jersey: Princeton University Press, 1946. (not referenced)

[13] F. Cribari-Neto, S.L.P. Ferrari. An improved Lagrange multiplier test for heteroscedasticity. Communications in Statistics–Simulation and Computation 24:31-44, 1995.

[14] C. Crnkovic, J. Drachman. Quality control. In: VAR: Understanding and Applying Value-at-Risk. London: Risk Publication. 1997.

[15] R.B. D'Agostino, M.A. Stephens. Goodness-of-Fit Techniques. New York: Marcel Dekker, 1986. (not referenced)

[16] F.X. Diebold, T.A. Gunther, A.S. Tay. Evaluating density forecasts with applications to financial risk management. International Economic Review 39:863-883, 1998.

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

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

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

[20] C.W.J Granger, M.H. Pesaran. Economic and Statistical Measures of Forecast Accuracy. Journal of Forecasting 19:537-560, 2000.

[21] P. Harris. An asymptotic expansion for the null distribution of the efficient score statistics. Biometrika 72:653-659, 1985.

[22] P. Harris. Correction to 'An asymptotic expansion for the null distribution of the efficient score statistic.' Biometrika 74:667, 1987.

[23] K. J. Kopecky, D.A. Pierce. Efficiency of smooth goodness-of-fit tests. Journal of the American Statistical Association 74: 393-397, 1979. (not referenced)

[24] T. Ledwina. Data-driven version of Neyman's smooth test of fit. Journal of the American Statistical Association 89:1000-1005, 1994.

[25] E.L. Lehmann. Testing Statistical Hypothesis. New York: John Wiley & Sons, 1959. (not referenced)

[26] G.S. Maddala. Econometrics. New York: McGraw-Hill, 1977. (not referenced)

[27] J. Neyman. "Smooth test" for goodness of fit. Skandinaviske Aktuarietidskrift 20:150-199, 1937.

[28] J. Neyman. Optimal asymptotic test of composite statistical hypothesis. In: U. Grenander, ed. Probability and Statistics, the Harold Cramér Volume. Uppsala: Almqvist and Wiksell. 1959, pp. 213-234.

[29] E.S. Pearson. The probability integral transformation for testing goodness of fit and combining independent tests of significance. Biometrika 30:134-148, 1938.

[30] K. Pearson. 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, 1900.

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

[32] C.R. Rao. Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, 1952. (not referenced)

[33] C.R. Rao. Linear Statistical Inference and Its Applications. New York: John Wiley & Sons, 1973.

[34] C.R. Rao, S.J. Poti. On locally most powerful tests when the alternatives are one-sided. Sankhyā 7:439, 1946.

[35] J.C.W. Rayner, D.J. Best. Smooth Tests of Goodness of Fit. New York: Oxford University Press, 1989.

[36] M. Rosenblatt. Remarks on a multivariate transformation. Annals of Mathematical Statistics 23:470-472, 1952.

[37] A.S. Tay, K.F. Wallis. Density forecasting: A Survey. Journal of Forecasting 19:235-254, 2000.

[38] D.R. Thomas, D.A. Pierce. Neyman's smooth goodness-of-fit test when the hypothesis is composite. Journal of the American Statistical Association 74:441-445, 1979.

[39] A. Weigend, S. Shi. Predicting Daily Probability Distributions of S&P500 Returns. Journal of Forecasting, v. 19:375-392, 2000.