7-2015

Density forecast evaluation for dependent financial data: Theory and applications

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

Anil K. BERA
University of Illinois at Urbana-Champaign

Density Forecast Evaluation for Dependent Financial Data: Theory and Applications

Aurobindo Ghosh
Lee Kong Chian School of Business,
Singapore Management University,
50 Stamford Road, #04-01,
Singapore-178899. Singapore.
e-mail: aurobindo@smu.edu.sg. Phone: +65-6828-0863

Anil K. Bera
Department of Economics, University of Illinois,
225E, David Kinley Hall, 1407 W. Gregory,
Urbana, Illinois 61801. U.S.A.
e-mail: abera@illinois.edu. Phone: 217-333-4596

JEL Classification: C12, C52, C53

Aug 31, 2016

Abstract

In this paper, we propose a *formal test* for density forecast evaluation in presense of dependent data. Apart from accepting or rejecting the tested model, our smooth test identifies the possible sources (such as the location, scale and shape of the distribution) of rejection, thereby helping in revising the initial model. We also propose how to augment the smooth test to investigate explicit forms of dependence in the data within the same test framework. An extensive application to S&P 500 returns indicate capturing time-varying volatility and non-gaussianity significantly improve the performance of the model. Although we are dealing with index returns, the proposed smooth test can be applied to other financial data for exchange rates, futures or forward markets, options prices, inflation rate, analyst forecasts among many others.

Highlights:

-Propose a smooth test for density forecast evaluation.
-The test can check for correct specification in the presence of both serial dependence (in higher order moments) and non-normality.
-The test can identify specific source(s) of mis-specification to address model revision.
-Monte Carlo results show good size and power properties of the proposed test in small samples.
- A substantive application using S&P 500 return data is provided.

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

1 Introduction and Motivation

In the statistical estimation literature 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 the need for a density-forecast, though construction of density forecast in empirical work is a relatively recent phenomenon (Ross, 2015, Liu, Shackleton, Taylor and Xu, 2007). While there are some empirical and Monte Carlo evidence of more acurate volatility estimates with semiparametruc models of diffusion for short term interest rate (Hou and Suardi, 2011), there has been only a few papers, we are aware of, that directly address the question of a formal test of evaluation of density forecasts; such as Diebold, Gunther and Tay (1998), Berkowitz (2001), Hong (2001), Bai (2003), Wallis (2003), Sarno and Valente (2004), Hong and Li (2005), see Corradi and Swanson (2006 a,b) for a review. 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. In particular, forecasting only certain moments of teh distribution are arguably more susceptible to challenging issues on model instability in return prediction (Pesaran and Timmermann, 2002, Paye and Timmermann, 2006). 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."

One important aspect of our test is that in the spirit of the test proposed by Neyman (1937) who coined the word "smooth" to represent *local or contiguous* alternative distribution, we use disjoint intervals to do estimation and evaluation of the density forecasts. The central issue driving the focus on out-of-sample evaluation is that the parameter estimates are independent of the data, so reducing the possibility that under the null hypothesis the probability integral transforms are not independent. In point forecast it has been noted that to compare tests of predictive accuracy out-of-sample behavior is of vital importance (Granger, 1980).

Our work is similar in spirit with Bontemps and Meddahi (2005), Hong and Li (2005) and Thompson (2008) but unlike them the proposed test enjoys the optimality benefits of a classical score test. We explicitly used moment conditions to capture dependencies, and the resultant augmented smooth test statistic has an asymptotic χ^2 distribution. After incorporating the type of depndence explicitly in the parametric model for the generalized residual (i.e., the probability integral transform), the proposed test alleviates the parameter estimation uncertainty and reduce the size distortion by allowing the sample size of the estimation diverge to infinity faster than evaluation sample sizes.

From a pure statistical perspective, density forecast evaluation is essentially a goodness-of-fit test problem. In a seminal paper, though rarely 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.

As an illustration from Hong (2001), consider a normal GARCH(1,1) formulation:

$$\begin{aligned} X_t &= \varepsilon_t \sqrt{h_t} \\ h_t &= \phi_0 + \phi_1 X_{t-1}^2 + \phi_2 h_{t-1} \\ \varepsilon_t &\overset{iid}{\sim} \mathcal{N}(0,1), \end{aligned}$$

where $\phi_i \geq 0, i = 0, 1, 2.$ We can write the conditional density function of X_t

$$f(X_t|\theta_0, \Omega_t) = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{X_t^2}{h_t}}, \text{ for } X_t \in (-\infty, \infty),$$

and the probability integral transform Y_t for "true" $\theta' = (\phi_0, \phi_1, \phi_2)'$ as

$$Y_t = \int_{-\infty}^{X_t} \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{z^2}{h_t}} dz = \Phi(\varepsilon_t) \overset{iid}{\sim} U(0,1).$$

If we have a special IGARCH(1,1) process with $\phi_2 = 1 - \phi_1$ and $\phi_0 = 0$, (J.P. Morgan Riskmetrics, 1996), h_t reduces to

$$h_t = (1 - \phi_1) \sum_{j=1}^{\infty} \phi_1^{j-1} X_{t-j}^2,$$

where we have only one parameter ϕ_1 to estimate. The joint test of independence and uniformity provides an opportunity of correctly sized optimal tests based on the score test principle after taking into account parameter estimation uncertainty. This model also illustrates how dependent structure in the volatility term can be incorporated in the model explicitly by including selected conditional moment conditions to accommodate for specific types of dependence.

The plan of the rest of the paper is as follows. In Section 2 we review Neyman (1937) smooth test approach, for a fuller account see Bera and Ghosh (2001). Section 3 uses the framework of Diebold et al. (1998) and proposes a smooth test for density forecast evaluation. We augment the smooth test to explicitly test for the failure of the independence assumption in Section 4. Section 5 provides Monte Carlo results to examine size properties of the proposed test. An application to S&P 500 returns data is given in Section 6. Section 7 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 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) dropping the condition under H_0 for notational convenience

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

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

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

where $Q(y) = F^{-1}(y)$ is the quantile function of Y. Therefore, the density of Y can be written as

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

Although this is the ratio of two densities, $h(y)$ is a proper density function when F and G are absolutely continuous distribution functions on $(0,1)$ (F is strictly increasing). 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

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

$$\Rightarrow \int_0^1 \pi_i(y)\, dy = 0 \text{ if } i \neq 0 \text{ as } \pi_0(y) = 1. \tag{6}$$

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 a couple of simple 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. Thus we see that departures from the null hypothesis can be tested using an appropriate function (or functions) estimating the RDF, $h(y)$.

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,

$$\lambda = \left(\theta_1^2 + ... + \theta_k^2\right)^{\frac{1}{2}}, \tag{7}$$

between H_0 and H_1. The test statistic is

$$\Psi_k^2 = \sum_{j=1}^{k} \frac{1}{n} \left[\sum_{i=1}^{n} \pi_j(y_i)\right]^2, \tag{8}$$

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.

In Appendix A1A, we show that Ψ_k^2 can simply be obtained using Rao's (1948) score (RS) test principle. Thus we can write

$$\Psi_k^2 = RS = \Sigma_{j=1}^{k} u_j^2, \tag{9}$$

where $u_j = \left(\frac{1}{\sqrt{n}}\right) \Sigma_{i=1}^{n} \pi_j(y_i)$.

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, $f(x;\gamma)$ where γ is a parameter to be estimated. 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

$$\int_0^1 \{\pi_i(F(x;\gamma))\}\{\pi_j(F(x;\gamma))\}\,dF(x;\gamma) = \int_0^1 \{q_i(x;\gamma)\}\{q_j(x;\gamma)\}\,f(x;\gamma)\,dx = \delta_{ij}. \tag{10}$$

Then, from (3) and (4) the density under the alternative hypothesis takes the form

$$\begin{aligned} g(x;\gamma,\theta) &= h(F(x;\gamma))\frac{dy}{dx} \\ &= c(\theta;\gamma)\exp\left[\sum_{j=1}^{k}\theta_j q_j(x;\gamma)\right] f(x;\gamma). \end{aligned} \tag{11}$$

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

$$\Psi_k^2 = \sum_{j=1}^{k}\frac{1}{n}\left[\sum_{i=1}^{n} q_j(x_i;\gamma)\right]^2. \tag{12}$$

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 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

$$y_t = \int_{-\infty}^{x_t} f_t(u)\,du. \tag{13}$$

which has the density function

$$h_t(y_t) = 1,\ 0 < y_t < 1, \tag{14}$$

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

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 (13) 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, if $F_i(.)$ denotes the distribution function of the i^{th} variable X_i, the n random variables defined by*

$$Y_1 = F_1(X_1), Y_2 = F_2(X_2|X_1 = x_1),$$
$$..., Y_n = F_n(X_n|X_1 = x_1, X_2 = x_2, ..., X_{n-1} = x_{n-1})$$

are IID $U(0,1)$.

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

$$\begin{aligned} P(Y_i \le y_i, i = 1, 2, ..., n) &= \int_0^{y_1}\int_0^{y_2} ... \int_0^{y_n} f(x_1)\,dx_1 f(x_2|x_1)\,dx_2 ... f(x_n|x_1, ..., x_{n-1})\,dx_n \\ &= \int_0^{y_1}\int_0^{y_2} ... \int_0^{y_n} dt_1 dt_2 ... dt_n \\ &= y_1 y_2 ... y_n. \end{aligned} \tag{15}$$

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

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 $\pi_j(y)$ as the orthogonal polynomials that 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). The normalized Legendre polynomials are the natural orthogonal polynomials derived from the uniform distribution (see for example, Bontemps and Meddahi, 2012). Solving these the first five $\pi_j(y)$ are (Neyman 1937, pp. 163-164) $\pi_0(y) = 1$, $\pi_1(y) = \sqrt{12}\left(y - \frac{1}{2}\right)$, $\pi_2(y) = \sqrt{5}\left(6\left(y - \frac{1}{2}\right)^2 - \frac{1}{2}\right)$, $\pi_3(y) = \sqrt{7}\left(20\left(y - \frac{1}{2}\right)^3 - 3\left(y - \frac{1}{2}\right)\right)$, $\pi_4(y) = 210\left(y - \frac{1}{2}\right)^4 - 45\left(y - \frac{1}{2}\right)^2 + \frac{9}{8}$.

There is one 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 (8), with

$$y_i = F(x_i;\beta) = \int_0^{x_i} f(u;\beta)\, du, \ i = 1, 2, ..., n. \tag{16}$$

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

$$\hat{\Psi}_k^2 = \sum_{j=1}^{k} \hat{u}_j^2 = \sum_{j=1}^{k} \frac{1}{n}\left(\sum_{i=1}^{n} \pi_j(\hat{y}_i)\right)^2, \tag{17}$$

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 A1B. ∎

4 Augmented Smooth Test

In order to give a formal test of the dependent structure for the graphical procedure suggested by Diebold et al. (1998), Berkowitz (2001) proposed a formal likelihood ratio test. An advantage of his proposed test is that it gives some indication of the nature of the violation when the goodness-of-fit test is rejected. Berkowitz used the likelihood ratio test based on the inverse standard normal transformation of the probability integral transforms of the data. The main driving force behind the proposed transformation is the tractability of the Gaussian distribution. If $z_t = \Phi^{-1}\left(\hat{F}(y_t)\right)$ is the standard normal transform, an AR(1) model can be written as

$$z_t - \mu = \rho(z_{t-1} - \mu) + \varepsilon_t. \tag{18}$$

To test for independence, we can test $H_0 : \rho = 0$ in the presence of nuisance parameters μ and σ^2 (the constant variance of the error term ε_t). A joint likelihood ratio test for the parameters $\mu = 0, \rho = 0$ and $\sigma^2 = 1$ is based on

$$LR = -2\left(l(0,1,0) - l\left(\hat{\mu}, \hat{\sigma}^2, \hat{\rho}\right)\right), \tag{19}$$

which is asymptotically distributed as a χ^2 with three degrees of freedom, where $l(\theta) = \ln L(\theta)$ is the log-likelihood function, and where $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2, \hat{\rho})$ are the unrestricted maximum likelihood estimators. Explicit but separate tests of dependence have been proposed by Egorov, Hong and Li (2005) based on the methodology of Hong and Li (2005) in the out-of-sample case, while Thompson (2008) proposed a GMM type technique where moment conditions for dependence were explicitly used.

The null hypothesis that we are keen on testing is whether the PIT's are *independently and identically distributed* as $U(0,1)$. The main drawback of the Berkowitz (2001) procedure is that it tests only the independence aspect through only a first-order dependence as an alternative hypothesis. We show if dependence is included in a constructive way, it is possible to identify the cause and nature of departures from the null hypothesis. We discuss this below.

Let $(X_1, X_2, ..., X_n)$ has a joint probability density function (PDF) $g(x_1, x_2, ..., x_n)$. Define $\tilde{X}_1 = \{X_1\}$, $\tilde{X}_2 = \{X_2|X_1 = x_1\}$, $\tilde{X}_3 = \{X_3|X_2 = x_2, X_1 = x_1\}$, ..., $\tilde{X}_n = \{X_n|X_{n-1} = x_{n-1},\ \ X_{n-2} = x_{n-2}\ \ ..., X_1 = x_1\}$. Then we have

$$g(x_1, x_2, ..., x_n) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) ... f_{X_n|X_{n-1}X_{n-2}...X_1}(x_n|x_{n-1}, x_{n-2}, ..., x_1).$$

Furthermore, using Theorem 1 if we define $(Y_1, Y_2, ..., Y_n)$ as conditional cumulative distribution functions (CDF) of $(X_1, X_2, ..., X_n)$ or the probability integral transforms (PIT) evaluated at $(x_1, x_2, ..., x_n)$,

$$Y_1 = F_{X_1}(x_1), Y_2 = F_{X_2|X_1}(x_2|x_1), ..., Y_n = F_{X_n|X_{n-1}X_{n-2}...X_1}(x_n|x_{n-1}, x_{n-2}, ..., x_1)$$

are then distributed as *IID* $U(0,1)$. Suppose now, under null hypothesis H_0 of the true specification of the model CDF $F(.)$ or PDF $f(.)$, $(Y_1, Y_2, ..., Y_n) = (U_1, U_2, ..., U_n)$ where $U_t \sim U(0,1)$, $t = 1, 2, ...n$, so the joint PDF is

$$\begin{aligned} h(y_1, y_2, ..., y_n|H_0) &= h_1(y_1) h_2(y_2|y_1) ...h_n(y_n|y_{n-1}, y_{n-2}, ..., y_1) \\ &= 1.1....1 = 1. \end{aligned}$$

Under the alternative H_1, $Y_i's$ are neither uniformly distributed nor are they *IID*. Let us suppose the conditional density function of Y_t depends on p lag terms, that is to say,

$$\begin{aligned} h(y_t|y_{t-1}, y_{t-2}, ..., y_1) &= h(y_t|y_{t-1}, y_{t-2}, ..., y_{t-p}) \\ &= c(\theta, \phi) \exp\left[\sum_{j=1}^{k} \theta_j \pi_j(y_t) + \sum_{l=1}^{q} \phi_l \delta_l(y_t, y_{t-1}, ..., y_{t-p})\right], \end{aligned} \quad (20)$$

where we have assumed for now $k \geq q$. For simplicity, we start with $p = 1$, this could be more general than it sounds in one-step-ahead forecasts as we can test pairwise dependence including models like AR(1), ARCH(1) etc.

Theorem 3 *If the conditional density function under the alternative hypothesis is given by equation (20) and* $p = 1$, *the augmented smooth test statistic is given by*

$$\hat{\Psi}_k^2 = \begin{bmatrix} U'U + U'BEB'U - V'EB'U \\ -U'BEV + V'EV \end{bmatrix} = U'U + (V - B'U)' E (V - B'U)$$

has a central χ^2 *distribution with* $k + q$ *degrees of freedom where* U *is a* $k-$*vector of components* $u_j = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_j(y_t)$, $j = 1, ..., k$, V *is a* $q-$*vector of components* $v_l = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \delta_l(y_t, y_{t-1})$, $l = 1, ..., q$, $B = E[\pi\delta]$, $D = E[\delta\delta]$ *are components of the information matrix defined in equation (84) in Addendum A and* $E = (D - B'B)^{-1}$.

Proof. See APPENDIX A2. ■

Example 1: As an illustration of Theorem 3, let us consider a simple example of the smooth test for autocorrelation for

$$y_t - \mu = \rho\left(y_{t-1} - \mu\right) + \sigma_t \varepsilon_t \ , \tag{21}$$

where $E\left(\varepsilon_t\right) = 0$, $V\left(\varepsilon_t\right) = 1$, $\sigma_t = \sigma$.We define,

$$\delta_1\left(y_t, y_{t-1}\right) = \left(y_t - 0.5\right) y_{t-1} = \frac{1}{\sqrt{12}}\pi_1\left(y_t\right) y_{t-1} = a_1 \pi_1\left(y_t\right) y_{t-1} \tag{22}$$

with $a_1 = \frac{1}{\sqrt{12}}$. Then, we can denote $v_1 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\delta_1\left(y_t, y_{t-1}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(y_t - 0.5\right) y_{t-1}$. Given information set $\Omega_t = \left\{y_{t-1}, y_{t-2}, ...\right\}$, applying the Law of Iterative Expectation,

$$\begin{aligned}
&\int_0^1 a_1 \pi_j\left(y_t\right)\left(y_t - 0.5\right) y_{t-1} dy_t \\
&= y_{t-1}\int_0^1 \pi_j\left(y_t\right)\pi_1\left(y_t\right) dy_t = \begin{cases} a_1 y_{t-1} & j = 1 \\ 0 & j \neq 1 \end{cases} , \\
&\Rightarrow E\left[E\left[\int_0^1 \pi_j\left(y_t\right)\left(y_t - 0.5\right) y_{t-1} dy_t | \Omega_t\right]\right] = \begin{cases} a_1 E\left[y_{t-1}\right] = a_1 \mu & j = 1 \\ 0 & j \neq 1 \end{cases} .
\end{aligned} \tag{23}$$

Applying the Law of Iterative Expectation once again

$$\begin{aligned}
&\int_0^1 \left(\left(y_t - 0.5\right) y_{t-1}\right)^2 dy_t \\
&= a_1^2\left(y_{t-1}\right)^2 \int_0^1 \pi_1^2\left(y_t\right) dy_t \\
&= a_1^2\left(y_{t-1}\right)^2 \\
&\Rightarrow E\left[E\left[\int_0^1 \left(\left(y_t - 0.5\right) y_{t-1}\right)^2 dy_t | \Omega_t\right]\right] = a_1^2 E\left[y_{t-1}^2\right] = a_1^2\left(\sigma^2 + \mu^2\right) .
\end{aligned} \tag{24}$$

Hence, it follows that

$$\begin{aligned}
E\left[\pi\delta\right] &= \begin{pmatrix} a_1 E\left[y_{t-1}\right] & 0 & 0 & ... & 0 \end{pmatrix}' = B \\
E\left[\delta\delta\right] &= a_1^2 E\left[y_{t-1}^2\right] = D,
\end{aligned} \tag{25}$$

which in turn gives the information matrix

$$\mathcal{I}=n\left[\begin{array}{ccc}1 & 0_{k-1}^{\prime} & a_1\mu \\ 0_{k-1} & I_{k-1} & 0_{k-1} \\ a_1\mu & 0_{k-1}^{\prime} & a_1^2\left(\sigma^2+\mu^2\right)\end{array}\right], \tag{26}$$

where I_p is the identity matrix of order p and 0_p is a p^{th} order vector of $0's$. In order to evaluate the inverse of the information matrix in (26) we use the following results:

$$\begin{aligned} D-B'B &= a_1^2\left[E\left(y_{t-1}^2\right)-\left(E\left(y_{t-1}\right)\right)^2\right]=a_1^2\sigma^2, \\ U'BEB'U &= a_1^2u_1^2\mu^2/\left(a_1^2\sigma^2\right), \\ V'EB'U &= v_1u_1\mu/\left(a_1^2\sigma^2\right), \\ V'EV &= v_1^2/\left(a_1^2\sigma^2\right). \end{aligned} \tag{27}$$

Hence, using the above results we have a correction term as an LM test for autocorrelation (Breusch, 1978)

$$\begin{aligned} \Psi_{k+1}^2 &= \sum_{j=1}^{k}u_i^2+\frac{1}{\left(a_1^2\sigma^2\right)}\left[a_1^2u_1^2\mu^2-2a_1v_1u_1\mu+v_1^2\right] \\ &= \sum_{j=1}^{k}u_i^2+\frac{\left(v_1-a_1u_1\mu\right)^2}{\left(a_1^2\sigma^2\right)}=\sum_{j=1}^{k}u_i^2+\frac{12\left(v_1-u_1\mu/\sqrt{12}\right)^2}{\sigma^2}\overset{a}{\sim}\chi_{k+1}^2 \text{ under } H_0. \end{aligned} \tag{28}$$

The sample counterpart of the second expression in (28) is given by

$$12\left(\frac{\sqrt{\frac{1}{n}}\sum_{t=2}^{n}\left(y_t-0.5\right)y_{t-1}-\sqrt{\frac{1}{12n}}\sum_{t=1}^{n}\left(y_t-0.5\right)\frac{1}{n}\sum_{t=1}^{n}y_t}{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(y_t-\overline{y}\right)^2}}\right)^2\overset{a}{\sim}\chi_1^2. \tag{29}$$

It is evident that this will give us a test for autocorrelation of the first order in a *global* sense.

Example 2: In order to further illustrate this technique, let us consider a test for ARCH (1) type alternative with mean equation (21),

$$\sigma_t^2=\alpha_0+\alpha_1\sigma_{t-1}^2\varepsilon_{t-1}^2, \tag{30}$$

where we can define a function δ_2 for testing dependence

$$\delta_2(y_t, y_{t-1}) = y_t^2 y_{t-1}^2 - \frac{1}{3} y_{t-1}^2 = y_{t-1}^2 \left(y_t^2 - \frac{1}{3} \right) = y_{t-1}^2 \left(a_1 \pi_1(y_t) + a_2 \pi_2(y_t) \right), \quad (31)$$

where $a_1 = \frac{1}{\sqrt{12}}$ and $a_2 = \frac{1}{6\sqrt{5}}$, and $a_3 = a_1^2 + a_2^2 = \frac{4}{45}$ for notational convenience.

Hence, the smooth test statistic incorporating an ARCH(1) effect is

$$\begin{aligned} \hat{\Psi}_k^2 &= \sum_{j=1}^{k} u_j^2 + \left(\frac{4}{45} \left(m_4 - m_2^2 \right) \right)^{-1} \left[v_1 - \frac{1}{\sqrt{12}} m_2 u_1 - \frac{1}{6\sqrt{5}} m_2 u_2 \right]^2 \\ &\sim \chi_{k+1}^2(0) \end{aligned} \quad (32)$$

where as defined before $m_j = E\left(y_{t-1}^j\right)$ (see Ghosh and Bera, 2004). Similarly, we can obtain a test only for leverage effect with

$$\delta_3(y_t, y_{t-1}) = y_{t-1} \left(y_t^2 - \frac{1}{3} \right) = y_{t-1} \left(a_1 \pi_1(y_t) + a_2 \pi_2(y_t) \right)$$

that yields the test statistic as a correction term

$$\begin{aligned} \hat{\Psi}_k^2 &= \sum_{j=1}^{k} u_j^2 + \left(\frac{4}{45} \left(m_2 - m^2 \right) \right)^{-1} \left[v_1 - \frac{1}{\sqrt{12}} m_2 u_1 - \frac{1}{6\sqrt{5}} m_2 u_2 \right]^2 \\ &\sim \chi_{k+1}^2(0). \end{aligned} \quad (33)$$

The joint test of both Leverage effect and ARCH type effects is more involved but can be derived from the shortcut matrix formula for the correction term given

$$E = \frac{1}{\Delta} \begin{bmatrix} a_3 \sigma_{22} & -a_3 \sigma_{12} \\ -a_3 \sigma_{12} & a_3 \sigma_{11} \end{bmatrix}, V - B'U = \begin{bmatrix} v_1 - (a_1 u_1 + a_2 u_2)\mu \\ v_2 - (a_1 u_1 + a_2 u_2)\mu_2 \end{bmatrix}$$

where $\sigma_{ij} = \mu_{i+j} - \mu_i \mu_j$ and $\Delta = a_3^2 \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2 \right)$ as

$$\frac{1}{\Delta} \begin{bmatrix} (v_1 - (a_1 u_1 + a_2 u_2)\mu)^2 a_3 \sigma_{22} \\ -2 (v_1 - (a_1 u_1 + a_2 u_2)\mu)(v_2 - (a_1 u_1 + a_2 u_2)\mu_2) a_3 \sigma_{12} \\ + (v_2 - (a_1 u_1 + a_2 u_2)\mu_2)^2 a_3 \sigma_{11} \end{bmatrix} \sim \chi_2^2. \quad (34)$$

Similarly, a joint test of AR(1) and ARCH(1) effects can be shown to be function of the first 4 raw moments m_j, $j = 1, ..., 4$ of y_{t-1}, besides the score functions $u_j's$ and $v_l's$ (not included here).

Unfortunately, the choice of the dependency function $\delta_l(y_t, y_{t-1}), l = 1, 2, ..., q$ (a moment condition to capture the dependent structure) involves a trade-off. On one hand, the smaller the number q there are fewer parameters to estimate, however, there will be a loss of power owing to the types of dependencies that are ignored; on the other, if q is large we will suffer from a curse of dimensionality as there will be several parameters to be estimated based on the same data. In the following examples we illustrate how to incorporate more general dependence structures like ARMA(1,1), GARCH (1,1) and several ARCH parameters, and not to increase the dimensionality of the problem substantially under certain regularity conditions.

Suppose we want to incorporate an ARMA (1,1) error term of the following form (See Bera and Ra, 1994; Andrews and Ploberger, 1996)

$$y_t - (\phi + \xi) y_{t-1} = \varepsilon_t - \phi\varepsilon_{t-1}, \tag{35}$$

where ε_t is *IID* $\mathcal{N}(0, \sigma_\varepsilon^2)$, ϕ and $(\phi + \xi) \in (-1, 1)$ and $t = 1, 2, ..., n$. Here, to test for white noise we can test $H_0 : \xi = 0$ against $H_1 : \xi \neq 0$. It is worth noting that under H_0, the parameter ϕ becomes unidentified, hence we have a nuisance parameter under the null which is often termed as the Davies' problem (Davies 1977, 1987). We will start of assuming that the parameter ϕ is fixed and then relax that assumption to do the test. Define the dependency function

$$\delta_1(y_t, y_{t-1}, ..., y_1) = \pi_1(y_t) \sum_{s=1}^{t-1} \phi^{s-1} y_{t-s}. \tag{36}$$

Hence, if ϕ is a known constant, as shown in the Addendum A subsection 7.2, the smooth test statistic incorporates the LM test similar to Andrews and Ploberger (1996),

$$\begin{aligned} \hat{\Psi}_k^2 &= \sum_{j=1}^{k} u_j^2 + v_1^2 \frac{(1 - \phi^2)}{\sigma_\varepsilon^2} \\ &\sim \chi_{k+1}^2 \text{ where } v_1 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_i(y_t) \sum_{s=1}^{t-1} \phi^{s-1} y_{t-s}. \end{aligned} \tag{37}$$

It is worth noting that putting $\phi = 0$ we get back the test using AR(1) terms.

Let us now work out the example for a GARCH(1,1) type dependent structure with the conditional variance of the form

$$\begin{aligned}h_t &= (1-\omega_1)\zeta + \omega_1 h_{t-1} + \alpha_1 u_{t-1}^2 \\ &\Rightarrow h_t + u_t^2 = (1-\omega_1)\zeta - \omega_1\left(u_{t-1}^2 - h_{t-1}\right) + \omega_1 u_{t-1}^2 + \alpha_1 u_{t-1}^2 + u_t^2 \\ &\Rightarrow u_t^2 - \gamma_1 u_{t-1}^2 = (1-\omega_1)\,\zeta + w_t - \omega_1 w_{t-1}, \end{aligned} \tag{38}$$

where $\gamma_1 = (\alpha_1 + \omega_1)$ and $w_t = u_t^2 - h_t$ is serially uncorrelated shows that u_t^2 is ARMA(1,1). In order to test whether the errors are simply white noise against the alternative that they are GARCH (1,1) we can test $H_0 : \alpha_1 = 0$ against $H_1 : \alpha_1 \neq 0$. It is easy to see that under H_0, (38) gives

$$\begin{aligned} u_t^2 - \omega_1 u_{t-1}^2 &= (\alpha_1 + \omega_1)\,\zeta + w_t - \omega_1 w_{t-1} \\ \Rightarrow \left(u_t^2 - \zeta\right) - \omega_1\left(u_{t-1}^2 - \zeta\right) &= w_t - \omega_1 w_{t-1} \text{ under } H_0 : \alpha_1 = 0 \\ \Rightarrow \tilde{u}_t^2 &= w_t \text{ where } \tilde{u}_t^2 = u_t^2 - \zeta. \end{aligned} \tag{39}$$

Although, we have shown that (39) gives us a valid test procedure for testing white noise against GARCH (1,1) errors however the nuisance parameter ω_1 only appears under the alternative. Let us first assume that ω_1 is a known constant. Then from (38) we can write

$$\begin{aligned} w_t &= \omega_1 w_{t-1} + \nu_t - \kappa, \text{ where } \kappa = (1-\omega_1)\zeta,\ \nu_t = u_t^2 - \gamma_1 u_{t-1}^2 \\ &= \sum_{s=0}^{\infty} \omega_1^s \nu_{t-s} - \sum_{s=0}^{\infty} \omega_1^s (1-\omega_1)\zeta \\ &= \sum_{s=0}^{\infty} \omega_1^s \nu_{t-s} - \zeta. \end{aligned} \tag{40}$$

Hence,

$$\begin{aligned} w_t &= \sum_{s=0}^{\infty} \omega_1^s \left(u_{t-s}^2 - \gamma_1 u_{t-s-1}^2\right) - \zeta = \tilde{u}_t^2 - (\gamma_1 - \omega_1) \sum_{s=1}^{\infty} \omega_1^{s-1} u_{t-s}^2 \\ &\Rightarrow h_t = \zeta + (\gamma_1 - \omega_1) \sum_{s=1}^{\infty} \omega_1^{s-1} u_{t-s}^2. \end{aligned} \tag{41}$$

Since, we have data $t = 1, ..., n$ we have to truncate $h_t = \zeta + (\gamma_1 - \omega_1) \sum_{s=1}^{t-1} \omega_1^{s-1} u_{t-s}^2$. We can derive a RS type test (like the supLM or AvgLM test , Andrews and Ploberger 1996) of white noise against GARCH (1,1) error or obtain a range of the p-value of a RS test for different values of ω_1 given that ω_1 is actually not known.

Let us now define the function

$$\delta_1\left(y_t, y_{t-1}, ..., y_1\right) = \pi_2\left(y_t\right)\sum_{s=1}^{t-1}\omega_1^{s-1}y_{t-s}^2 \tag{42}$$

to capture GARCH (1,1) type dependence in the data. Hence, we have the score function related to δ_1 is given by

$$v_1 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\pi_2\left(y_t\right)\sum_{s=1}^{t-1}\omega_1^{s-1}y_{t-s}^2. \tag{43}$$

In Addendum A, subsection 7.3, we have worked out the corresponding smooth test. If ω_1 is a known constant, the test statistic is

$$\begin{aligned}\hat{\Psi}_k^2 &= \sum_{j=1}^{k}u_j^2 + \left(v_1 - \frac{E\left(y_t^2\right)u_2}{1-\omega_1}\right)^2\left(\frac{E\left(y_t^4\right)}{1-\omega_1^2} - \left(\frac{E\left(y_t^2\right)}{1-\omega_1}\right)^2\right)^{-1} \\ &\sim \chi_{k+1}^2.\end{aligned}$$

In the operational form of $\hat{\Psi}_k^2$ we repalce the population moments $E\left(y_t^j\right)$ by their sample counterparts. If we put $\omega_1 = 0$, we would get back a test for ARCH(1) dependence. However, in general since ω_1 is unknown we will have to use methods described for the ARMA(1,1) case to handle this problem of nuisance parameter which exists only under the alternative in this setup. Although this setup is quite general for dealing with the GARCH(1,1) case, however, the problem with the nuisance parameter existing only under the alternative makes the decision problem non-trivial, whether to use a maximum RS statistic or give the p-values over a whole range of values of ω_1 or equivalently give probability bounds for the test statistic. We would finally suggest a procedure inspired by Engle (1982, 1983) where we considered a weighted ARCH type alternative. The conditional variance function suggested by Engle (1982, 1983) was

$$h_t = \alpha_0 + \alpha_1\sum_{s=1}^{r}w_s u_{t-s}^2, \text{ where } w_s = \frac{(r+1)-s}{\frac{1}{2}r\left(r+1\right)} \text{ for some fixed } r. \tag{44}$$

We derive the RS test for testing $H_1 : \alpha_1 = 0$ against $H_1 : \alpha_1 \neq 0$ in Addendum A, subsection 7.4. The smooth test statistic is given by

$$\begin{aligned}\hat{\Psi}_k^2 &= \sum_{j=1}^{k} u_j^2 + \left(v_1 - E\left(y_t^2\right) u_2\right)^2 \\ &\times \left(\frac{2}{3}\frac{(2r+1)}{r(r+1)} E\left(y_t^4\right) - \left(E\left(y_t^2\right)\right)^2\right)^{-1} \\ &\sim \chi_{k+1}^2.\end{aligned}$$

5 Monte Carlo Study

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.

Insert Figure 6 here.

Insert Figure 7 here

.

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:

$$\begin{aligned}y_t &= \sqrt{\frac{5h_t}{7}} t_7 \\ h_t &= 0.2 + 0.15 y_{t-1}^2 + 0.65 h_{t-1}.\end{aligned} \tag{45}$$

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 $m/4$.

Insert Figure 8 here.

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.

We further investigated the size and power properties of the smooth test statistic under different hypothesis and data generating process. We start of with the following $MA(1) - GARCH(1,1)$ model with error distributed as t_7 given by the model

$$
\begin{aligned}
y_t &= \varepsilon_t + 0.2\varepsilon_{t-1}, \varepsilon_t = \sqrt{\frac{5h_t}{7}}t_7, \\
h_t &= 0.2 + 0.15y_{t-1}^2 + 0.80h_{t-1}.
\end{aligned} \tag{46}
$$

Here, keeping with the empirical results in the literature, we take a stronger form of GARCH, where the sum of the two coefficients is close to 1, i.e., $0.15 + 0.80 = 0.95$. The distributions of the unmodified smooth test statistic $\widehat{\Psi}_k^2$, and the augmented smooth test is given in Figure 9 below. We compare the kernel density estimates with normal kernel (default for optimal bandwidth selection on MATLAB), and compare with the χ_4^2 and χ_6^2 distributions respectively under the null hypoethesis $H_0 : F = G$. We have used the joint test for AR(1) and ARCH(1), and weighted ARCH with $\omega_1 = 0.1$ from (42). We also truncated the infinite lag series at 20 lags. We observe that with sample size $m = 5000$ and $n = 1000$, gave close forecast distribution of the model for the unmodified smooth test in Figure 9. However, Figure 9 shows that the augmented (modified) smooth test has more size distortion which is true for many score-type tests with estimated parameters, and can be adjusted using finite sample correction. One further note is that the weighted AR and ARCH models given in Table 5 and Figure 9 are using fixed value of ϕ and ω, ideally due to the the Davis' problem we should look at a Sup LM or Sup LR type test which should substantially reduce size distortion (see Andrews and Ploberger, 1996)

Insert Figure 9 here.

Now, lets generate the data from a MA(1)-GJR-GARCH(1,1) where

$$
\begin{aligned}
y_t &= \varepsilon_t + 0.2\varepsilon_{t-1}, \varepsilon_t = \sqrt{\frac{5h_t}{7}} t_7, \\
h_t &= 0.2 + (0.15 + 0.2I\{y_{t-1} < 0\})y_{t-1}^2 + 0.70h_{t-1}.
\end{aligned} \tag{47}
$$

using the *leverage* coefficient $L = 0.2$ multiplying $I\{y_{t-1} < 0\}$ We estimated a naive GARCH(1,1) model with gaussian error

$$
\begin{aligned}
y_t &= \varepsilon_t \sqrt{h_t}, \varepsilon_t \tilde{} N(0,1), \\
h_t &= a_0 + a_2 y_{t-1}^2 + a_1 h_{t-1}.
\end{aligned} \tag{48}
$$

to calculate the probability integral transforms. The distribution of the smooth test statistic $\widehat{\Psi}_4^2$ and the central χ_4^2 distribution under the null hypothesis are given in Figure 10.

Insert Figure 10 here.

It is fairly obvious that the smooth test statistic $\widehat{\Psi}_4^2$ has very good power properties overall. Further, with 2000 replications at 5% level of significance we observe the component tests.

6 Application to Asset Return on S&P 500 Index

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). The summary statistics of the data are given in Table 1. 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.

Insert Figure1 here.

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, 2013). In order to better fit the model for forecasting future observations, we use a naive MA(1) (Figure 3), followed by MA(1)-normal-GARCH(1,1) (Figure 4), MA(1)-GJR-GARCH(1,1), MA(1)-EGARCH(1,1), MA(1)-t-GARCH(1,1) (Figure 5), MA(1)-t-GJR-GARCH(1,1) and finally, MA(1)-t-EGARCH(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 (Figures 2, 3,4 and 5) 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).

Insert Figure 2 here.

Insert Figure 3 here.

Insert Figure 4 here.

Insert Figure 5 here.

Insert Table 1 here.

Insert Table 2 here.

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 2 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 3 and Table 4, we report the results of the smooth test and the augmented smooth test respectively

Insert Table 3 here

Insert Table 4 here

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 possible deviations in the directions of the second and fourth order polynomials that can be related to second and fourth moments (Neyman, 1937, Bera and Ghosh, 2001) One caveat to the above statement is that the normalized Legendre polynomials indicate that the second order term is present in the fourth order polynomial, hence it would be difficult to identify 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 17.9702), 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 16.5625 compared to the earlier value of 203.3362). Therefore, introduction of conditional heteroskedasticty into the forecast density model substantially improves its performance.

It has been noted that returns to financial indices are often conditionally asymmetric distribution and also reflect effects of significant "leverage effect." We introduce two special types of GARCH models GJR-GARCH (Glosten, Jagannathan and Runkle, 1993) and Nelson's EGARCH (Nelson, 1991) models that take account of the leverage effect and asymmetry simultaneously. Introduction of leverage effect term to the MA(1)-Normal GARCH model (often called the GJR-GARCH) does not make the significance of the overall of the overall $\hat{\Psi}^2$ change substantially, in fact in our sample it increases marginally ($\hat{\Psi}_4^2$ =17.9702 to $\hat{\Psi}_4^2$ =22.6123). However, the third moment now becomes more significant ($\hat{u}_3^2$ = 0.086 to $\hat{u}_3^2$ = 5.0333), which probably indicates that leverage effect does not play a very significant role in our sample period. We also fit the EGARCH model with normal errors that introduces significant asymmetry in the original GARCH model. The overall $\hat{\Psi}^2$ changes from 17.9702 for the MA(1)-GARCH(1,1) to $\hat{\Psi}^2$ = 75.075 in the EGARCH(1,1) model. This indicates that asymmetry in the form of the EGARCH(1,1) model is also not the "best" model for the density forecast with divergence in the direction of the second and fourth moments. It is worth noting that there is no effect in the direction of the third moment i.e., $\hat{u}_3^2 = 0.2219$ is not significant.

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 $\hat{\Psi}_k^2$ =1.6993, 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 do further investigate the effect of asymmetry and leverage effects in the model with GJR-GARCH and EGARCH models with Student's $t-$error. GJR-GARCH(1,1) is still an acceptable model with overall $\hat{\Psi}^2 = 8.3782$ is not statistically significant with the main departure coming from the third order polynomial $\hat{u}_3^2 = 5.6706$. Finally, for the smooth test that allows from Student's t errors the EGARCH specification does not seem to be an overall good fit ($\hat{\Psi}^2 = 17.9108$).

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 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 (see Bera, Ghosh and Xiao, 2013). 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 surmise that in a formal score type test the true null hypothesis would be rejected more frequently than the nominal 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 3, 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 16.5625 to statistically insignificant 0.0002 (see Table 3). 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 heteroskedasticty.

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.

We used the augmented smooth test to explicitly incorporate dependence into the model, the results of our tests are given in Table 4. It should be bourne in mind that the beauty of the smooth test technique in particular, and using the orthonormal polynomials in general, means that we do not have to recalculate the individual components $\hat{u}_i^2$. We just need to calculate the additional term we would call the correction term. The resulting test statistic is distributed as χ^2 with $k+q$ degrees of freedom where q is the number of moment conditions of dependence included. We have incorporated an explicit test for GARCH(1,1) type disturbance in the augmented smooth test by introducing a weighted ARCH framework (linear or $r-$weighting-Engle 1982, 1983; exponential or $\phi-$weighting-Bollerslev, 1986). Due to the existence of the non-regular Davies' problem we face, we have to look at a SupLM or AveLM type statistic and relevant method for calculating p-values through simulation suggested as proposed by Andrews and Ploberger (1996), If we use the linear weighting proposed first by Engle (1982), we avoid the severity of the Davies' problem as now the only issue is selection of the number of terms used r.

When we use the EDF, the overall augmented $\hat{\Psi}^2$ is strongly rejected due to ARCH effects using both exponential and linear weighting (Linear weighted Augmented $\hat{\Psi}^2$ =750.1082), while the exponential weighted adjustment term (aveLM statistic) 4773.425 which is strongly statistically significant. A naive MA(1) model is also strongly rejected (Linear weighted Augmented $\hat{\Psi}^2$ =394.9528). So far the results from the different weighting schemes has been similar, however, while a MA(1)-GARCH(1,1) seemed to be reasonable using a linear weighting scheme (Linear weighted ARCH score is 3.1251, p-value=0.0771), but the adjustment with exponential weights rejects the the model strongly. This could be an indication of a more complex dependence structure including stochastic volatility (Kim, Shephard and Chib, 1998).

We have also used component tests for leverage effects and a joint test for leverage effect and ARCH(1) terms in the probability intergral transform or generalized residuals (Table 4). When testing jointly, we find there is unaccounted for leverage effect in the naive EDF based (score term=2909.57, pvalue=0) and normal MA(1) models (score=61.424, p-value=4.59×10^{-014}). The joint augmented smooth test of leverage and ARCH(1) effects in the generalized residual strongly indicates that these affects have been unaccounted for in the MA(1)-GARCH (1,1) model (Joint test statistic=48.7539, p-value=2.59×10^{-011}). Including leverage effect term in the estimating model like GJR GARCH(1,1) reduces the overall augmented $\hat{\Psi}^2$, the exponential weighted squared ARCH lags is still significant but to a lesser degree ($score$ =12.0132, p-value=0) and the linear weighted squared lags is not statistically significant ($score$=2.2001, p-value=0.138). One surprising result is that leverage effect score in the gereneralized residuals by itself doesn't seem to be significant in either MA(1)-GARCH(1,1) or GJR-GARCH(1,1) models. This seemingly is an anomaly from the existing literature as GJR models are better at handling asymmetry and leverage effects although the former is probably a better model (see Bao, Lee and Saltoglu, 2007). MA(1) $-$EGARCH is also rejected strongly as the true model with linear weighted ARCH dependence (Augmented $\hat{\Psi}^2$ =.77.5246, p-value=2.78×10^{-015}). We also observe that the MA(1)-EGARCH(1,1) model doesn't seem to capture leverage effect completely, either individually (score=128.5709) or jointly with ARCH (1) type errors (score=45.6711). This might be an indication of unexplained and unmodeled volatility like a stochastic volatility model. As before, introduction of a conditional Student's $t-$distribution remarkably improves the performance almost thrughout the board with both exponential and linear weights with an overall augmnted $\hat{\Psi}_4^2 = 4.822$ (p-value=0.438) which is an acceptable "true" model. Accounting for leverage effect and asymmetry GJR-GARCH with Student's t-error is also an acceptable model although it doesn't fully capture the joint effect of leverage and ARCH(1) type errors (score=52.99, pvalue=3.11×10^{-012}). However, EGARCH with t-error does not seem to fit the data well, in particular, the effects of leverage effect although the linear weighted ARCH dependence is accounted for. Our results indicate that there could be possibility of a more involved volatility process that cannot be modeled in this framework.

7 Conclusion and Future 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 works-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. Although we are dealing with index returns, the proposed smooth test can be applied to other financial data for exchange rates, futures or forward markets, options prices, inflation rate, analyst forecasts among many others.

Acknowledgement: We would like to thank the participants at various conferences where this paper and its earlier versions were presented. We are thankful to Torben Anderson, Dick Van Dyk, Yongmiao Hong, Roberto Mariano, Adrian Pagan, Yiu Kuen Tse, Aman Ullah, Ken Wallis and Zhijie Xiao for helpful discussions and suggestions. However, any remaining errors are our responsibility. The first author would like to thank SMU-Wharton School Office of Research, Research Grant Number. 03-C208-SMU-025 for the financial support.

References

[1] Andrews, D.W.K., Ploberger, W., 1996. Testing for serial correlation against an ARMA(1,1) process. Journal of the American Statistical Association 91, 1331-1342.

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

[3] Bera, A. K., Ghosh, A., Xiao, Z., 2013. A Smooth test for the equality of distributions. Econometric Theory 29, 419-426.

[4] Bera, A. K., Ra., S., 1994. A test for white noise against ARMA (1,1). Working Paper, University of Illinois.

[5] Berkowitz, J., 2001. Testing density forecasts with applications to risk management. Journal of Business & Economic Statistics 19, 465-474.

[6] Bai, J., 2003, Testing parametric conditional distributions of dynamic models. Review of Economics and Statistics 85, 531-549.

[7] Bao, Y.,Lee, T.-H., Saltoglu, B., 2007. A Test for density forecast comparisons with applications to risk management. Journal of Forecasting 26, 203–225.

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

[9] Bontemps, C., Meddahi, N., 2005. Testing normality: A GMM approach. Journal of Econometrics 124, 149-186.

[10] Bontemps, C., Meddahi, N., 2012. Testing distributional assumptions: A GMM approach. Journal of Applied Econometrics 27, 978–1012.

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

[12] Breusch,T., 1978. Testing for autocorrelation in dynamic linear models. Australian Economic Papers 17, 334–355.

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

[14] Corradi, V., Swanson, N., 2006a. Predictive density and conditional confidence interval accuracy tests. Journal of Econometrics 135, 187–228.

[15] Corradi, V., Swanson N., 2006b. Predictive density evaluation. In: Elliott, G., Granger, C.W. J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, Elsevier: Amsterdam, 197-284.

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

[17] D'Agostino, R. B., Stephens, M. A., 1986. Goodness-of-Fit techniques. Marcel Dekker, New York.

[18] Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247-254.

[19] Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33-43.

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

[21] Diebold, F. X., Hahn, J., 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.

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

[23] Egorov, A., Hong, Y., Li, H., 2005. Validating forecasts of the joint probability density of bond yields: can affine models beat random walk? Journal of Econometrics 135, 255-284.

[24] Engle, R.F., 1982. Autoregressive conditional heteroskedasticty with estimates of the variance of UK inflation. Econometrica 50, 987-1008.

[25] Engle, R.F., 1983. Estimates of the variance of US inflation based upon the ARCH model. Journal of Money, Credit and Banking 15, 286-301.

[26] Glosten, L. R., Jahannathan, R. , Runkle, D.E., 1993. On the relationship between expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

[27] Granger, C.W.J., 1980. Testing for causality: A personal viewpoint. Journal of Economics and Dynamic Controls 2, 329-352.

[28] Granger, C.W.J, Pesaran, M.H., 2000. Economic and statistical measures of forecast accuracy. Journal of Forecasting 19, 537-560.

[29] Hong, Y., 2001. Evaluation of out-of-sample probability density forecasts with applications to stock prices, *Mimeo*, Department of Economics, Cornell University.

[30] Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Review of Financial Studies 18, 37-84.

[31] Hou, A. J, Suardi, S., 2011. Modelling and forecasting short-term interest rate volatility: A semiparametric approach. Journal of Empirical Finance 18, 692-710.

[32] Kim, S., Shephard, N., Chib S, 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361-393.

[33] Liu, X., Shackleton, M.B., Taylor, S.J., Xu, X., 2007. Closed-form transformations from risk-neutral to real-world distributions. Journal of Banking and Finance 31, 1501–1520.

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

[35] Morgan, J.P., 1996. RiskMetrics–Technical Document 4rd Edition. Morgan Guarranty Trust Company: New York.

[36] Nelson, D., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347-370.

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

[38] Paye, B., Timmermann, A., 2006. Instability of return prediction models. Journal of Empirical Finance 13, 274-315.

[39] 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.

[40] Pesaran, M.H., Timmermann, A., 2002. Market timing and return prediction under model instability. Journal of Empirical Finance, 9, 495-510.

[41] 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.

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

[43] Ross, S., 2015. The recovery theorem. The Journal of Finance 70, 615-648.

[44] Sarno, L., Valente, G., 2004. Comparing the accuracy of density forecasts from competing models. Journal of Forecasting 23, 541-557.

[45] Thomas, D. R., 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.

[46] Thompson, S., 2008. Identifying term structure volatility from the LIBOR-swap curve. Review of Financial Studies 21, 819-854.

[47] 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.

[48] Weigend, A., Shi., S.,2000. Predicting daily probability distributions of S&P500 returns. Journal of Forecasting 19, 375-392.

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 1. Summary Statistics for return distributions for estimation and test samples of S&P 500 returns for density forecast evaluation

	Test Statistic	Critical Values Upper .1%
D^+	4.19843	1.859
D^-	4.89182	1.859
KS	4.89182	1.95
CvM	10.62024	1.167
A-D	94.37819	6.0

Table 2. Goodness-of-Fit statistics based on EDF with $m = 8431$ and $n = 2016$, Critical values are from D'Agostino and Stephens (1986), shows statistically significant difference in distributions using Kolmogorov-Smirnov statistics (D^+,D^-, max(D^+,D^-)=KS), Cramer-von Mises Statistics (CvM) and Anderson-Darling statistics (A-D).

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)-	17.9702***	1.0806	16.5625***	0.086	0.2411
Normal GARCH (1,1)	(0.00125)	(0.29856)	(0.00005)	(0.76937)	(0.62339)
MA(1)-	22.6123***	1.68	14.8608***	5.0333**	1.0381
GJRGARCH (1,1)	(0.00015)	(0.19493)	(0.00012)	(0.02486)	(0.30826)
MA(1)-	73.075***	1.3444	27.557***	0.2219	43.9517
EGARCH(1,1)	5.107×10^{-15}	(0.24626)	(0.00000)	(0.63757)	(0.00000)
MA(1)-	1.6993	1.0727	0.0002	0.32275	0.3036
t_8 GARCH (1,1)	(0.79085)	(0.30034)	(0.9879)	(0.57)	(0.58164)
MA(1)-	8.3782	1.8233	0.0512	5.6706	0.8331
t_8 GJRGARCH (1,1)	(0.07867)	(0.17692)	(0.82097)	(0.01725)	(0.3614)
MA(1)-	17.9108***	0.9332	14.3319***	0.2252	2.4206
t_8 EGARCH (1,1)	(0.00128)	(0.33403)	(0.00015)	(0.63513)	(0.11975)

*** *significant at* 1% *level.*** *significant at* 5% *level.*

Table 3.Smooth statistics and components (p-values are in parenthesis). Column 1 shows the parametric model of the evaluation sample estimation, and density forecast of the test sample. If the model is correct, the sample statistic for column 2 would follow a Chi-squared with 4 degrees of freedom and each component in the following 4 columns will follow Chi-squared with 1 degree of freedom. If the model is not correct, then the overall smooth test will be rejected. The components will show the direction of departure is in which moment-direction.

Hypothesis	Aug. $\hat{\Psi}_4^2$	L. Effect	Lev-ARCH	$\phi-$ARCH	$r-$ARCH
EDF	750.1082*** (0.0000)	865.2929*** (0.0000)	2909.57*** (0.0000)	4773.425*** (0.0000)	141.8506*** (0.0000)
MA(1)	394.9528*** (0.0000)	476.265*** (0.0000)	61.42434*** (4.59×10^{-014})	1008.378*** (0.0000)	4.5796** (0.03235)
MA(1)-Normal GARCH(1,1)	21.0953*** (0.0008)	0.0234 (0.8783)	48.7539*** (2.59×10^{-011})	208.7223*** (0.0000)	3.1251 (0.0771)
MA(1)-GJR GARCH(1,1)	24.81235*** (0.0002)	0.7434 (0.3886)	52.9411*** (3.19×10^{-012})	12.0132*** (0.0000)	2.2001 (0.138)
MA(1)-EGARCH (1,1)	77.5246*** (2.78×10^{-015})	128.5709*** (0.0000)	45.6711*** (1.21×10^{-010})	180.1875*** (0.0000)	4.4496** (0.0349)
MA(1)-t_8GARCH (1,1)	4.822 (0.438)	34.1214 (5.18×10^{-009})	55.34966 (9.57×10^{-013})	2.1067 (0.232)	3.1228 (0.0772)
MA(1)-t_8GJR GARCH (1,1)	10.6468 (0.0589)	32.3637*** (1.28×10^{-008})	52.9959*** (3.11×10^{-012})	0.8905 (0.972)	2.2685 (0.132)
MA(1)-t_8 EGARCH (1,1)	20.3171*** (0.0011)	114.2006*** (0.0000)	55.8364*** (7.504×10^{-013})	103.7763*** (0.0000)	2.4063 (0.1209)

$^{***}significant\ at\ 1\%\ level.^{**}significant\ at\ 1\%\ level.$

Table 4.Neyman's smooth statistics with leverage effect and weighted ARCH type dependence (p-values in parenthesis). Col. 1 shows the distributional assumption of the density forecast. Overall smooth test in col. 2 gives the smooth test which is Chi-squared with 4 degrees of freedom under H_0. In column 3 and 4, we see the augmented smooth test statistic for leverage effect and joint leverage-ARCH effect, respectively with chisquared 1 d.f. Cols. 6 and 7 give augmented smooth test with multiplicative and linear weighted ARCH models, respectively.

Source ($r = 2000$)	$\widehat{\Psi}_4^2$	$\hat{u}_1^2$	$\hat{u}_2^2$	$\hat{u}_3^2$	$\hat{u}_4^2$	$\widehat{\Psi}_5^2(AR(1))$	$\widehat{\Psi}_6^2(joint)$	$\widehat{\Psi}_5^2(wtd.ARCH)$
Empirical Size	0.08	0.08	0.09	0.05	0.06	0.09	0.35	0.35
Source ($r = 2000$)	$\widehat{\Psi}_4^2$	$\hat{u}_1^2$	$\hat{u}_2^2$	$\hat{u}_3^2$	$\hat{u}_4^2$	$\widehat{\Psi}_5^2(AR(1))$	$\widehat{\Psi}_6^2(joint)$	$\widehat{\Psi}_5^2(wtd.ARCH)$
Empirical Power	0.92	0.16	0.9	0.05	0.55	0.9	0.97	1.0

Table 5: Size and Power Properties of the Unmodified and Augmented Smooth Test ($\alpha = 5\%,$)

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

Figure 2: Histogram for the probability integral transforms using EDF.

Figure 3: Histogram for the probability integral transform with MA(1)-normal model.

Figure 4: Histogram for the probability integral transform with MA(1)-normal GARCH(1,1) model.

Figure 5: Histogram for the probability integral transform with MA(1)-t-GARCH (1,1).

Figure 6: Histogram and distribution of $\hat{\Psi}_4^2$ under the null hypothesis.

Figure 7: Distribution of individual $\hat{u}^2$ under the null hypothesis.

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

Figure 9: Smooth Test Statistic under the Null (Size) ($m = 8000, n = 1000$)

Figure 10: Smooth test statistic under the Alternative (Power) ($m = 5000, n = 1000$)

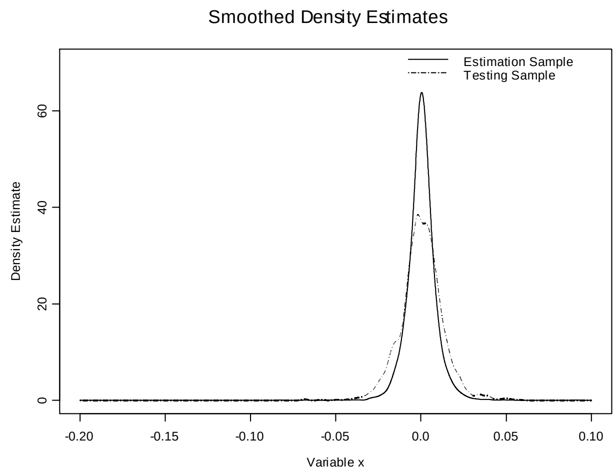

Figure 1: We consider the daily returns on the value-weighted S&P 500 Composite Index from July 3, 1962 to December 31, 2003. The sample is split into in-sample and out-of-sample periods for model estimation and density forecast evaluation. There are 8431 in-sample observations (07/03/62 - 12/29/95) and 2016 out-of-sample observations (01/02/96 - 12/31/2003). This graph shows the kernel density estimates of S&P 500 returns for estimation and test samples.

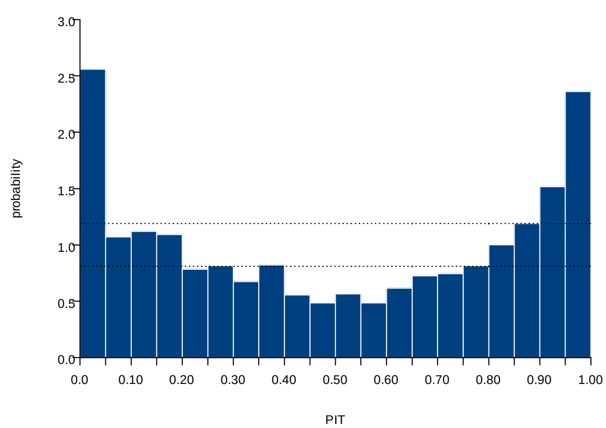

Figure 2: Histogram for the Probability Integral Transforms (PIT) using the Empirical Distribution Function (EDF). We consider the daily returns on the value-weighted S&P 500 Composite Index from July 3, 1962 to December 31, 2003. The sample is split into in-sample and out-of-sample periods for model 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).

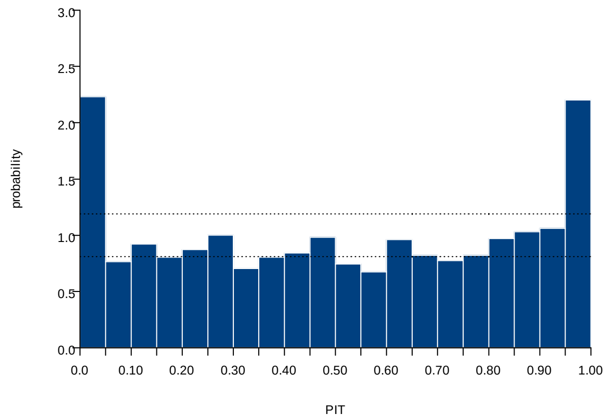

Figure 3: Histogram for the Probability Integral Transform (PIT) with naive MA(1)-normal model for estimation. We consider the daily returns on the value-weighted S&P 500 Composite Index from July 3, 1962 to December 31, 2003. The sample is split into in-sample and out-of-sample periods for model estimation 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).

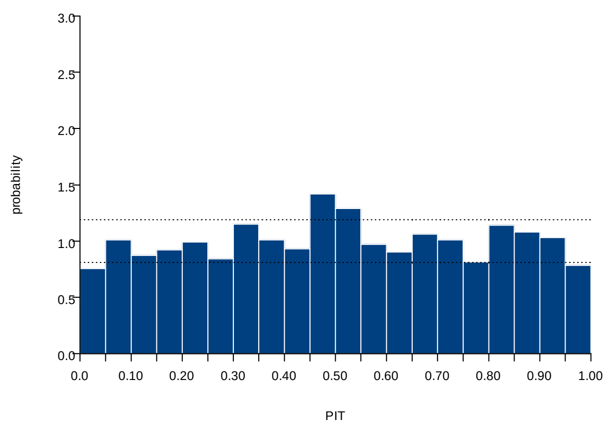

Figure 4: Histogram for the Probability Integral Transform (PIT) with MA(1)-normal GARCH (1,1) model for estimation. We consider the daily returns on the value-weighted S&P 500 Composite Index from July 3, 1962 to December 31, 2003. The sample is split into in-sample and out - of - sample periods for 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).

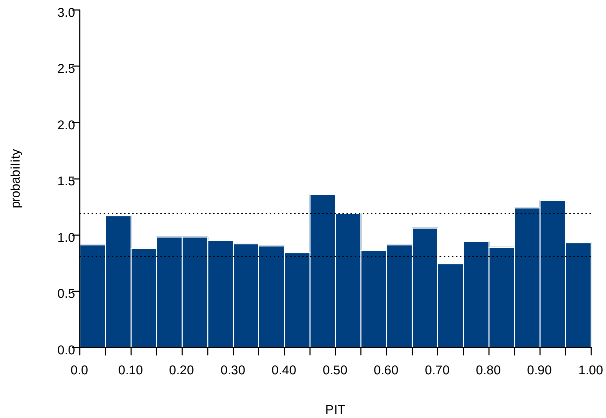

Figure 5: Histogram for the PIT with MA(1) - t - GARCH (1,1) for estimation. We consider the daily returns on the value-weighted S&P 500 Composite Index from July 3, 1962 to December 31, 2003. The sample is split into in-sample and out-of-sample periods for model estimation and density forecast evaluation. There are 8431 in-sample observations (07/03/62 - 12/29/95) and 2016 out-of-sample observations (01/02/96 - 12/31/2003).

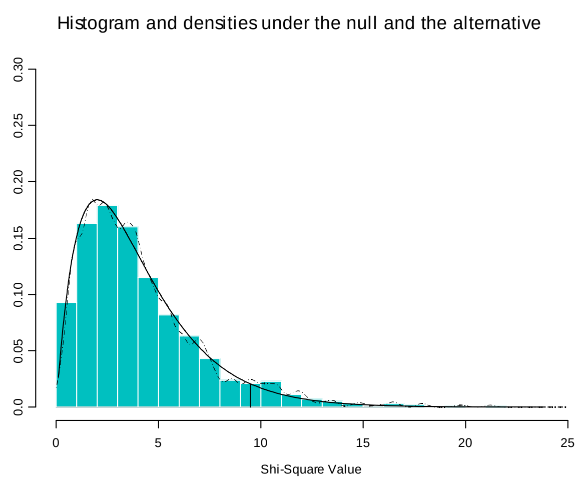

Figure 6: Histogram of Ψ_4^2, the overall smooth test statistics, under the null hypothesis that the estimation and the test distributions are the same. We also see distribution of the χ_4^2 distribution.

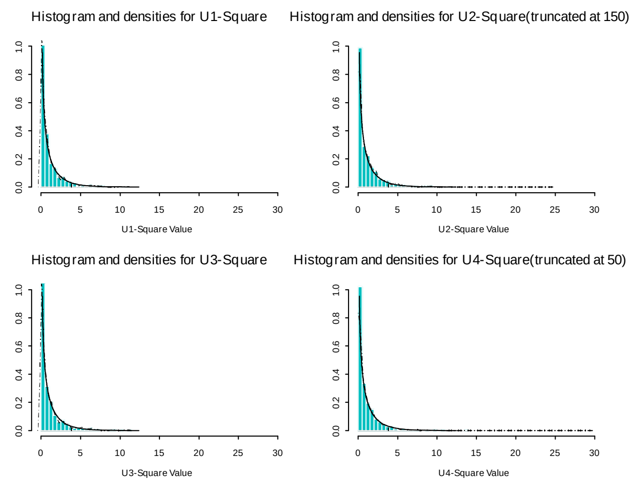

Figure 7: Histograms of $u_i^2,\ i = 1, ..., 4,$ the four components of the overall Ψ_4^2 statistics, under the null hypothesis that the estimation and the test distributions are the same. We also see distribution of the χ_1^2 distribution.

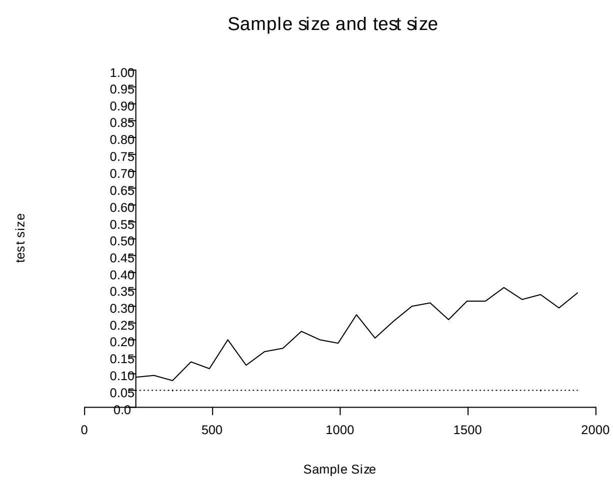

Figure 8: Empirical size of the the overall smooth test statistics $\Psi_4^2,$ under the null hypothesis that the estimation and the test distributions are the same. We plot of the size of the test as a function of n $(m = 2000)$.

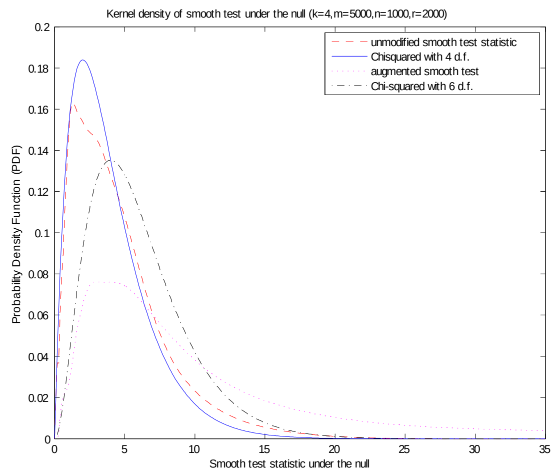

Figure 9: The empirical size of the overall modified and unmodified smooth test statistic Ψ_4^2 under the null hypothesis. We also plot the χ_4^2 and χ_1^2 charts. We use $m = 8000$ and $n = 1000$ for the simulation exercise.

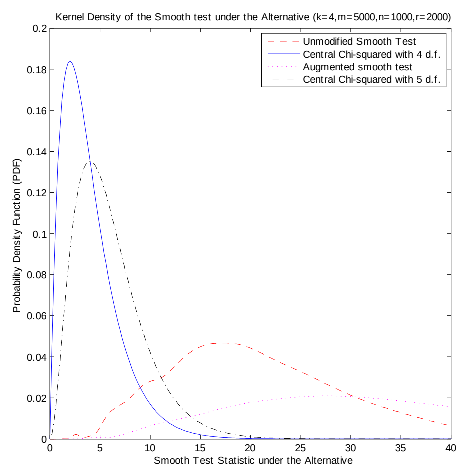

Figure 10: The empirical size/power properties of the overall modified and unmodified smooth test statistic Ψ_4^2 under the alternative hypothesis. We also plot the χ_4^2 and χ_1^2 charts. We choose estimation sample size $m = 5000$ and test sample size $n = 1000$.

Part I

Appendices:

APPENDIX A1A:

Proof. Taking (4) as the PDF under the alternative hypothesis, the log-likelihood function $l(\theta)$ can be written as

$$l(\theta) = n \ln c(\theta) + \sum_{j=1}^{k} \theta_j \sum_{i=1}^{n} \pi_j(y_i). \tag{49}$$

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

$$RS = s(\theta_0)' \mathcal{I}(\theta_0)^{-1} s(\theta_0), \tag{50}$$

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

$$\begin{aligned} s(\theta_j) &= \frac{\partial l(\theta)}{\partial \theta_j} \\ &= n\frac{\partial \ln c(\theta)}{\partial \theta_j} + \sqrt{n}u_j,\ \ j = 1, 2, ..., k, \\ \text{with } u_j &= \sum_{i=1}^{n} \pi_j(y_i)/\sqrt{n}. \end{aligned} \tag{51}$$

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

$$\frac{\partial c(\theta)}{\partial \theta_j} \int_0^1 \exp\left[\sum_{j=1}^{k} \theta_j \pi_j(y)\right] dy + c(\theta) \int_0^1 \exp\left[\sum_{j=1}^{k} \theta_j \pi_j(y)\right] \pi_j(y)\,dy = 0. \tag{52}$$

Evaluating (52) 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

$$s(\theta_j) = \sqrt{n}u_j. \tag{53}$$

To get the information matrix, let us first note from (51) that

$$\frac{\partial^2 l(\theta)}{\partial\theta_j \partial\theta_l} = n\frac{\partial^2 \ln c(\theta)}{\partial\theta_j \partial\theta_l}, \tag{54}$$

which is deterministic. 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 (52) with respect to θ_l and evaluating it at $\theta = \mathbf{0}$, after some simplification, we have

$$\left.\frac{\partial^2 c(\theta)}{\partial\theta_j \partial\theta_l}\right|_{\theta=0} + \int_0^1 \pi_j(y)\,\pi_l(y)\,dy = 0. \tag{55}$$

Using the orthonormal property in (5)

$$\left.\frac{\partial^2 c(\theta)}{\partial\theta_j \partial\theta_l}\right|_{\theta=0} = -\delta_{jl}. \tag{56}$$

Further, using (52), $c(\theta) = 1$ 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

$$\mathcal{I}(\theta_0) = nI_k, \tag{57}$$

where I_k is a $k \times k$ identity matrix. Combining (50), (53) and (57) the RS test statistic has the simple form

$$\Psi_k^2 = \sum_{j=1}^{k} u_j^2 = RS. \tag{58}$$

■

APPENDIX A1B (PROOF OF THEOREM 2)

Proof. From equations (8), (16) and (17)

$$\begin{aligned}\hat{\Psi}_k^2 - \Psi_k^2 &= \sum_{j=1}^{k} \frac{1}{n}\left[\left(\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\hat{\beta}\right)\right)\right)^2 - \left(\sum_{i=1}^{n} \pi_j\left(F\left(x_i;\beta\right)\right)\right)^2\right] \\ &= \sum_{j=1}^{k}\left[\hat{u}_j^2 - u_j^2\right]. \end{aligned} \tag{59}$$

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 $\left|\hat{\beta}-\beta\right| \geq |\beta^*-\beta|$.

$$\text{Hence, } \hat{u}_j^2 - u_j^2 = \frac{2}{n}\left(\hat{\beta}-\beta\right)\left[\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right)\right]\left[\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right]$$
$$= 2\left(\frac{n}{\sqrt{m}}\right)\left[\sqrt{m}\left(\hat{\beta}-\beta\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \times \left[\frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right]. \tag{60}$$

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

$$E\left(\pi_j\left(y_i\right)\right) = \int_0^1 \pi_j(u)\,du = 0. \tag{61}$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta\right)\right) \xrightarrow{p} E\left(\pi_j\left(y_i\right)\right) = 0. \tag{62}$$

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)

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

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

$$\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right) \xrightarrow{p} E\left[\pi_j\left(F\left(x_i;\beta^*\right)\right)\right]. \tag{63}$$

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. Since, $\pi_j\left(F\left(x;\beta\right)\right)$ is a continuous function of β at $\beta = \beta^*$, we have

$$E\left[\pi_j\left(F\left(x;\beta^*\right)\right)\right] \xrightarrow{p} E\left[\pi_j\left(F\left(x;\beta\right)\right)\right], \ j = 1, 2, ..., k. \tag{64}$$

Hence, as m and n go to infinity, using results in (61), (62), (63) and (64), we have

$$\begin{aligned} &\tfrac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right) \xrightarrow{p} E\left[\pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \xrightarrow{p} E\left[\pi_j\left(F\left(x;\beta\right)\right)\right] = 0, \\ &\text{i.e.,} \tfrac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right) = a_1 = o_p\left(1\right). \end{aligned} \tag{65}$$

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,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta} &\xrightarrow{p} E\left[\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \\ &\xrightarrow{p} E\left[\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] < \infty \\ \Rightarrow \frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta} &= a_2 = O_p\left(1\right). \end{aligned} \tag{66}$$

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

$$\sqrt{m}\left(\hat{\beta} - \beta\right) = a_3 = O_p\left(1\right). \tag{67}$$

Hence from equation (60) using the results in (65), (66) and (67), we obtain

$$\begin{aligned} \hat{u}_j^2 - u_j^2 &= 2\left(\frac{n}{\sqrt{m}}\right)\left[\sqrt{m}\left(\hat{\beta} - \beta\right)\right]\left[\frac{1}{n}\sum_{i=1}^{n}\pi_j\left(F\left(x_i;\beta^*\right)\right)\right] \\ &\times \left[\frac{1}{n}\sum_{i=1}^{n}\frac{d\pi_j\left(F\left(x_i;\beta^*\right)\right)}{d\beta}\right] \\ &= 2\frac{n}{\sqrt{m}}a_1 a_2 a_3 \\ &= \frac{n}{\sqrt{m}}o_p\left(1\right). \end{aligned} \tag{68}$$

From (59) using (68) for fixed k,

$$\hat{\Psi}_k^2 - \Psi_k^2 = \frac{n}{\sqrt{m}} o_p(1). \tag{69}$$

which proves Theorem 2. ■

APPENDIX A2 (Proof of Theorem 3)

7.1 Proof of Theorem 3

Proof. In order to test for uniformity and as well as for dependence, one would test $H_0 : \theta_1 = \theta_2 = ... = \theta_k = 0; \phi_1 = \phi_2 = ... = \phi_q = 0$ against the alternative $H_1 : \theta_j \neq 0$ for at least one j or $\phi_l \neq 0$ for at least one l. However, we have not specified the forms of the functions $\pi_j(.)$ and $\delta_l(.)$. The log-likelihood function is

$$\begin{aligned}\sum_{t=1}^{n} \ln\left(h\left(y_t|y_1, y_2, ..., y_{t-1}\right)\right) &= \sum_{t=1}^{n} \ln f\left(y_t, y_{t-1}\right) \\ &= \sum_{t=1}^{n} \ln c(\theta, \phi) + \sum_{t=1}^{n}\sum_{j=1}^{k} \theta_j \pi_j(y_t) + \sum_{t=1}^{n}\sum_{l=1}^{q} \phi_l \delta_l(y_t, y_{t-1}) \\ &= n \ln c(\theta, \phi) + \sum_{j=1}^{k} \theta_j \sum_{t=1}^{n} \pi_j(y_t) + \sum_{l=1}^{q} \phi_l \sum_{t=1}^{n} \delta_l(y_t, y_{t-1}) \\ &= \ln L = l, \ \ say. \end{aligned} \tag{70}$$

So, if we use $\theta = (\theta_1, \theta_2, ..., \theta_k)'$ and $\phi = (\phi_1, \phi_2, ..., \phi_q)'$ then under the null hypothesis H_0

$$\begin{aligned} \left.\frac{\partial l}{\partial \theta_j}\right|_{\theta=0,\phi=0} &= n \left.\frac{\partial \ln c(\theta, \phi)}{\partial \theta_j}\right|_{\theta=0,\phi=0} + \sum_{t=1}^{n} \pi_j(y_t) \\ \Rightarrow \frac{1}{\sqrt{n}} \left.\frac{\partial l}{\partial \theta_j}\right|_{\theta=0,\phi=0} &= \sqrt{n} \left.\frac{\partial \ln c(\theta, \phi)}{\partial \theta_j}\right|_{\theta=0,\phi=0} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_j(y_t). \end{aligned} \tag{71}$$

Similarly, we have

$$\begin{aligned} \left.\frac{\partial l}{\partial \phi_l}\right|_{\theta=0,\phi=0} &= n \left.\frac{\partial \ln c(\theta, \phi)}{\partial \phi_l}\right|_{\theta=0,\phi=0} + \sum_{t=1}^{n} \delta_l(y_t, y_{t-1}) \\ \Rightarrow \frac{1}{\sqrt{n}} \left.\frac{\partial l}{\partial \phi_l}\right|_{\theta=0,\phi=0} &= \sqrt{n} \left.\frac{\partial \ln c(\theta, \phi)}{\partial \phi_l}\right|_{\theta=0,\phi=0} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \delta_l(y_t, y_{t-1}). \end{aligned} \tag{72}$$

Further, if we take derivative twice and evaluate at $H_0 : \theta = 0, \phi = 0$, from (70)

$$\left.\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right|_{\theta=0,\phi=0} = n \left.\frac{\partial^2 c\left(\theta,\phi\right)}{\partial\theta_i\partial\theta_j}\right|_{\theta=0,\phi=0}, \tag{73}$$

$$\left.\frac{\partial^2 l}{\partial\phi_l\partial\theta_j}\right|_{\theta=0,\phi=0} = n \left.\frac{\partial^2 c\left(\theta,\phi\right)}{\partial\phi_l\partial\theta_j}\right|_{\theta=0,\phi=0}, \tag{74}$$

$$\left.\frac{\partial^2 l}{\partial\phi_i\partial\phi_l}\right|_{\theta=0,\phi=0} = n \left.\frac{\partial^2 c\left(\theta,\phi\right)}{\partial\phi_i\partial\phi_l}\right|_{\theta=0,\phi=0}. \tag{75}$$

Since (20) is a density function under H_1, we have for each value of y_{t-1}

$$c\left(\theta,\phi\right)\int_0^1 \exp\left[\sum_{j=1}^{k}\theta_j\pi_j\left(y_t\right)+\sum_{l=1}^{q}\phi_l\delta_l\left(y_t,y_{t-1}\right)\right]dy_t = 1. \tag{76}$$

Evaluating the identity in (76) at $\theta_j = 0$, $j = 1, ..., k$ and $\phi_l = 0$, $l = 1, ..., q$, $c\left(0,0\right) = 1$. Also, if we differentiate (76) and evaluate at $\theta = 0$, $\phi = 0$ the following results are obtained:[1]

$$(i) \qquad \left.\frac{\partial c\left(\theta,\phi\right)}{\partial\theta_j}\right|_{\theta=0,\phi=0} + c\left(0,0\right)\int_0^1 \pi_j\left(y_t\right)dy_t = 0$$

$$\Rightarrow \left.\frac{\partial c\left(\theta,\phi\right)}{\partial\theta_j}\right|_{\theta=0,\phi=0} = 0, \text{ since } \int_0^1 \pi_j\left(y_t\right)dy_t = 0,\ j \neq 0. \tag{77}$$

$$(ii) \qquad \left.\frac{\partial c\left(\theta,\phi\right)}{\partial\phi_l}\right|_{\theta=0,\phi=0} + c\left(0,0\right)\int_0^1 \delta_l\left(y_t,y_{t-1}\right)dy_t = 0$$

$$\Rightarrow \left.\frac{\partial c\left(\theta,\phi\right)}{\partial\phi_l}\right|_{\theta=0,\phi=0} = -\int_0^1 \delta_l\left(y_t,y_{t-1}\right)dy_t = 0. \tag{78}$$

[1] For (ii) we can choose δ_l appropriately to make $\int_0^1 \delta_l\left(y_t,y_{t-1}\right)dy_t = 0$, this can be achieved by using $\tilde{\delta}_l\left(y_t,y_{t-1}\right) = \delta_l\left(y_t,y_{t-1}\right) - \int_0^1 \delta_l\left(y_t,y_{t-1}\right)dy_t$ if indeed $\int_0^1 \delta_l\left(y_t,y_{t-1}\right)dy_t \neq 0$.

$$(iii) \qquad \frac{\partial^2 c(\theta,\phi)}{\partial\theta_i\partial\theta_j} + \frac{\partial c(\theta,\phi)}{\partial\theta_j}\int_0^1 \pi_i(y_t)\,dy_t +$$
$$\frac{\partial c(\theta,\phi)}{\partial\theta_i}\int_0^1 \pi_j(y_t)\,dy_t + c(\theta,\phi)\int_0^1 \pi_i(y_t)\,\pi_j(y_t)\,dy_t = 0$$
$$\Rightarrow c_{\theta_i\theta_j} + c_{\theta_j}.0 + c_{\theta_i}.0 + \int_0^1 \pi_i(y_t)\,\pi_j(y_t)\,dy_t = 0$$
$$\Rightarrow c_{\theta_i\theta_j} = -\delta_{ij}, \tag{79}$$

where $\delta_{ij} = 1$ if $i = j$; $\delta_{ij} = 0$ if $i \neq j$, $c_{\theta_i\theta_j} \equiv \left.\frac{\partial^2 c(\theta,\phi)}{\partial\theta_i\partial\theta_j}\right|_{\theta=0,\phi=0}$ and $c_{\theta_j} = \left.\frac{\partial c(\theta,\phi)}{\partial\theta_j}\right|_{\theta=0,\phi=0}$. Similarly, it can be shown that

$$(iv) \qquad \frac{\partial^2 c(\theta,\phi)}{\partial\phi_l\partial\theta_j} + \frac{\partial c(\theta,\phi)}{\partial\theta_j}\int_0^1 \delta_l(y_t, y_{t-1})\,dy_t +$$
$$\frac{\partial c(\theta,\phi)}{\partial\phi_l}\int_0^1 \pi_j(y_t)\,dy_t + c(\theta,\phi)\int_0^1 \pi_j(y_t)\,\delta_l(y_t, y_{t-1})\,dy_t = 0$$
$$\Rightarrow c_{\phi_l\theta_j} = -\int_0^1 \pi_j(y_t)\,\delta_l(y_t, y_{t-1})\,dy_t, \text{ where } c_{\phi_l\theta_j} = \left.\frac{\partial^2 c(\theta,\phi)}{\partial\phi_l\partial\theta_j}\right|_{\theta=0,\phi=0}. \tag{80}$$

Finally, using the same procedure we can obtain

$$(v) \qquad c_{\phi_i\phi_l} = \left.\frac{\partial^2 c(\theta,\phi)}{\partial\phi_i\partial\phi_l}\right|_{\theta=0,\phi=0} = -\int_0^1 \delta_i(y_t, y_{t-1})\,\delta_l(y_t, y_{t-1})\,dy_t. \tag{81}$$

Using $(i) - (v)$, the score functions under the null are given by

$$\frac{\partial l}{\partial\theta_j} = \sum_{t=1}^{n} \pi_j(y_t), \ j = 1, ..., k,$$
$$\frac{\partial l}{\partial\phi_l} = \sum_{t=1}^{n} \delta_l(y_t, y_{t-1}), \ l = 1, ..., q. \tag{82}$$

The information matrix under H_0, $\mathcal{I}$ is given by

$$\mathcal{I} = -\left.\begin{bmatrix} E\left[\frac{\partial^2 l}{\partial\theta\partial\theta'}\right] & E\left[\frac{\partial^2 l}{\partial\phi_l\partial\theta_j'}\right] \\ E\left[\frac{\partial^2 l}{\partial\phi_l\partial\theta_j'}\right]' & E\left[\frac{\partial^2 l}{\partial\phi_i\partial\phi_l}\right] \end{bmatrix}\right|_{\theta=0,\phi=0}, \tag{83}$$

where given I_k is the $k \times k$ identity matrix

$$E\left[-\frac{\partial^2 l}{\partial\theta\partial\theta'}\right] = nI_k,$$
$$E\left[-\frac{\partial^2 l}{\partial\phi_l\partial\theta_j'}\right] = n\left(E\left[\int_0^1 \pi_j\left(y_t\right)\delta_l\left(y_t,y_{t-1}\right)dy_t\right]\right)_{j=1,...,k;\ l=1,...,q}$$
$$= E\left[\pi\delta\right],$$
$$E\left[-\frac{\partial^2 l}{\partial\phi_i\partial\phi_l}\right] = n\left(E\left[\int_0^1 \delta_i\left(y_t,y_{t-1}\right)\delta_l\left(y_t,y_{t-1}\right)dy_t\right]\right)_{j=1,...,k;\ l=1,...,q}$$
$$= E\left[\delta\delta\right]. \tag{84}$$

So, using the well-known results of the Rao score test, defining $u_j = \frac{1}{\sqrt{n}}\sum_{t=1}^n \pi_j\left(y_t\right),$ $j = 1, ..., k$ and $v_l = \frac{1}{\sqrt{n}}\sum_{t=1}^n \delta_l\left(y_t, y_{t-1}\right), \ l = 1, ..., q,$

$$\left[\begin{array}{c}\sqrt{n}U\\ \sqrt{n}V\end{array}\right]'\left[\begin{array}{cc}nI_k & nE\left[\pi\delta\right]\\ nE\left[\pi\delta\right]' & nE\left[\delta\delta\right]\end{array}\right]^{-1}\left[\begin{array}{c}\sqrt{n}U\\ \sqrt{n}V\end{array}\right] \sim \chi^2_{k+q}\left(0\right)$$
$$\Rightarrow \left[\begin{array}{c}U\\ V\end{array}\right]'\left[\begin{array}{cc}I_k & E\left[\pi\delta\right]\\ E\left[\pi\delta\right]' & E\left[\delta\delta\right]\end{array}\right]^{-1}\left[\begin{array}{c}U\\ V\end{array}\right] \sim \chi^2_{k+q}\left(0\right), \tag{85}$$

where $U = \left(u_1, u_2, ..., u_k\right)'$, $V = \left(v_1, v_2, ..., v_q\right)'$ and $\chi^2_d\left(0\right)$ means a central χ^2 distribution with d degrees of freedom. Simplifying the notation further, and defining $B = E\left[\pi\delta\right]$, $D = E\left[\delta\delta\right]$,from results on block matrices we have

$$\left[\begin{array}{cc}I_k & B\\ B' & D\end{array}\right]^{-1} = \left[\begin{array}{cc}I_k + BEB' & -BE\\ -EB' & E\end{array}\right] \tag{86}$$

where $E = \left(D - B'B\right)^{-1}$. From (85) and (86),

$$\left[\begin{array}{c}U\\ V\end{array}\right]'\left[\begin{array}{cc}I_k & E\left[\pi\delta\right]\\ E\left[\pi\delta\right]' & E\left[\delta\delta\right]\end{array}\right]^{-1}\left[\begin{array}{c}U\\ V\end{array}\right]$$
$$= \left[\begin{array}{c}U'U + U'BEB'U - V'EB'U\\ -U'BEV + V'EV\end{array}\right] \overset{a}{\sim} \chi^2_{k+l}. \tag{87}$$

As E is non-singular there exists a non-singular matrix L, such that $E = LL'$. Substituting this in equation (87), we can rewrite as

$$\begin{aligned}
& U'BEB'U - V'EB'U - U'BEV + V'EV \\
& = U'BLL'B'U - V'LL'B'U - U'BLL'V + V'LL'V \\
& = (L'V)' L'V - (L'V)' L'B'U \\
& - (L'B'U)' L'V + (L'B'U)' L'B'U \\
& = (L'V - L'B'U)' (L'V - L'B'U) \\
& = (L' (V - B'U))' (L' (V - B'U)) \\
& = (V - B'U)' LL' (V - B'U) \\
& = (V - B'U)' E (V - B'U) \qquad (88)
\end{aligned}$$

From (87) this gives

$$U'U + (V - B'U)' E (V - B'U) \overset{a}{\sim} \chi^2_{k+l}. \qquad (89)$$

■

APPENDIX B (Illustrative: Examples of Weights)

7.2 Case 1: (Fixed ϕ) Test for Weighted Autoregressive Terms

In our usual formulation with $q = 1$, $\delta_1 (y_t, y_{t-1}, ..., y_1) = \pi_1 (y_t) \sum_{s=1}^{t-1} \phi^{s-1} y_{t-s}$, we can obtain

$$v_1 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_1 (y_t) \sum_{s=1}^{t-1} \phi^{s-1} y_{t-s}. \qquad (90)$$

as the score function related to δ_1. Furthermore, given the model in (35), $E (y_t) = \mu = 0$,

$$\begin{aligned}
E \left[\int_0^1 \pi_j (y_t) \delta_1 (y_t, y_{t-1}, ..., y_1) dy_t \right] &= E \left[\int_0^1 \pi_j (y_t) \pi_1 (y_t) \sum_{s=1}^{t-1} \phi^{s-1} y_{t-s} dy_t \right] \\
&= \begin{cases} E \left[\sum_{s=1}^{t-1} \phi^{s-1} y_{t-s} \right] & \text{if } j = 1 \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} \frac{1-\phi^{t-1}}{1-\phi} \mu = 0 & \text{if } j = 1 \\ 0 & \text{otherwise.} \end{cases} \qquad (91)
\end{aligned}$$

since under H_0, $E(y_{t-s}) = E(y_t) = \mu = 0$, for all s. Similarly,

$$E\left[\int_0^1 [\delta_1(y_t, y_{t-1}, ..., y_1)]^2 \, dy_t\right] = E\left[\int_0^1 \pi_1^2(y_t)\left(\sum_{s=1}^{t-1} \phi^{s-1} y_{t-s}\right)^2 dy_t\right]$$
$$= E\left[\sum_{s=1}^{t-1} \phi^{s-1} y_{t-s}\right]^2$$
$$= \frac{1-\phi^{2(t-1)}}{1-\phi^2}\sigma_\varepsilon^2 \tag{92}$$

since under H_0 all $y_t s$ are independent and $E\left(y_{t-s}^2\right) = E\left(y_t^2\right) = \sigma_\varepsilon^2$. Hence, the asymptotic information matrix is given by

$$\mathcal{I} = \begin{bmatrix} 1 & \mathbf{0}' & 0 \\ \mathbf{0} & I_{k-1} & \mathbf{0} \\ 0 & \mathbf{0}' & \frac{\sigma_\varepsilon^2}{1-\phi^2} \end{bmatrix} = \begin{bmatrix} I_k & B \\ B' & D \end{bmatrix}. \tag{93}$$

Using the same notations as before we obtain the following results:

$(i)\, D - B'B = \frac{\sigma_\varepsilon^2}{1-\phi^2} - \left(\frac{0}{1-\phi}\right)^2 = \frac{\sigma_\varepsilon^2}{(1-\phi^2)} \Rightarrow E = (D - B'B)^{-1} = \frac{\left(1-\phi^2\right)}{\sigma_\varepsilon^2}$.

$(ii)\, U'BEB'U = \left(\frac{\mu^2 u_1^2}{(1-\phi)^2}\right) E = \frac{(1+\phi)\mu^2 u_1^2}{\sigma_\varepsilon^2(1-\phi)-2\phi\mu^2} = 0$.

$(iii)\, U'BEV = \left(\frac{\mu u_1 v_1}{(1-\phi)}\right) E = \frac{(1-\phi)(1+\phi)\mu u_1 v_1}{\sigma_\varepsilon^2(1-\phi)-2\phi\mu^2} = 0$.

$(iv)\, V'EV = v_1^2 E = \frac{\left(1-\phi^2\right) v_1^2}{\sigma_\varepsilon^2}$.

Hence, if ϕ is a known constant,

$$\sum_{j=1}^k u_j^2 + \left[\left(\frac{\mu u_1}{(1-\phi)}\right)^2 - 2\left(\frac{\mu u_1 v_1}{(1-\phi)}\right) + v_1^2\right] E$$
$$= \sum_{j=1}^k u_j^2 + v_1^2 \frac{(1-\phi^2)}{\sigma_\varepsilon^2}$$
$$\sim \chi_{k+1}^2. \tag{94}$$

7.3 Case 2: (Fixed ϕ) Test for Weighted Autregressive Conditional Heteroscdasticty Terms (GARCH(1,1))

$$E\left[\int_0^1 \pi_j(y_t)\,\delta_1(y_t,y_{t-1},...,y_1)\,dy_t\right] = E\left[\int_0^1 \pi_j(y_t)\,\pi_2(y_t)\sum_{s=1}^{t-1}\omega_1^{s-1}y_{t-s}^2 dy_t\right]$$
$$= \begin{cases} E\left[\sum_{s=1}^{t-1}\omega_1^{s-1}y_{t-s}^2\right] & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases}$$
$$= \begin{cases} \frac{1-\delta_1^{t-1}}{1-\omega_1}E\left(y_t^2\right) & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \tag{95}$$

since under H_0, $E\left(y_{t-s}^2\right) = E\left(y_t^2\right)$, for all s. Similarly,

$$E\left[\int_0^1 \left[\delta_1(y_t,y_{t-1},...,y_1)\right]^2 dy_t\right] = E\left[\int_0^1 \pi_2^2(y_t)\left(\sum_{s=1}^{t-1}\omega_1^{s-1}y_{t-s}^2\right)^2 dy_t\right]$$
$$= \begin{cases} E\left[\sum_{s=1}^{t-1}\omega_1^{s-1}y_{t-s}^2\right]^2 & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases}$$
$$= \begin{cases} \frac{1-\omega_1^{2(t-1)}}{1-\omega_1^2}E\left(y_t^4\right) & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \tag{96}$$

since under H_0 all y_ts are independent and $E\left(y_{t-s}^4\right) = E\left(y_t^4\right)$. Hence, the asymptotic information matrix is given by

$$\mathcal{I} = \begin{bmatrix} 1 & 0 & \mathbf{0}' & 0 \\ 0 & 1 & 0 & \frac{E\left(y_t^2\right)}{1-\omega_1} \\ \mathbf{0} & 0 & I_{k-1} & \mathbf{0} \\ 0 & \frac{E\left(y_t^2\right)}{1-\omega_1} & \mathbf{0}' & \frac{E\left(y_t^4\right)}{1-\omega_1^2}\end{bmatrix} = \begin{bmatrix} I_k & B \\ B' & D\end{bmatrix}. \tag{97}$$

Using the same notations as before we obtain the following results:

$(i)\; D - B'B = \frac{E\left(y_t^4\right)}{1-\omega_1^2} - \left(\frac{E\left(y_t^2\right)}{1-\omega_1}\right)^2 \Rightarrow E = \left(D - B'B\right)^{-1}.$

$(ii)\; U'BEB'U = \left(\frac{E\left(y_t^2\right)u_2}{1-\omega_1}\right)^2 E.$

$(iii)\; U'BEV = \left(\frac{E\left(y_t^2\right)u_2v_2}{1-\omega_1}\right) E.$

$(iv)\; V'EV = v_1^2 E.$

Hence, if ω_1 is a known constant,

$$\begin{aligned}
&\sum_{j=1}^{k} u_j^2 + \left[\left(\frac{E\left(y_t^2\right)u_2}{1-\omega_1}\right)^2 - 2\left(\frac{E\left(y_t^2\right)u_2 v_2}{1-\omega_1}\right) + v_1^2\right] E \\
&= \sum_{j=1}^{k} u_j^2 + \left(v_1 - \frac{E\left(y_t^2\right)u_2}{1-\omega_1}\right)^2 \left(\frac{E\left(y_t^4\right)}{1-\omega_1^2} - \left(\frac{E\left(y_t^2\right)}{1-\omega_1}\right)^2\right)^{-1} \\
&\sim \chi^2_{k+1}. \qquad (98)
\end{aligned}$$

7.4 Case 3: Weighted ARCH Model

The log-likelihood function is

$$\begin{aligned}
L &= Const - \frac{1}{2}\sum_{t=1}^{n} \ln h_t - \frac{1}{2}\sum_{t=1}^{n} \frac{u_t^2}{h_t} \\
&= Const - \frac{1}{2}\sum_{t=1}^{n} \ln\left[\alpha_0 + \alpha_1 \sum_{s=1}^{r} w_s u_{t-s}^2\right] - \frac{1}{2}\sum_{t=1}^{n} \frac{u_t^2}{\alpha_0 + \alpha_1 \sum_{s=1}^{r} w_s u_{t-s}^2}. \qquad (99)
\end{aligned}$$

Differentiating (99) with respect to α_1,

$$\begin{aligned}
\frac{\partial L}{\partial \alpha_1} &= -\frac{1}{2}\sum_{t=1}^{n} \frac{\left[\sum_{s=1}^{r} w_s u_{t-s}^2\right]}{\left[\alpha_0 + \alpha_1 \sum_{s=1}^{r} w_s u_{t-s}^2\right]} + \frac{1}{2}\sum_{t=1}^{n} \frac{u_t^2\left[\sum_{s=1}^{r} w_s u_{t-s}^2\right]}{\left[\alpha_0 + \alpha_1 \sum_{s=1}^{r} w_s u_{t-s}^2\right]^2} \\
\Rightarrow \left.\frac{\partial L}{\partial \alpha_1}\right|_{H_0} &= \frac{1}{2\alpha_0^2}\sum_{t=1}^{n} \tilde{u}_t^2 \sum_{s=1}^{r} w_s u_{t-s}^2, \text{ where } \tilde{u}_t^2 = u_t^2 - \alpha_0. \qquad (100)
\end{aligned}$$

From (100) differentiating again

$$\begin{aligned}
\frac{\partial^2 L}{\partial \alpha_1^2} &= \frac{1}{2}\sum_{t=1}^{n} \frac{\left[\sum_{s=1}^{r} w_s u_{t-s}^2\right]^2}{\left[\alpha_0 + \alpha_1 \sum_{s=1}^{r} w_s u_{t-s}^2\right]^2} - \frac{2}{2}\sum_{t=1}^{n} \frac{u_t^2\left[\sum_{s=1}^{r} w_s u_{t-s}^2\right]^2}{\left[\alpha_0 + \alpha_1 \sum_{s=1}^{r} w_s u_{t-s}^2\right]^3} \\
\Rightarrow -\left.\frac{\partial^2 L}{\partial \alpha_1^2}\right|_{H_0} &= \frac{1}{2\alpha_0^3}\sum_{t=1}^{n} \left[2u_t^2 - \alpha_0\right]\left[\sum_{s=1}^{r} w_s u_{t-s}^2\right]^2. \qquad (101)
\end{aligned}$$

Now taking expectation of (101) as $n \to \infty$ the asymptotic information matrix is

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} E\left[- \left.\frac{\partial^2 L}{\partial \alpha_1^2}\right|_{H_0} \right] &= \lim_{n\to\infty} \frac{1}{n} \frac{\alpha_0}{2\alpha_0^3} \sum_{t=1}^{n} E\left[\sum_{s=1}^{r} w_s u_{t-s}^2 \right]^2 \\
&= \frac{1}{2\alpha_0^2} \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{r} w_s^2 E\left[u_{t-s}^4\right], \mathit{IID} \text{ under } H_0 \\
&= \frac{1}{2\alpha_0^2} E\left[u_t^4\right] \sum_{s=1}^{r} \frac{\left[(r+1)-s\right]^2}{\left[\frac{1}{2} r (r+1)\right]^2} \\
&= \frac{1}{2\alpha_0^2} E\left[u_t^4\right] \frac{\frac{1}{6} r (r+1)(2r+1)}{\left[\frac{1}{2} r (r+1)\right]^2} \\
&= \frac{1}{2\alpha_0^2} E\left[u_t^4\right] \frac{2}{3} \frac{(2r+1)}{r(r+1)} \\
&= \frac{(2r+1)}{r(r+1)} \text{ putting } E\left[u_t^4\right] = 3\alpha_0^2. \qquad (102)
\end{aligned}$$

Hence, the Rao Score Statistic is

$$RS = n^{-1} \left[\frac{1}{2\hat{\alpha}_0^2} \sum_{t=1}^{n} \tilde{u}_t^2 \sum_{s=1}^{r} w_s u_{t-s}^2 \right]^2 \frac{r(r+1)}{(2r+1)} = n \frac{r(r+1)}{(2r+1)} \frac{\left[\sum_{t=1}^{n} \tilde{u}_t^2 \sum_{s=1}^{r} w_s u_{t-s}^2\right]^2}{4\left[\sum_{t=1}^{n} u_t^4\right]^2} \qquad (103)$$

which for testing for ARCH(1) becomes

$$RS = \frac{n}{12} \frac{\left[\sum_{t=1}^{n} \tilde{u}_t^2 u_{t-1}^2\right]^2}{\left[\sum_{t=1}^{n} u_t^4\right]^2}. \qquad (104)$$

Now let us setup the augmented Neyman Smooth test for incorporating several ARCH effects using a linear weighting scheme suggested by Engle (1982, 1983). We choose

$$\begin{aligned}
\delta_1 (y_t, y_{t-1}, ..., y_1) = \pi_2 (y_t) \sum_{s=1}^{r} w_s y_{t-s}^2, \text{ where } w_s = \frac{(r+1)-s}{\frac{1}{2} r (r+1)} \\
\Rightarrow v_1 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi_2 (y_t) \sum_{s=1}^{r} w_s y_{t-s}^2 \text{ is the score function.} \qquad (105)
\end{aligned}$$

Furthermore,

$$
\begin{aligned}
E\left[\int_0^1 \pi_j(y_t)\,\delta_1(y_t, y_{t-1}, ..., y_1)\,dy_t\right] &= E\left[\int_0^1 \pi_j(y_t)\,\pi_2(y_t)\sum_{s=1}^{r} w_s y_{t-s}^2 dy_t\right] \\
&= \begin{cases} E\left[\sum_{s=1}^{r} w_s y_{t-s}^2\right] & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \\
&= \begin{cases} E\left(y_t^2\right)\sum_{s=1}^{r} w_s & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \\
&= \begin{cases} E\left(y_t^2\right) & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \qquad (106)
\end{aligned}
$$

since under H_0, $E\left(y_{t-s}^2\right) = E\left(y_t^2\right)$, for all s. Similarly,

$$
\begin{aligned}
E\left[\int_0^1 \left[\delta_1(y_t, y_{t-1}, ..., y_1)\right]^2 dy_t\right] &= E\left[\int_0^1 \pi_2^2(y_t)\left(\sum_{s=1}^{r} w_s y_{t-s}^2\right)^2 dy_t\right] \\
&= \begin{cases} E\left[\sum_{s=1}^{r} w_s y_{t-s}^2\right]^2 & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \\
&= \begin{cases} \sum_{s=1}^{r} w_s^2 E\left(y_t^4\right) & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \\
&= \begin{cases} E\left(y_t^4\right)\frac{2}{3}\frac{2r+1}{r(r+1)} & \text{if } j=2 \\ 0 & \text{otherwise.}\end{cases} \qquad (107)
\end{aligned}
$$

since under H_0 all $y_t s$ are independent and $E\left(y_{t-s}^4\right) = E\left(y_t^4\right)$. Hence, the asymptotic information matrix is given by

$$
\mathcal{I} = \begin{bmatrix} 1 & 0 & \mathbf{0}' & 0 \\ 0 & 1 & 0 & E\left(y_t^2\right) \\ \mathbf{0} & 0 & I_{k-1} & \mathbf{0} \\ 0 & E\left(y_t^2\right) & \mathbf{0}' & \frac{2}{3}\frac{(2r+1)}{r(r+1)}E\left(y_t^4\right) \end{bmatrix} = \begin{bmatrix} I_k & B \\ B' & D \end{bmatrix}. \qquad (108)
$$

Using the same notations as before we obtain the following results:

$(i)\ D - B'B = \frac{2}{3}\frac{(2r+1)}{r(r+1)}E\left(y_t^4\right) - \left(E\left(y_t^2\right)\right)^2 \Rightarrow E = \left(D - B'B\right)^{-1}.$

$(ii)\ U'BEB'U = \left(E\left(y_t^2\right)u_2\right)^2 E.$

$(iii)\ U'BEV = \left(E\left(y_t^2\right)u_2 v_2\right)E.$

$(iv)\ V'EV = v_1^2 E.$

Hence, if r is a known constant,

$$\begin{aligned}
&\sum_{j=1}^{k} u_j^2 + \left[\left(E\left(y_t^2\right)u_2\right)^2 - 2\left(E\left(y_t^2\right)u_2v_2\right) + v_1^2\right] E \\
&= \sum_{j=1}^{k} u_j^2 + \left(v_1 - E\left(y_t^2\right)u_2\right)^2 \left(\frac{2}{3}\frac{(2r+1)}{r\,(r+1)}E\left(y_t^4\right) - \left(E\left(y_t^2\right)\right)^2\right)^{-1} \\
&\sim \chi^2_{k+1}. \qquad (109)
\end{aligned}$$