Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models With Student t Innovations

Gabriele Fiorentini
Università di Firenze
Viale Morgagni 59, I-50134 Firenze, Italy
<fiorentini@ds.unifi.it>

Enrique Sentana
CEMFI
Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

Giorgio Calzolari
Università di Firenze
Viale Morgagni 59, I-50134 Firenze, Italy
<calzolar@ds.unifi.it>

July 2000
Revised: January 2003

Abstract

We provide numerically reliable analytical expressions for the score, Hessian, and information matrix of conditionally heteroskedastic dynamic regression models when the conditional distribution is multivariate t. We also derive one-sided and two-sided Lagrange Multiplier tests for multivariate normality versus multivariate t based on the first two moments of the squared norm of the standardised innovations evaluated at the Gaussian pseudo-maximum likelihood estimators of the conditional mean and variance parameters. Finally, we illustrate our techniques through both Monte Carlo simulations, and an empirical application to 26 U.K. sectorial stock returns, which confirms that their conditional distribution has fat tails.

KEY WORDS: Financial Returns, Inequality Constraints, Kurtosis, Normality Test, Value at Risk, Volatility.

JEL: C51, C52, C22, C32

1 INTRODUCTION

Many empirical studies with financial time series data indicate that the distribution of asset returns is usually rather leptokurtic, even after controlling for volatility clustering effects (see e.g. Bollerslev, Chou and Kroner 1992 for a survey). This has been long realised, and two main alternative inference approaches have been proposed. The first one uses a "robust" estimation method, such as the Gaussian pseudo-maximum likelihood procedure advocated by Bollerslev and Wooldridge (1992), which remains consistent for the parameters of the conditional mean and variance functions even if the assumption of conditional normality is violated. The second one, in contrast, specifies a parametric leptokurtic distribution for the standardised innovations, such as the Student t distribution employed by Bollerslev (1987). While the second procedure will often yield more efficient estimators than the first if the assumed conditional distribution is correct, it has the disadvantage that it may end up sacrificing consistency when it is not (Newey and Steigerwald 1997). Nevertheless, a non-Gaussian distribution may be indispensable when we are interested in features of the distribution of asset returns, such as its quantiles, which go beyond its conditional mean and variance. For instance, empirical researchers and financial market practitioners are often interested in the so-called Value at Risk of an asset, which is the positive threshold value V such that the probability of the asset suffering a reduction in wealth larger than V equals some pre-specified level $\varkappa < 1/2$. Similarly, in the context of multiple financial assets, one may be interested in the probability of the joint occurrence of several extreme events, which is regularly underestimated by the multivariate normal distribution, especially in larger dimensions.

Notwithstanding such considerations, a significant advantage of the pseudo-maximum likelihood approach in Bollerslev and Wooldridge (1992) is that they derived convenient closed-form expressions for the Gaussian log-likelihood score and the conditional information matrix, which can be used to obtain numerically accurate extrema of the objective function, as well as reliable standard errors. In contrast, estimation under an alternative distribution typically relies on numerical approximations to the derivatives, which are often poor. One of the objectives of our paper is to partly close the gap between the two approaches by providing numerically reliable analytical expressions for the score vector, the Hessian matrix and its expected value for the multivariate conditionally heteroskedastic dynamic regression model considered by Bollerslev and Wooldridge (1992) when the distribution of the innovations is assumed to be proportional to a multivariate t. As is well known, the multivariate t distribution nests the normal as a limiting case, but the marginal distributions of its components have generally fatter tails, and it also allows for cross-sectional "tail dependence". In this respect, our results generalise the expressions in appendix B of Lange, Little and Taylor (1989), who only analysed an independent and identically distributed set up in which there is separation between unconditional mean and variance parameters.

We also use our analytical expressions to develop a test for multivariate normality when a dynamic model for the conditional mean and variance is fully specified, but the model is estimated under the Gaussianity null. We compare our proposed test with the kurtosis component of Mardia's (1970) test for multivariate normality, which reduces to the well-known Jarque and Bera (1980) test in univariate contexts. Importantly, we take into account the one-sided nature of the alternative hypothesis to derive the more powerful Kuhn-Tucker multiplier test, which is asymptotically equivalent to the Likelihood Ratio and Wald tests.

The rest of the paper is organised as follows. First, we obtain closed-form expressions for the log-likelihood score vector, the Hessian matrix and its conditional expected value in Section 2. Then, in Section 3, we introduce our proposed test for multivariate normality, and relate it to the existing literature. A Monte Carlo evaluation of different parameter and standard error estimation procedures can be found in Section 4. Finally, we include an illustrative empirical application to 26 U.K. sectorial stock returns in Section 5, followed by our conclusions. Proofs and auxiliary results are gathered in appendices.

2 MAXIMUM LIKELIHOOD ESTIMATION

2.1 The model

In a multivariate dynamic regression model with time-varying variances and covariances, the vector of N dependent variables, $\mathbf{y}_t$, is typically assumed to be generated by the following equations:

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu}_t(\boldsymbol{\theta}_0) + \boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*, \\
\boldsymbol{\mu}_t(\boldsymbol{\theta}) &= \boldsymbol{\mu}(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}), \\
\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}(\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}),
\end{aligned}$$

where $\boldsymbol{\mu}()$ and $vech\,[\boldsymbol{\Sigma}()]$ are N and $N(N+1)/2$-dimensional vectors of functions known up to the $p\times 1$ vector of true parameter values $\boldsymbol{\theta}_0$, $\mathbf{z}_t$ are k contemporaneous conditioning variables, I_{t-1} denotes the information set available at $t-1$, which contains past values of $\mathbf{y}_t$ and $\mathbf{z}_t$, $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})$ is an $N\times N$ "square root" matrix such that $\boldsymbol{\Sigma}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{1/2\prime}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t^*$ is a vector martingale difference sequence satisfying $E(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{0}$

and $V(\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) = \mathbf{I}_N$. As a consequence,

$$\begin{array}{rcl} E(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) & = & \boldsymbol{\mu}_t(\boldsymbol{\theta}_0), \\ V(\mathbf{y}_t|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0) & = & \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0). \end{array}$$

As in Bollerslev (1987) in a univariate context, and Harvey, Ruiz and Sentana (1992) in a multivariate one, followed by many others, our approach is based on the t distribution. In particular, we assume hereinafter that conditional on $\mathbf{z}_t$ and I_{t-1}, $\boldsymbol{\varepsilon}_t^*$ is independent and identically distributed as a standardised multivariate t with ν_0 degrees of freedom, or *i.i.d.* $t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ for short. That is,

$$\boldsymbol{\varepsilon}_t^* = \sqrt{\frac{(\nu_0 - 2)\,\zeta_t}{\xi_t}}\mathbf{u}_t,$$

where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_t is a chi-square random variable with N degrees of freedom, ξ_t is a Gamma variate with mean $\nu_0 > 2$ and variance $2\nu_0$, and $\mathbf{u}_t$, ζ_t and ξ_t are mutually independent (see Appendix A). As is well known, the multivariate Student t approaches the multivariate normal as $\nu_0 \to \infty$, but has generally fatter tails. For that reason, it is often more convenient to use the reciprocal of the degrees of freedom parameter, $\eta_0 = 1/\nu_0$, as a measure of tail thickness, which will always remain in the finite range $0 \leq \eta_0 < 1/2$ under our assumptions.

2.2 The log-likelihood function

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}',\eta)'$ denote the $p+1$ parameters of interest, which we assume variation free for simplicity (cf. expression (53) in Harvey et al. 1992). The log-likelihood function of a sample of size T (ignoring initial conditions) takes the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^T l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = c(\eta) + d_t(\boldsymbol{\theta}) + g\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]$:

$$\begin{aligned} c(\eta) = \ln\left[\Gamma\left(\frac{N\eta + 1}{2\eta}\right)\right] - \ln\left[\Gamma\left(\frac{1}{2\eta}\right)\right] \\ - \frac{N}{2}\ln\left(\frac{1-2\eta}{\eta}\right) - \frac{N}{2}\ln\pi, \end{aligned}$$

$$d_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln\left|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right|,$$

and

$$g\left[\varsigma_t(\boldsymbol{\theta}), \eta\right] = -\left(\frac{N\eta + 1}{2\eta}\right)\ln\left[1 + \frac{\eta}{1-2\eta}\varsigma_t(\boldsymbol{\theta})\right],$$

where $\Gamma()$ is Euler's gamma (or generalised factorial) function, $\varsigma_t(\boldsymbol{\theta})=\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})$, $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$, and $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})=\mathbf{y}_t-\boldsymbol{\mu}_t(\boldsymbol{\theta})$. Nevertheless, it is important to stress that since both $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are often recursively defined (as in ARMA or GARCH models), it may be necessary to choose some initial values to start up the recursions. As pointed out by Fiorentini, Calzolari and Panattoni (1996), this fact should be taken into account in computing the analytic score, in order to make the results exactly comparable with those obtained by using numerical derivatives. Not surprisingly, it can be readily verified that $L_T(\boldsymbol{\theta}, 0)$ collapses to a conditionally Gaussian log-likelihood.

Given the nonlinear nature of the model, a numerical optimisation procedure is usually required to obtain maximum likelihood (ML) estimates of $\boldsymbol{\phi}$, $\hat{\boldsymbol{\phi}}_T$ say. Assuming that all the elements of $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ are twice continuously differentiable functions of $\boldsymbol{\theta}$, we can use a standard gradient method in which the first derivatives are numerically approximated by reevaluating $L_T(\boldsymbol{\phi})$ with each parameter in turn shifted by a small amount, with an analogous procedure for the second derivatives. Unfortunately, such numerical derivatives are sometimes unstable, and moreover, their values may be rather sensitive to the size of the finite increments used. This is particularly true in our case, because even if the sample size T is large, the Student t-based log-likelihood function is often rather flat for small values of η. As we shall show in the next subsections, though, in this case it is also possible to obtain simple analytical expressions for the score vector and Hessian matrix. The use of analytical derivatives in the estimation routine, as opposed to their numerical counterparts, should considerably improve the accuracy of the resulting estimates (McCullough and Vinod 1999). Moreover, a fast and numerically reliable procedure for the computation of the score for any value of η is of paramount importance in the implementation of the score-based indirect inference procedures introduced by Gallant and Tauchen (1996) (see Calzolari, Fiorentini and Sentana 2003 for an application to a discrete-time, stochastic volatility model).

The analytic derivatives that we shall obtain could also be used even if the coefficients of the model were reparametrised as $\boldsymbol{\phi} = f(\boldsymbol{\varphi})$, with $\boldsymbol{\varphi}$ unconstrained, in order to maximise the unrestricted log-likelihood function $\mathfrak{L}_T(\boldsymbol{\varphi}) = L_T[f(\boldsymbol{\varphi})]$. In particular, by virtue of the chain rule for Jacobian matrices (Magnus and Neudecker 1988, thm. 5.8), we would have that

$$D[\mathfrak{L}_T(\boldsymbol{\varphi})] = D[L_T(\boldsymbol{\phi})] \cdot D[f(\boldsymbol{\varphi})],$$

where the symbol $D[.]$ denotes the corresponding Jacobian matrix. Similarly, we can use the chain rule for Hessian matrices (Magnus and Neudecker 1988, thm. 6.9) to write:

$$\begin{aligned} H[\mathfrak{L}_T(\boldsymbol{\varphi})] = &\, D[f(\boldsymbol{\varphi})]' \, H[L_T(\boldsymbol{\phi})] \, D[f(\boldsymbol{\varphi})] \\ &+ (D[L_T(\boldsymbol{\phi})] \otimes I_{p+1}) \, H[f(\boldsymbol{\varphi})], \end{aligned}$$

where the symbol $H[.]$ denotes the corresponding Hessian matrix.

2.3 The score vector

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $s_{\eta t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and η respectively. Given that

$$\frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = -\frac{1}{2}vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'}$$

and

$$\begin{aligned}\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\boldsymbol{\theta}'} &= \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = \\ &-\frac{N\eta+1}{2\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})\right]}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} &= -2\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \\ &-vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'},\end{aligned}$$

we can immediately show that:

$$\begin{aligned}\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\boldsymbol{\theta}} = \\ &\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}) \\ &+\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \\ &\times vec\left[\frac{N\eta+1}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right], \qquad (1)\end{aligned}$$

where the Jacobian matrices $\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}'$ depend on the particular specification adopted.

Similarly, it is straightforward to see that

$$s_{\eta t}(\boldsymbol{\phi}) = \frac{\partial c(\eta)}{\partial\eta} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\eta},$$

which for $\eta > 0$ are given by:

$$\frac{\partial c(\eta)}{\partial\eta} = \frac{N}{2\eta\left(1-2\eta\right)} - \frac{1}{2\eta^2}\left[\psi\left(\frac{N\eta+1}{2\eta}\right) - \psi\left(\frac{1}{2\eta}\right)\right] \qquad (2)$$

and

$$\begin{aligned}\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\eta} &= -\frac{N\eta+1}{2\eta\left(1-2\eta\right)}\frac{\varsigma_t(\boldsymbol{\theta})}{1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})} \\ &+\frac{1}{2\eta^2}\ln\left[1+\frac{\eta}{1-2\eta}\varsigma_t(\boldsymbol{\theta})\right], \qquad (3)\end{aligned}$$

where $\psi(x) = \partial \ln \Gamma(x)/\partial x$ is the so-called di-gamma function (or Gauss' psi function; Abramowitz and Stegun 1964), which can be computed using standard routines.

If we then take limits as $\eta \to 0$ from above, we can once more show that $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},0)$ does indeed reduce to the multivariate normal expression in Bollerslev and Wooldridge (1992). Unfortunately, both $\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]/\partial\eta$ and especially $\partial c(\eta)/\partial\eta$ are numerically unstable for η small. When $N=1$, for instance, Figure 1 shows that the numerical accuracy in the computation of (2) is poor for small η, and eventually breaks down. In those cases, we recommend the evaluation of (2) and (3), which in the limit should be understood as right derivatives, by means of the (directional) Taylor expansions around $\eta = 0$ in Appendix B.

In practice, the log-likelihood score is often used not only as the input to a steepest ascent, BHHH or quasi-Newton numerical optimisation routine, but also to estimate the asymptotic covariance matrix of the ML parameter estimators. Nevertheless, both these uses could be problematic. First, the results in Fiorentini et al. (1996) and many others suggest that alternative gradient methods, such as scoring or Newton-Raphson, usually show much better convergence properties, particularly when the parameter values reach the neighbourhood of the optimum. Similarly, it is well known that the outer-product-of-the-score standard errors and test statistics can be very badly behaved in finite samples, especially in dynamic models (Davidson and MacKinnon 1993). For both these reasons, we follow Bollerslev and Wooldridge (1992), and derive in the next two sections analytic expressions for the Hessian matrix and its expected value conditional on $\mathbf{z}_t, I_{t-1}$.

2.4 The Hessian matrix

Let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$, and partition it into four blocks, $\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi})$, $\mathbf{h}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi})(=\mathbf{h}'_{\eta\boldsymbol{\theta}t}(\boldsymbol{\phi}))$ and $h_{\eta\eta t}(\boldsymbol{\phi})$, whose row and column dimensions conform to those of $\boldsymbol{\theta}$ and η.

Let us start with the first block, which will be given by

$$\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}.$$

It is then straightforward to see that

$$\begin{aligned}\frac{\partial^2 d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= \frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'} \\ &\quad -\frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_p\right\}\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\right\}.\end{aligned}$$

In addition, if we use again the chain rule for Hessian matrices, we obtain:

$$\frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma^2}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma}\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},$$

where

$$\frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]}{\partial\varsigma^2} = \frac{(N\eta+1)\,\eta}{2\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})\right]^2}.$$

We can then show that

$$\begin{aligned}
\frac{\partial^2\varsigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= 2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + 2\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}} \\
&\times\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'} \\
&+2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}'} \\
&+2\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \\
&-2\left[\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\mathbf{I}_p\right]\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\} \\
&-\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_p\right\}\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\right\}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) = \frac{\partial^2 g(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}\partial\eta} &= \frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\frac{N+2-\varsigma_t(\boldsymbol{\phi})}{\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\phi})\right]^2} \\
&+\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \\
&\times vec\left\{\frac{N+2-\varsigma_t(\boldsymbol{\phi})}{\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\phi})\right]^2}\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t'(\boldsymbol{\theta})\right\}.
\end{aligned}$$

Similarly, for $\eta > 0$

$$\begin{aligned}
\frac{\partial^2 c(\eta)}{\partial\eta^2} &= \frac{N}{2}\frac{4\eta-1}{\eta^2\left(1-2\eta\right)^2} \\
&-\frac{1}{\eta^3}\left[\psi\left(\frac{1}{2\eta}\right)-\psi\left(\frac{N\eta+1}{2\eta}\right)\right] \\
&+\frac{1}{4\eta^4}\left[\psi'\left(\frac{N\eta+1}{2\eta}\right)-\psi'\left(\frac{1}{2\eta}\right)\right],
\end{aligned}\tag{4}$$

where $\psi'(x) = \partial^2\ln\Gamma(x)/\partial x^2$ is the so-called tri-gamma function (Abramowitz and Stegun 1964). Finally, we have that when

$\eta > 0$

$$
\begin{aligned}
\frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}), \eta\right]}{\partial \eta^2} &= -\frac{1}{\eta^3} \ln\left[1 + \frac{\eta}{1-2\eta}\varsigma_t(\boldsymbol{\theta})\right] \\
&+ \frac{1}{\eta^2} \frac{\varsigma_t(\boldsymbol{\theta})}{\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})\right](1-2\eta)} \\
-\frac{N\eta+1}{2\eta} &\left\{\frac{4(1-2\eta)\varsigma_t(\boldsymbol{\theta}) + (4\eta-1)\,\varsigma_t^2(\boldsymbol{\theta})}{(1-2\eta)^2\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta})\right]^2}\right\}. \qquad (5)
\end{aligned}
$$

Again, both (4) and (5) can be numerically unstable when η is small. Hence, we also recommend to evaluate them in those cases by means of the (directional) Taylor expansions in Appendix B.

2.5 The conditional information matrix

Given correct specification, the results in Crowder (1976) imply that the score vector $\mathbf{s}_t(\boldsymbol{\phi})$ evaluated at the true parameter values has the martingale difference property. His results also imply that, under suitable regularity conditions, the asymptotic distribution of the ML estimator will be given by the following expression

$$
\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \to N\left[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)\right], \qquad (6)
$$

where

$$
\mathcal{I}(\boldsymbol{\phi}_0) = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathcal{I}_t(\boldsymbol{\phi}_0),
$$

and

$$
\mathcal{I}_t(\boldsymbol{\phi}_0) = V\left[\mathbf{s}_t(\boldsymbol{\phi}_0)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0\right] = -E\left[\mathbf{h}_t(\boldsymbol{\phi}_0)|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0\right].
$$

In this respect, we show in Appendix A the following result:

Proposition 1

$$
\mathcal{I}_t(\boldsymbol{\phi}) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) \\ \mathcal{I}'_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) & \mathcal{I}_{\eta\eta t}(\boldsymbol{\phi}) \end{pmatrix}
$$

$$
\begin{aligned}
\mathcal{I}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\nu\,(N+\nu)}{(\nu-2)\,(N+\nu+2)} \frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\
+\frac{(N+\nu)}{2(N+\nu+2)} &\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} \left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}'} \\
&-\frac{1}{2(N+\nu+2)} \frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} vec\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \\
&\times vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right] \frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}'},
\end{aligned}
$$

$$\mathcal{I}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi}) = -\frac{(N+2)\,\nu^2}{(\nu-2)\,(N+\nu)\,(N+\nu+2)} \times \frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} vec\left\{\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right\},$$

$$\mathcal{I}_{\eta\eta t}(\boldsymbol{\phi}) = \frac{\nu^4}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{N+\nu}{2}\right)\right] - \frac{N\nu^4\left[\nu^2 + N(\nu-4) - 8\right]}{2\,(\nu-2)^2\,(N+\nu)\,(N+\nu+2)}.$$

It is important to note that unlike the Hessian, the above expressions only require first derivatives of the conditional mean and variance functions (see Fiorentini et al. 1996 for the required derivatives in univariate linear regression models with GARCH(p,q) innovations, and Sentana in press for analogous expressions in multivariate conditionally heteroskedastic in mean factor models). It also shares with the outer-product-of-the-score formula the convenient property of being positive semi-definite, usually with full rank.

3 AN LM TEST FOR MULTIVARIATE NORMALITY

3.1 Implementation Details

We can easily compute an LM (or efficient score) test for multivariate normality versus multivariate t distributed innovations on the basis of the value of the score of the log-likelihood function evaluated at the restricted parameter estimates $\tilde{\boldsymbol{\phi}}_T = (\tilde{\boldsymbol{\theta}}_T', 0)'$. To do so, it is necessary to find the value of $s_{\eta t}(\boldsymbol{\theta}, 0)$. In this sense, we can use the results in Appendix B to prove that

$$s_{\eta t}(\boldsymbol{\theta}, 0) = \frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta}),$$

where $s_{\eta t}(\boldsymbol{\theta}, 0)$ should be understood as a directional derivative. In addition, it turns out that the information matrix is block-diagonal between $\boldsymbol{\theta}$ and η when $\eta_0 = 0$, which means that as far as $\boldsymbol{\theta}$ is concerned, there is no asymptotic efficiency loss in estimating η in that case. More formally:

Proposition 2 *If $\eta_0 = 0$, then*

$$V\left[\mathbf{s}_{\phi}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, 0\right] = \begin{bmatrix} V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, 0\right] & \mathbf{0} \\ \mathbf{0}' & N(N+2)/2 \end{bmatrix},$$

where

$$V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, 0\right] = -E\left[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, 0\right].$$

Therefore, we can compute the information matrix version of the LM test, $LM_{2T}^{I}(\tilde{\boldsymbol{\theta}}_T)$ say, as the square of:

$$\tau_T^I(\tilde{\boldsymbol{\theta}}_T) = \frac{T^{-1/2}\sum_t s_{\eta t}(\tilde{\boldsymbol{\theta}}_T, 0)}{\sqrt{N(N+2)/2}}, \tag{7}$$

which, importantly, only depends on the first two sample moments of $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$. Note also that the block-diagonality of the information matrix implies that a joint LM test of multivariate normality and any other restrictions on the conditional mean and variance parameters $\boldsymbol{\theta}$, can be decomposed in two additive components, the first of which would be precisely our proposed test (Bera and McKenzie 1987). If $H_0 : \eta = 0$ is true, then $LM_{2t}^{I}(\tilde{\boldsymbol{\theta}}_T)$ will have an asymptotic chi-square distribution with one degree of freedom. The limiting distribution can be obtained directly from (7) by combining the block-diagonality of the information matrix under the null with the following result:

Proposition 3 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1} \sim i.i.d.\ t(\mathbf{0}, \mathbf{I}_N, \nu_0)$ with $\nu_0 > 8$, then*

$$\frac{\sqrt{T}}{T}\sum_t \left[\frac{s_{\eta t}(\boldsymbol{\theta}_0, 0) - E\left[s_{\eta t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, \nu_0\right]}{V^{1/2}\left[s_{\eta t}^2(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, \nu_0\right]}\right] \xrightarrow{d} N(0,1),$$

where

$$E\left[s_{\eta t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, \nu_0\right] = \frac{N(N+2)}{4}\left(\frac{\nu_0 - 2}{\nu_0 - 4} - 1\right)$$

$$\begin{aligned} E\left[s_{\eta t}^2(\boldsymbol{\theta}_0, 0)|\boldsymbol{\theta}_0, \nu_0\right] = &-\frac{3N^2(N+2)^2}{16} \\ &+\frac{N(N+2)^2(3N+4)}{8}\frac{\nu_0-2}{\nu_0-4} \\ &-\frac{N(N+2)^2(N+4)}{4}\frac{(\nu_0-2)^2}{(\nu_0-4)(\nu_0-6)} \\ &+\frac{N(N+2)(N+4)(N+6)}{16}\frac{(\nu_0-2)^3}{(\nu_0-4)(\nu_0-6)(\nu_0-8)}. \end{aligned}$$

Two asymptotically equivalent test, both under the null and under local alternatives, are given by (i) the usual outer product version of the LM test, $LM_{2T}^{O}(\tilde{\boldsymbol{\theta}}_T)$, which can be computed as T times the uncentred R^2 from the regression of 1 on $s_{\eta t}(\tilde{\boldsymbol{\theta}}_T, 0)$, and (ii) its Hessian version:

$$LM_{2T}^{H}(\tilde{\boldsymbol{\theta}}_T) = \frac{\left\{T^{-1/2}\sum_t s_{\eta t}(\tilde{\boldsymbol{\theta}}_T, 0)\right\}^2}{-T^{-1}\sum_t h_{\eta\eta t}(\tilde{\boldsymbol{\theta}}_T, 0)}, \tag{8}$$

with

$$\begin{aligned} h_{\eta\eta t}(\boldsymbol{\theta}, 0) = &-\frac{N(N+2)(N-5)}{6} - (4+2N)\varsigma_t(\boldsymbol{\theta}) \\ &+\frac{N+4}{2}\varsigma_t^2(\boldsymbol{\theta}) - \frac{1}{3}\varsigma_t^3(\boldsymbol{\theta}). \end{aligned}$$

Given that the numerators of those three LM tests coincide, while the denominators of $LM^H_{2T}(\tilde{\boldsymbol{\theta}}_T)$ and $LM^O_{2T}(\tilde{\boldsymbol{\theta}}_T)$ converge in probability to the denominator of $LM^I_{2T}(\tilde{\boldsymbol{\theta}}_T)$, which contains no stochastic terms, we would expect a priori that $LM^I_{2T}(\tilde{\boldsymbol{\theta}}_T)$ would be the version of the test with the smallest size distortions, followed by $LM^H_{2T}(\tilde{\boldsymbol{\theta}}_T)$, whose denominator involves the first three sample moments of $\varsigma_t(\boldsymbol{\theta})$, and finally $LM^O_{2T}(\tilde{\boldsymbol{\theta}}_T)$, whose calculation also requires its fourth sample moment (Davidson and MacKinnon 1983).

It is important to mention that the fact that $\eta = 0$ lies at the boundary of the admissible parameter space invalidates the usual χ^2_1 distribution of the likelihood ratio (LR) and Wald (W) tests, which under the null will be more concentrated towards the origin (see Andrews 2001 and the references therein, as well as the simulation evidence in Bollerslev 1987). The intuition can be perhaps more easily obtained in terms of the W test. Given that $\hat{\eta}_T$ cannot be negative, $\sqrt{T}\hat{\eta}_T$ will have a half-normal asymptotic distribution under the null (Andrews 1999). As a result, the W test will be an equally weighted mixture of a chi-square distribution with 0 degrees of freedom (by convention, χ^2_0 is a degenerate random variable that equals zero with probability 1), and a chi-square distribution with 1 degree of freedom. In practice, obviously, we simply need to compare the t-statistic $\sqrt{TN(N+2)/2}\hat{\eta}_T$ with the appropriate one-sided critical value from the normal tables. For analogous reasons, the asymptotic distribution of the LR test will also be degenerate half the time, and a chi-square with one degree of freedom the other half.

Although the above argument does not invalidate the distribution of the LM test statistic, intuition suggests that the one-sided nature of the alternative hypothesis should be taken into account to obtain a more powerful test. For that reason, we also propose a simple one-sided version of the LM test for multivariate normality. In particular, since $E\left[s_{\eta t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\phi}_0\right] > 0$ when $\eta_0 > 0$ in view of Proposition 3, we suggest to use

$$LM^I_{1T}(\tilde{\boldsymbol{\theta}}_T) = \left\{\max\left[\tau^I_T(\tilde{\boldsymbol{\theta}}_T), 0\right]\right\}^2$$

as our one-sided LM test, and to compare it to the same 50:50 mixture of chi-squares 0 and 1. In this context, we would reject H_0 at the $100\varkappa\%$ significance level if the average score with respect to η evaluated at the Gaussian pseudo-ML (PML) estimators $\tilde{\boldsymbol{\phi}}_T = (\tilde{\boldsymbol{\theta}}'_T, 0)'$ is positive *and* $LM^I_{1T}(\tilde{\boldsymbol{\theta}}_T)$ exceeds the $100(1-2\varkappa)$ percentile of a χ^2_1 distribution. Since the Kuhn-Tucker (KT) multiplier associated with the inequality restriction $\eta \geq 0$ is equal to $\max[-T^{-1}\sum_t s_{\eta t}(\tilde{\boldsymbol{\theta}}_T, 0), 0]$, our proposed one-sided LM test is equivalent to the KT multiplier test introduced by Gourieroux, Holly and Monfort (1980), which in turn is equivalent in large samples to the LR and W tests. As we argued before, the reason is that those tests are implicitly one-sided in our context. In this respect, it is important to mention that when there is a single restriction, such as in our case, those one-sided tests would be asymptotically locally more powerful (Andrews 2001).

Nevertheless, it is still interesting to compare the power properties of the one-sided and two-sided LM statistics. But given that the block-diagonality of the information matrix is generally lost under the alternative of $\eta_0 > 0$, and its exact form is unknown, we can only get closed form expressions for the case in which the standardised innovations $\boldsymbol{\varepsilon}_t^*$ are directly observed. In more realistic cases, though, the results are likely to be qualitatively similar. On the basis of Proposition 3, we can obtain the asymptotic power of the one-sided and two-sided variants of the information matrix version of the LM test for any possible significance level $\varkappa$. The results at the usual 5% level are plotted in Figures 2a, 2b and 2c for η_0 in the range $0 \leq \eta_0 \leq .04$, that is $\nu_0 \geq 25$. Not surprisingly, the power of both tests uniformly increases with the sample size T for a fixed alternative, and as we depart from the null for a given sample size. Importantly, their power also increases with the number of series N. As expected, the one-sided test is more powerful than the usual two-sided one. The difference is particularly noticeable for small departures from the null, which is precisely when power is generally low. For instance, when $\nu_0 = 100$, $T = 500$ and $N = 10$, the power of the one-sided test is almost 60% while the power of its two-sided counterpart is less than 50% (see Figure 2c). Similarly, the one-sided tests for $N = 1$ and $N = 5$ are initially more powerful than the two-sided tests for $N = 2$ and $N = 10$ respectively. However, as η_0 approaches 1/8 from below, the one-sided test looses power for fixed N and T, and eventually the two-sided test becomes more powerful. This is due to the fact that the variance of the score goes to infinity as $\nu_0 \rightarrow 8$ from Proposition 3

Although in view of Lemma 1 in Appendix A, our proposed LM test can be regarded as a test of whether $\varsigma_t(\boldsymbol{\theta}_0)$ is χ^2_N against the alternative that it is proportional to an F_{N,ν_0}, it is also possible to re-interpret (7) as a specification test of the restriction on the first two moments of $\varsigma_t(\boldsymbol{\theta}_0)$ implicit in $E\left[s_{\eta t}(\boldsymbol{\theta}_0, 0)|\boldsymbol{\phi}_0\right] = 0$. More specifically:

$$E\left[\frac{N(N+2)}{4} - \frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta})\right| \boldsymbol{\phi}_0\Bigg] = 0. \qquad (9)$$

Hence, the two-sided version has non-trivial power against any spherically symmetric distribution for which $s_{\eta t}(\boldsymbol{\theta}_0, 0)$ has expected value different from zero (see Theorem 1 in Appendix A for a characterisation of spherically symmetric distributions). For instance, if we consider the extreme case in which the true standardised disturbances were in fact uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, so that $\varsigma_t(\boldsymbol{\theta}_0) = N\ \forall t$, then $s_{\eta t}(\boldsymbol{\theta}_0, 0) = -N(N+1)/4$, which means that we would reject the null hypothesis with probability approaching one as T goes to infinite. On the other hand, the one-sided LM test only has power for the leptokurtic subclass of spherically symmetric distributions. Nevertheless, as we shall see in Section 5, standardised residuals are frequently leptokurtic and rarely platykurtic in practice.

3.2 Relationship with existing kurtosis tests

Following Mardia (1970), we can define the population coefficient of multivariate excess kurtosis as:

$$\kappa = \frac{E\left[\varsigma_t^2(\boldsymbol{\theta}_0)\right]}{N(N+2)} - 1, \tag{10}$$

which equals $2/(\nu_0 - 4)$ for the multivariate t distribution, as well as its sample counterpart:

$$\bar{\kappa}_T(\boldsymbol{\theta}) = \frac{T^{-1}\sum_{t=1}^{T}\varsigma_t^2(\boldsymbol{\theta})}{N(N+2)} - 1. \tag{11}$$

On this basis, we can write $\tau_T^I(\tilde{\boldsymbol{\theta}}_T)$ in (7) as

$$\sqrt{\frac{N(N+2)}{8}}\left\{\frac{\sqrt{T}}{T}\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T) - \frac{2\sqrt{T}}{NT}\sum_{t=1}^{T}\left[\varsigma_t(\tilde{\boldsymbol{\theta}}_T) - N\right]\right\}.$$

If we ignored the term

$$\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\left[\varsigma_t(\tilde{\boldsymbol{\theta}}_T) - N\right], \tag{12}$$

then (7) would coincide with the kurtosis component of Mardia's (1970) test for multivariate normality, which in turn reduces to the popular Jarque and Bera (1980) test in the univariate case. Hence, given that if $T^{-1}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t^*(\tilde{\boldsymbol{\theta}}_T)\boldsymbol{\varepsilon}_t^{*\prime}(\tilde{\boldsymbol{\theta}}_T) = \mathbf{I}_N$ then (12) is identically 0, it follows from (1) that their tests are *numerically* identical to ours in nonlinear regression models with conditionally homoskedastic disturbances estimated by Gaussian PML, in which the covariance matrix of the innovations, $\boldsymbol{\Sigma}$, is unrestricted and does not affect the conditional mean, and the conditional mean parameters, $\boldsymbol{\delta}$ say, and the elements of $vech(\boldsymbol{\Sigma})$ are variation free. However, ignoring (12) in more general contexts may lead to size distortions, because it is precisely the inclusion of such a term what makes $s_{\eta t}(\boldsymbol{\theta}_0, 0)$ orthogonal to the other elements of the score. The same point was forcefully made by Davidson and MacKinnon (1993) in a univariate context in Section 16.7 of their textbook, and not surprisingly, their suggested test for excess kurtosis turns out to be equal to the outer product version of our LM test. Similarly, the term (12) also appears explicitly in the Kiefer and Salmon (1983) LM test for univariate excess kurtosis based on a Hermite polynomial expansion of the density, which coincides in their context with the information matrix version of our test (7) (see Hall 1990 for an extension to models in which the higher order moments depend on the information set).

Nevertheless, the exclusion of the additional term (12) does not necessarily lead to asymptotic size distortions. In particular, there will be no size distortions if (12) is $o_p(1)$. A necessary and sufficient condition for this to happen is that $\varsigma_t(\boldsymbol{\theta}_0) - N$ can be written as an exact, time-invariant, linear combination of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)$ (Fiorentini, Sentana and Calzolari 2003). Given that such a condition involves a rather complicated system of nonlinear differential equations, it is not possible to explicitly characterise which models for $\boldsymbol{\mu}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ will satisfy it, so we have to proceed on a model by model basis. In this respect, Fiorentini et al. (2003) establish that the condition is indeed satisfied for the family of GARCH-M models analysed by Hentschel (1995). Nevertheless, it is possible to find examples of other ARCH models in which the aforementioned is not satisfied (e.g. the variant of the EGARCH model proposed in Barndorf-Nielsen and Shephard 2001, chap. 13). Therefore, the conclusion to draw from the above analysis is that even though the asymptotic size of the tests commonly employed by practitioners is often correct, it is safer to use the LM test (7) because its limiting null distribution never depends on the particular parametrisation used, and the additional computational cost is negligible.

Finally, several authors have recently suggested alternative multivariate generalisations of the Jarque-Bera test, which as far as kurtosis is concerned, consist in adding up the univariate kurtosis tests for each element of $\boldsymbol{\varepsilon}_t^*(\tilde{\boldsymbol{\theta}}_T)$ (see Lütkepohl 1993; Doornik and Hansen 1994; Kilian and Derimoglou 2000). But apart from the issue discussed in the previous paragraphs, another potential shortcoming of those tests is that they are not invariant to the way in which the residuals $\boldsymbol{\varepsilon}_t(\tilde{\boldsymbol{\theta}}_T)$ are orthogonalised to obtain $\boldsymbol{\varepsilon}_t^*(\tilde{\boldsymbol{\theta}}_T)$. For instance, while Doornik and Hansen (1994) obtain $\Sigma_t^{1/2}(\tilde{\boldsymbol{\theta}}_T)$ from the spectral decomposition of $\Sigma_t(\tilde{\boldsymbol{\theta}}_T)$, the other authors use a Cholesky decomposition. In this respect, note that by implicitly assuming that the excess kurtosis is the same for all possible linear combinations of the true standardised innovations $\boldsymbol{\varepsilon}_t^*$, we obtain a test statistic which is numerically invariant to orthogonal rotations of $\Sigma_t^{1/2}(\tilde{\boldsymbol{\theta}}_T)$ (see also Mardia 1970). If $\boldsymbol{\varepsilon}_t^*$ were directly observed, the relative power of the two testing procedures would depend on the exact nature of the alternative hypothesis. Given that the ε_{it}^*'s are independent across $i = 1, \ldots, N$ under the null, the situation is completely analogous to the comparison between the one-sided tests for ARCH(q) of Lee and King (1993) and Demos and Sentana (1998). In particular, if we define $\kappa_i = E(\varepsilon_{it}^{*4}/3) - 1$ for $i = 1, \ldots, N$, our test would be more powerful against alternatives close to $\kappa_i = \kappa$ for all i, while the additive test would have more power when the $\kappa_i's$ were rather dispersed.

4 A MONTE CARLO COMPARISON OF ALTERNATIVE ESTIMATION PROCEDURES AND STANDARD ERROR ESTIMATORS

In this section, we assess the performance of two alternative ways of obtaining ML estimates of $\boldsymbol{\phi}$, and three common ways of estimating the corresponding standard errors. The first estimation procedure employs the following mixed approach: initially, we use a scoring algorithm with a fairly large tolerance criterion; then, after "convergence" is achieved, we switch to a Newton-Raphson algorithm to refine the solution. Both stages are implemented by means of the NAG Fortran 77 Mark 19 library E04LBF routine (see Numerical Algorithm Group 2001 for details), with the analytic expressions for $\mathbf{s}_t(\boldsymbol{\phi})$, $\mathcal{I}_t(\boldsymbol{\phi})$ and $\mathbf{h}_t(\boldsymbol{\phi})$ derived in Section 2. The second procedure, in contrast, uses a quasi-Newton algorithm that computes the score on the basis of finite difference procedures, as implemented by the NAG E04JYF routine. Importantly, both routines allow for fixed upper and lower bounds in the elements of $\boldsymbol{\phi}$. In this respect, we should mention that when η is close to zero, the quasi-Newton algorithm that uses function values only sometimes fails to converge. For that reason, and in accordance with standard practice, we set $\hat{\eta}_T$ to 0 whenever its estimate falls below a minimum threshold, $\eta_{\min}$. In particular, we follow Microfit 4.0 in choosing $\eta_{\min} = .04$ (Pesaran and Pesaran 1997, p. 457). Then, we maximise a Gaussian pseudo-log likelihood function using the combined scoring plus Newton-Raphson algorithm explained above.

As for the estimators of the asymptotic covariance matrix of the ML parameter estimators, we consider the three standard approaches: outer-product of the gradient (OPG), Hessian (H) and conditional information (CI) matrix. In order to replicate what an empirical researcher would do in practice, though, we do not employ numerical expressions when analytic expressions are used in the optimisation algorithm, and vice versa. Therefore, we end up with five different combinations of estimators and standard errors.

We assess their performance by means of an extensive Monte Carlo analysis, with an experimental design borrowed from Bollerslev and Wooldridge (1992). Specifically, the model that we simulate and estimate is given by the following equations:

$$\begin{aligned}
y_t &= \mu_t(\boldsymbol{\delta}_0) + \sigma_t(\boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)\varepsilon_t^* \\
\mu_t(\boldsymbol{\delta}) &= \upsilon + \rho y_{t-1} \\
\sigma_t^2(\boldsymbol{\delta}, \boldsymbol{\gamma}) &= \vartheta + \alpha\left[y_{t-1} - \mu_{t-1}(\boldsymbol{\delta})\right]^2 + \beta\sigma_{t-1}^2(\boldsymbol{\delta}, \boldsymbol{\gamma}) \\
\varepsilon_t^*|I_{t-1} &\sim i.i.d.\ t(0, 1, \nu_0)
\end{aligned}$$

where $\boldsymbol{\delta}' = (\upsilon, \rho)$, $\boldsymbol{\gamma}' = (\vartheta, \alpha, \beta)$, $\upsilon_0 = 1$, $\rho_0 = .5$, $\vartheta_0 = .05$, $\alpha_0 = .15$ and $\beta_0 = .8$. As for η_0, we consider three different values: $0, .04$ and $.1$, which correspond to the Gaussian limit, and two Student t's with 25 and 10 degrees of freedom respectively.

Given the large number of parameters involved, we summarise the performance of the estimates of the asymptotic covariance matrix of the estimators by computing the experimental distribution of a very simple W test statistic. In particular, the null hypothesis that we test is that all six parameters are equal to their true values. When $\eta_0 > 0$, the asymptotic distribution of such a test will be χ_6^2. In contrast, when $\eta_0 = 0$, it follows from the discussion in Section 3.1 imply that its asymptotic distribution will be a 50:50 mixture of χ_5^2 and χ_6^2. Our results, which are based on $10,000$ samples of $1,000$ observations each, are summarised in Figures 3a-3c using Davidson and MacKinnon's (1998) **p-value discrepancy plots**, which show the difference between actual and nominal test sizes for every possible nominal size. As expected, the CI standard errors seem to be the most reliable, followed by the H-based ones, and finally, the OPG versions, which tend to show the largest size distortions. In addition, there is a marked difference between numeric and analytic expressions. In Figure 3a, for instance, the performance of the numerical H standard errors is as distorted as the performance of the analytic OPG ones. But the most striking difference arises when $\eta_0 = .04$ (see Figure 3b). In this case, the two numerical approaches lead to much larger size distortions. This is due to two different reasons. First, the loss of accuracy in the computation of first and second derivatives by relative numerical increments of the log-likelihood function can be substantial when η is small, as illustrated in Figure 4 for a randomly selected replication. But our setting η to 0 whenever $\eta \leq \eta_{\min}$ has an even stronger impact. As Figure 5 illustrates, if we reduce $\eta_{\min}$ from .04 to .00001, then the behaviour of numerical and analytical methods is more in line. However, the problem with using such a small value of $\eta_{\min}$ is that convergence failures occur much more frequently. On the basis of these results, our practical recommendation would be to use the mixed optimisation algorithm described above with analytical derivatives, and to compute standard errors with the formulae for the conditional information matrix in Proposition 1.

5 AN EMPIRICAL APPLICATION TO UK STOCK RETURNS

In this section, we investigate the practical performance of the procedures discussed above. To do so, we substantially extend the analysis in Sentana (1991), who considered both Gaussian and t distributions in his empirical characterisation of multivariate leverage effects by means of a conditionally heteroskedastic latent factor model for the monthly excess returns on 26 U.K. sectorial indices for the period 1971:2 to 1990:10 (237 observations), with a GQARCH(1,1) parametrisation for the common factor, and a constant diagonal covariance matrix for the idiosyncratic terms. In order to concentrate on the modelling of the second and higher order moments of the conditional distribution of returns, all the data was demeaned prior to estimation. Nevertheless, a more explicit modelling of the mean has little impact on the remaining parameters (Sentana 1995). Specifically, the model he initially estimated by Gaussian PML is:

$$\mathbf{y}_t = \mathbf{c}f_t + \mathbf{w}_t,$$

$$\begin{pmatrix} f_t \\ \mathbf{w}_t \end{pmatrix} | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots \sim N\left[\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \lambda_t & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Gamma} \end{pmatrix} \right],$$

$$\lambda_t = \vartheta + \alpha \left[\left(f_{t-1|t-1} - \upsilon \right)^2 + \omega_{t-1|t-1} \right] + \beta \lambda_{t-1},$$

where $\mathbf{y}_t$ is the vector of returns, $\mathbf{c}$ the vector of factor loadings, $\mathbf{\Gamma}$ the diagonal matrix of idiosyncratic variances, $f_{t|t} = \omega_{t|t}\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{y}_t$ is the Kalman-filter based estimate of the latent factor, and $\omega_{t|t} = [\lambda_t^{-1} + (\mathbf{c}'\mathbf{\Gamma}^{-1}\mathbf{c})]^{-1}$ the corresponding conditional mean square error. Note that λ_t differs from a standard GQARCH(1,1) specification in that the unobserved factors are replaced by their expected value $f_{t-1|t-1}$, and the term $\omega_{t-1|t-1}$ is included to reflect the uncertainty in the factor estimates (Harvey et al. 1992). We solved the usual scale indeterminacy of the factor by fixing $E(\lambda_t) = 1$. To do so, we set $\vartheta = (1-\alpha-\beta) - \alpha\upsilon$, $\upsilon = \sqrt{(1-\alpha-\beta)/\alpha\varrho}$, and estimated the model subject to the inequality constraints $0 \leq \beta \leq 1-\alpha \leq 1$ and $-1 \leq \varrho \leq 1$, which also ensure that $\lambda_t \geq 0 \;\forall t$. PML estimates for α, β, ϱ and η can be found in Table 1, together with robust standard errors a la Bollerslev and Wooldridge (1992) calculated with analytical derivatives. On the basis of those estimates, we generated the time series of squared Euclidean norms of the standardised innovations, $\varsigma_t(\tilde{\boldsymbol{\theta}}_T)$, and computed the information matrix version of the LM tests for multivariate normality described in Section 3. Since $\tau_T^I(\tilde{\boldsymbol{\theta}}_T)$ equals 54.43, we can easily reject the null hypothesis regardless of whether we use a one-sided or a two-sided critical value, which suggests that it is worth estimating the same model with the student t. Given that $\tilde{\boldsymbol{\theta}}_T$ is a consistent estimator, we used it as initial values for $\boldsymbol{\theta}$. As for η, we used .106, which is the value of a consistent, two-stage method of moments estimator obtained from the sample coefficient of excess kurtosis of the standardised residuals $\bar{\kappa}_T(\tilde{\boldsymbol{\theta}}_T)$ by exploiting the theoretical relationship $\eta = \kappa/(4\kappa + 2)$. ML parameter estimates, together with standard errors based on Proposition 1, are also reported in Table 1. Apart from the marked improvement in fit, as measured by the increase in the likelihood function and the decrease in standard errors, and the fact that the estimated ϱ is now at the boundary of the admissible parameter space, the most noticeable difference is the drastic reduction in the parameter α, which measures the immediate effect of shocks to the level of the conditional variance, and the slight increase in the parameter β, which measures the rate at which the impact of those shocks decays over time.

In order to compare the two models from a graphical perspective, we have estimated the conditional standard deviations that they generate for an equally weighted portfolio. In this respect, note that conditional variance of $\boldsymbol{\iota}'\mathbf{y}_t/N$ implied by our single factor model is $(\mathbf{c}'\boldsymbol{\iota}/N)^2\lambda_t + \boldsymbol{\iota}'\boldsymbol{\Gamma}\boldsymbol{\iota}/N^2$, where $\boldsymbol{\iota}$ is a $N \times 1$ vector of ones. Although the correlation between both series is high (97.6%), the results depicted in Figure 6 indicate that the t distribution tends to produce less extreme values for λ_t. This is particularly true around the two most significant episodes in the sample: the October 1987 crash (a 23.6% drop in stock prices), and the January 1975 bounce back (a 51.5% surge).

As mentioned in the introduction, one of the reasons for using the t distribution is to compute the quantiles of the one-period-ahead predictive distributions of portfolio returns required in Value at Risk calculations. To determine to what extent the t is more useful than the normal in this respect across all conceivable quantiles, we have computed the empirical cumulative distribution function of the probability integral transforms of the equally weighted portfolio returns generated by the two fitted distributions (see Diebold, Gunther and Tay 1998). Figure 7 shows the difference between those two cumulative distributions and the 45° degree line. Under correct specification, those differences should tend to 0 asymptotically. Unfortunately, a size-corrected version of the usual Kolmogorov-type test that takes into account the sample uncertainty in the estimates of the underlying parameters is rather difficult to obtain in this case. Nevertheless, the graph clearly suggests that the multivariate t distribution does indeed provide a better fit than the normal, especially in the tails. In this respect, it is important to emphasize that the estimating criterion is multivariate, and not targeted to this particular portfolio. The observed differences are partly due to the fact that the t distribution has both fatter tails than the normal and more density around its mean. However, this cannot be the only reason, for a standardised univariate t distribution with 9.71 degrees of freedom and a standard normal share not only the median, but also the 3.6 and 96.4 percentiles. The other reason for the differing results are the differences in estimated volatilities plotted in Figure 6.

6 CONCLUSIONS

In the context of the general multivariate dynamic regression model with time-varying variances and covariances considered by Bollerslev and Wooldridge (1992), our main contributions are:

1. We provide numerically reliable analytical expressions for the score vector, the Hessian matrix, and its conditional expected value when the distribution of the innovations is assumed to be proportional to a multivariate t.

2. We conduct a detailed Monte Carlo experiment in which we demonstrate that a mixed scoring-Newton-Raphson algorithm with analytical derivatives constitutes the best way to maximise the log-likelihood function. In addition, we show that our analytic expressions for the conditional information matrix provide the most reliable standard errors.

3. We derive an LM test for multivariate normal versus multivariate t innovations, and relate it to the kurtosis component of the traditional tests proposed by Mardia (1970) and Jarque and Bera (1980). Since the limiting null distribution of our proposed LM test is correct regardless of the model used, and the additional computational cost is negligible, we recommend its use.

4. We also derive a one-sided version of the LM test previously discussed, which apart from being more powerful than its two-sided counterpart, is asymptotically equivalent to the LR and W tests.

5. We show that the multivariate t distribution provides not only a much better fit to the distribution of U.K. sectorial returns than the normal, but also more reliable quantiles to be used in portfolio Value at Risk calculations.

Since the existing simulation evidence indicates that the finite sample size properties of many LM tests could be significantly different from the nominal levels, a fruitful avenue for future research would be to consider bootstrap procedures in order to reduce size distortions (see e.g. Kilian and Demiroglu 2000). Similarly, given that we are ruling out by assumption any asymmetries in the conditional distribution of asset returns, it would be interesting to explore asymmetric extensions of the multivariate t distribution (see Bauwens and Laurent 2002). Relatedly, it would also be worth exploring ways in which our LM test for multivariate excess kurtosis can be complemented with tests for multivariate skewness. One possibility would be to use the asymmetry component of Mardia's (1970) test for multivariate normality, which is also numerically invariant to the way in which the residuals are orthogonalised. As argued in Section 3.2, though, if the conditional mean and variance parameters have to be estimated, it may be necessary to modify his test statistic to make it orthogonal to all the elements of $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}_0, 0)$ (see Davidson and MacKinnon 1993 for the correction involved in the univariate case).

7 ACKNOWLEDGEMENTS

We are grateful to seminar audiences at Carlos III (Madrid), Federico II (Naples), and the 2001 European Meeting of the Econometric Society for very helpful comments and suggestions. Special thanks are due to Alastair Hall and Jeffrey Wooldridge for their input in revising the paper. Of course, the usual caveat applies. Financial support from CICYT, CNR, IVIE and MURST-MIUR through the projects "Stochastic models and simulation methods for dependent data" and "Statistical models for time series analysis" is gratefully acknowledged.

[Received July 2000; Revised January 2003]

Appendix

A PROOFS AND AUXILIARY RESULTS

Let us first state the three following auxiliary results, which correspond to Theorem 2.5 (iii), and Examples 2.4 and 2.5, respectively, in Fang, Kotz and Ng (1990):

Theorem 1 *$\boldsymbol{\varepsilon}_t^\circ$ is distributed as a spherically symmetric multivariate random vector of dimension N if and only if $\boldsymbol{\varepsilon}_t^\circ = e_t \mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and e_t is an non-negative random variable which is independent of $\mathbf{u}_t$.*

Example 1 *$\boldsymbol{\varepsilon}_t^\dagger$ is distributed as a standardised multivariate normal random vector of dimension N if and only if $\boldsymbol{\varepsilon}_t^\dagger = \sqrt{\zeta_t}\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, and ζ_t is an independent chi-square random variable with N degrees of freedom.*

Example 2 *$\boldsymbol{\varepsilon}_t^*$ is distributed as a standardised multivariate Student t random vector of dimension N if and only if $\boldsymbol{\varepsilon}_t^* = \sqrt{\nu_0 - 2} \times \sqrt{\zeta_t/\xi_t}\mathbf{u}_t$, where $\mathbf{u}_t$ is uniformly distributed on the unit sphere surface in $\mathbb{R}^N$, ζ_t is a chi-square random variable with N degrees of freedom, and ξ_t is a Gamma variate with mean ν_0 and variance $2\nu_0$, with $\mathbf{u}_t$, ζ_t and ξ_t mutually independent.*

The variables e_t and $\mathbf{u}_t$ are usually referred to as the generating variate and the uniform base of the spherical distribution. On this basis, we can prove the following auxiliary result:

Lemma 1 *The squared Euclidean norm of the true standardised innovations, $\varsigma_t(\boldsymbol{\theta}_0)$, is independently and identically distributed as $N(\nu_0 - 2)/\nu_0$ times an F variate with N and ν_0 degrees of freedom when $\nu_0 < \infty$, and as a chi-square random variable with N degrees of freedom under Gaussianity.*

Proof. The general result follows immediately from the fact that

$$\varsigma_t(\boldsymbol{\theta}_0) = \boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) = \frac{(\nu_0 - 2)\zeta_t \mathbf{u}_t'\mathbf{u}_t}{\xi_t} = \frac{N(\nu_0 - 2)}{\nu_0}\frac{\zeta_t/N}{\xi_t/\nu_0}.$$

The special case follows from the well known fact that ξ_t/ν_0 converges in probability to 1 as $\nu_0 \to \infty$. $\square$

Proposition 1

For our purposes, it is convenient to re-write the score function as

$$
\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}_0) = & \frac{\partial \boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \frac{N\eta_0+1}{1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\
& + \frac{1}{\sqrt{2}} \frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial \boldsymbol{\theta}} \left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0) \otimes \boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right] \\
& \times \frac{1}{\sqrt{2}} vec\left[\frac{N\eta_0+1}{1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N\right]
\end{aligned}
$$

and

$$
\begin{aligned}
s_{\eta t}(\boldsymbol{\phi}_0) = & \frac{1}{2\eta_0^2} \ln\left[1+\frac{\eta_0}{1-2\eta_0}\varsigma_t(\boldsymbol{\theta}_0)\right] \\
& - \frac{1}{2\eta_0^2}\left[\psi\left(\frac{N\eta_0+1}{2\eta_0}\right) - \psi\left(\frac{1}{2\eta_0}\right)\right] \\
& + \frac{N}{2\eta_0(1-2\eta_0)} - \frac{N\eta_0+1}{2\eta_0(1-2\eta_0)} \frac{\varsigma_t(\boldsymbol{\theta}_0)}{1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)}.
\end{aligned}
$$

In view of Lemma 1, we will have that

$$
\begin{aligned}
& \frac{N\eta_0+1}{1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0) \\
= & \frac{(\nu_0+N)}{\sqrt{(\nu_0-2)}} \sqrt{\left(\frac{\zeta_t}{\zeta_t+\xi_t}\right)\left(1-\frac{\zeta_t}{\zeta_t+\xi_t}\right)} \mathbf{u}_t,
\end{aligned}
$$

$$
\frac{N\eta_0+1}{1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)} \boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) - \mathbf{I}_N = \frac{(N+\nu_0)\,\zeta_t}{\xi_t+\zeta_t} \mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_N,
$$

$$
\begin{aligned}
& \ln\left[1+\frac{\eta_0}{1-2\eta_0}\varsigma_t(\boldsymbol{\theta}_0)\right] - \left[\psi\left(\frac{N\eta_0+1}{2\eta_0}\right) - \psi\left(\frac{1}{2\eta_0}\right)\right] \\
= & \ln\left(\frac{\xi_t+\zeta_t}{\xi_t}\right) - \left[\psi\left(\frac{\nu_0+N}{2}\right) - \psi\left(\frac{\nu_0}{2}\right)\right],
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1+N\eta_0}{2\eta_0(1-2\eta_0)} \frac{\varsigma_t(\boldsymbol{\theta}_0)}{1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)} - \frac{1}{2\eta_0} \frac{N}{(1-2\eta_0)} \\
= & \frac{\nu_0^2(\nu_0+N)}{2(\nu_0-2)} \left[\frac{\zeta_t}{\xi_t+\zeta_t} - \frac{N}{(\nu_0+N)}\right].
\end{aligned}
$$

Importantly, we only need to compute unconditional moments because $\mathbf{u}_t$, ζ_t and ξ_t are independent of $\mathbf{z}_t$ and I_{t-1} by assumption. In this respect, note that the expectation of the first term is clearly zero because all the variables involved are mutually independent, and $E(\mathbf{u}_t) = \mathbf{0}$ from Fang et al. (1990), thm. 2.7. The same theorem also implies that $E(\mathbf{u}_t\mathbf{u}_t') = N^{-1}\mathbf{I}_N$. In addition, since $\zeta_t/(\xi_t+\zeta_t)$ is an independent beta variate with parameters $N/2$ and $\nu_0/2$, whose expected value is $N/(\nu_0+N)$, then the second and fourth terms will also be 0 in expectation. Finally, we can use the results in Johnson (1949) to show that the mean of the third term is also 0.

As for the conditional information matrix, it is also convenient to write the required expressions as

$$\frac{\partial g\left[\varsigma_t(\boldsymbol{\theta}_0),\eta\right]}{\partial\varsigma} = -\frac{N\eta+1}{2\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta}_0)\right]} = -\frac{(N+\nu_0)\,\xi_t}{2\,(\nu_0-2)\,(\zeta_t+\xi_t)},$$

$$\begin{aligned}\frac{\partial\varsigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'} &= -2\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'} - vec'\left[\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\right]\\
&\quad\times\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'}\\
&= -2\sqrt{\frac{(\nu_0-2)\,\zeta_t}{\xi_t}}\mathbf{u}_t'\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'} - \frac{(\nu_0-2)\,\zeta_t}{\xi_t}vec'\left(\mathbf{u}_t\mathbf{u}_t'\right)\\
&\quad\times\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'},\end{aligned}$$

$$\frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}_0),\eta\right]}{\partial\varsigma^2} = \frac{(N\eta+1)\,\eta}{2\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\theta}_0)\right]^2} = \frac{(N+\nu_0)\,\xi_t^2}{2(\nu_0-2)^2\,(\zeta_t+\xi_t)^2},$$

and

$$\begin{aligned}\frac{\partial^2\varsigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= 2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'} + 2\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\\
&\times\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'}\\
&+2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'}\\
&+2\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\\
&-2\left[\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\mathbf{I}_p\right]\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'\left[\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\right\}\\
&-\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*}(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\otimes\mathbf{I}_p\right\}\\
&\times\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\right\} = 2\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\\
&+2\frac{(\nu_0-2)\,\zeta_t}{\xi_t}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\\
&\times\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\mathbf{u}_t\mathbf{u}_t'\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'}\\
&+2\sqrt{\frac{(\nu_0-2)\,\zeta_t}{\xi_t}}\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\left[\mathbf{u}_t'\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]\\
&\times\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'} + 2\sqrt{\frac{(\nu_0-2)\,\zeta_t}{\xi_t}}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\\
&\times\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\mathbf{u}_t\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\end{aligned}$$

$$-2\sqrt{\frac{(\nu_0-2)\,\zeta_t}{\xi_t}}\left[\mathbf{u}_t'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\otimes\mathbf{I}_p\right]\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'\left[\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\right\}$$
$$-\frac{(\nu_0-2)\,\zeta_t}{\xi_t}\left\{vec'\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\mathbf{u}_t\mathbf{u}_t'\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta}_0)\right]\otimes\mathbf{I}_p\right\}$$
$$\times\frac{\partial vec}{\partial\boldsymbol{\theta}'}\left\{\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}\right\}.$$

In order to compute the expected value of $-\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi})$ conditional on $\mathbf{z}_t$ and I_{t-1}, the only extra element that we need are the fourth moments of the uniform distribution on the unit sphere surface in $\mathbb{R}^N$, which are given by

$$E\left[vec(\mathbf{u}_t\mathbf{u}_t')vec'(\mathbf{u}_t\mathbf{u}_t)\right]=E\left(\mathbf{u}_t\mathbf{u}_t'\otimes\mathbf{u}_t\mathbf{u}_t'\right)$$
$$=\frac{1}{N(N+2)}\left[(\mathbf{I}_{N^2}+\mathbf{K}_{NN})+vec(\mathbf{I}_N)vec'(\mathbf{I}_N)\right],$$

where we have used the fact that $\mathbf{u}_t\zeta_t=\boldsymbol{\varepsilon}_t^{\dagger}$ say, is a spherical multivariate normal random vector whose fourth moment can be found in Balestra and Holly (1990), and $E(\zeta_t^2)=N(N+2)$. Tedious but otherwise simple calculations show that we are eventually left with

$$E\left[-\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}_0)\right|\mathbf{z}_t,I_{t-1};\boldsymbol{\phi}_0]=\frac{\nu_0\,(N+\nu_0)}{(\nu_0-2)\,(N+\nu_0+2)}\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$$
$$\times\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}+\frac{(N+\nu_0)}{2(N+\nu_0+2)}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}$$
$$\times\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\otimes\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'}$$
$$-\frac{1}{2(N+\nu_0+2)}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}}$$
$$\times vec\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]vec'\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}_0)\right]\frac{\partial vec\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)\right]}{\partial\boldsymbol{\theta}'},$$

which converges to the usual expression under Gaussianity.

Similarly, we can write $\mathbf{h}_{\boldsymbol{\theta}\eta t}(\boldsymbol{\phi})$ as

$$\frac{\partial^2 g(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}\partial\eta}=\frac{\partial\boldsymbol{\mu}_t'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\frac{N+2-\varsigma_t(\boldsymbol{\phi})}{\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\phi})\right]^2}$$
$$+\frac{1}{2}\frac{\partial vec'\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}\left[\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\otimes\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{\theta})\right]$$
$$\times vec\left\{\frac{N+2-\varsigma_t(\boldsymbol{\phi})}{\left[1-2\eta+\eta\varsigma_t(\boldsymbol{\phi})\right]^2}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta})\right\}.$$

The expected value of the first term (conditional on $\mathbf{z}_t$ and I_{t-1}) is clearly zero when evaluated at $\boldsymbol{\phi}_0$ because $\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)$ is proportional to $\mathbf{u}_t$. As for the second term, we can show that

$$\begin{aligned}
& \frac{N+2-\varsigma_t(\boldsymbol{\phi}_0)}{[1-2\eta+\eta\varsigma_t(\boldsymbol{\phi}_0)]^2}\boldsymbol{\varepsilon}_t^*(\boldsymbol{\theta}_0)\boldsymbol{\varepsilon}_t^{*\prime}(\boldsymbol{\theta}_0) \\
= & \frac{\nu_0^2}{\nu_0-2}\left[(N+2)\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right)-(N+\nu_0)\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right)^2\right]\mathbf{u}_t\mathbf{u}_t',
\end{aligned}$$

whose expected value is

$$\frac{2\,(N+2)\,\nu_0^2}{(\nu_0-2)\,(\nu_0+N)\,(N+\nu_0+2)}\mathbf{I}_N,$$

which clearly goes to 0 as $\nu_0 \to \infty$.

Finally, let us look at the term

$$\begin{aligned}
\frac{\partial^2 g\left[\varsigma_t(\boldsymbol{\theta}_0),\eta\right]}{\partial\eta^2} &= -\frac{1}{\eta_0^3}\ln\left[1+\frac{\eta_0}{1-2\eta_0}\varsigma_t(\boldsymbol{\theta}_0)\right] \\
&+\frac{1}{\eta_0^2}\frac{\varsigma_t(\boldsymbol{\theta}_0)}{\left[1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)\right](1-2\eta_0)} \\
&-\frac{2\,(N\eta_0+1)\,\varsigma_t(\boldsymbol{\theta}_0)}{(1-2\eta_0)\,\eta_0\left[1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)\right]^2} \\
&-\frac{N\eta_0+1}{2\eta_0}\frac{(4\eta_0-1)\,\varsigma_t^2(\boldsymbol{\theta}_0)}{(1-2\eta_0)^2\left[1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)\right]^2}.
\end{aligned}$$

Taking expectations element by element we get

$$\begin{aligned}
-\frac{1}{\eta_0^3}E\left\{\ln\left[1+\frac{\eta_0\varsigma_t(\boldsymbol{\theta}_0)}{1-2\eta_0}\right]\right\} &= \nu_0^3 E\left[\ln\left(\frac{\xi_t}{\xi_t+\zeta_t}\right)\right] \\
&= \nu_0^3\left[\psi\left(\frac{\nu_0}{2}\right)-\psi\left(\frac{\nu_0+N}{2}\right)\right],
\end{aligned}$$

$$\begin{aligned}
E\left\{\frac{1}{\eta_0^2}\frac{\varsigma_t(\boldsymbol{\theta}_0)}{\left[1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)\right](1-2\eta_0)}\right\} &= \frac{\nu_0^4}{(\nu_0-2)}E\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right) \\
&= \frac{\nu_0^4 N}{(\nu_0-2)\,(N+\nu_0)},
\end{aligned}$$

$$\begin{aligned}
& -E\left\{\frac{2\,(N\eta_0+1)\,\varsigma_t(\boldsymbol{\theta}_0)}{(1-2\eta_0)\,\eta_0\left[1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)\right]^2}\right\} \\
&= -\frac{2(N+\nu_0)\nu_0^3}{(\nu_0-2)}E\left[\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right)-\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right)^2\right] \\
&= -\frac{2\nu_0^4 N}{(\nu_0-2)^2\,(N+\nu_0+2)},
\end{aligned}$$

and

$$-\frac{N\eta_0+1}{2}E\left(\frac{(4\eta_0-1)\,\varsigma_t^2(\boldsymbol{\theta}_0)}{(1-2\eta_0)^2\,\eta_0\,[1-2\eta_0+\eta_0\varsigma_t(\boldsymbol{\theta}_0)]^2}\right)$$
$$=\frac{1}{2}\frac{\nu_0^3\,(\nu_0-4)\,(N+\nu_0)}{(\nu_0-2)^2}E\left(\frac{\zeta_t}{\xi_t+\zeta_t}\right)^2$$
$$=\frac{1}{2}\frac{\nu_0^3\,(\nu_0-4)}{(\nu_0-2)^2}\frac{N\,(N+2)}{N+\nu_0+2}.$$

If we now add the expression for $\partial^2 c(\eta_0)/\partial\eta^2$, it is clear that the terms of order η^{-3} vanish, and the rest is equal to

$$\frac{\nu_0^4}{4}\left[\psi'\left(\frac{N+\nu_0}{2}\right)-\psi'\left(\frac{\nu_0}{2}\right)\right]$$
$$+\frac{\nu_0^3 N}{(\nu_0-2)}\left[\frac{\nu_0}{(N+\nu_0)}-\frac{2\nu_0}{(N+\nu_0+2)\,(\nu_0-2)}\right.$$
$$\left.+\frac{1}{2}\frac{(\nu_0-4)\,(N+2)}{(\nu_0-2)\,(N+\nu_0+2)}-\frac{(\nu_0-4)}{2(\nu_0-2)}\right]$$
$$=\frac{\nu_0^4}{4}\left[\psi'\left(\frac{N+\nu_0}{2}\right)-\psi'\left(\frac{\nu_0}{2}\right)\right]$$
$$+\frac{\nu_0^4 N}{(\nu_0-2)^2\,(N+\nu_0+2)}\left[\nu_0-\frac{4(N+2)}{(N+\nu_0)}\right].$$

Proposition 2

Most of the expressions can be obtained by simply taking limits as $\nu_0\to\infty$ of the formula for the conditional information matrix in Proposition 1. Nevertheless, the expression for $V\left[s_{\eta t}(\boldsymbol{\theta}_0,0)|\mathbf{z}_t,I_{t-1},\boldsymbol{\theta}_0,0\right]$ is in fact easier to derive by taking expected values of the outer product of the score. Specifically, we can show that

$$E\left[s_{\eta t}^2(\boldsymbol{\theta}_0,0)|\mathbf{z}_t,I_{t-1},\boldsymbol{\theta}_0,0\right]=\frac{N^2(N+2)^2}{16}$$
$$-\frac{N(N+2)^2}{4}E\left[\varsigma_t(\boldsymbol{\theta}_0)|\mathbf{z}_t,I_{t-1},\boldsymbol{\theta}_0,0\right]$$
$$+\left[\frac{(N+2)^2}{5}+\frac{N(N+2)}{8}\right]E\left[\varsigma_t^2(\boldsymbol{\theta}_0)|\mathbf{z}_t,I_{t-1},\boldsymbol{\theta}_0,0\right]$$
$$-\frac{N+2}{4}E\left[\varsigma_t^3(\boldsymbol{\theta}_0)|\mathbf{z}_t,I_{t-1},\boldsymbol{\theta}_0,0\right]$$
$$+\frac{1}{16}E\left[\varsigma_t^4(\boldsymbol{\theta}_0)|\mathbf{z}_t,I_{t-1},\boldsymbol{\theta}_0,0\right]=\frac{N(N+2)}{2},$$

where we have used the fact that under the null $\varsigma_t(\boldsymbol{\theta}_0)$ is an *i.i.d.* chi-square variate with N degrees of freedom (see Lemma 1), whose uncentred moment of integer order r is

$$E(\zeta_t^r)=2^r\left(r-1+\frac{N}{2}\right)\left(r-2+\frac{N}{2}\right)\cdots\left(1+\frac{N}{2}\right)\frac{N}{2}$$

(Mood, Graybill and Boes 1973).

Proposition 3

The expressions for conditional first and second moment of $s_{\eta t}(\boldsymbol{\theta}_0, 0)$ given $\mathbf{z}_t, I_{t-1}$ and $\boldsymbol{\phi}_0$ are obtained as in the proof of Proposition 2, except for the fact that under the alternative, $\varsigma_t(\boldsymbol{\theta}_0)$ is proportional to an *i.i.d.* F variate with N and ν_0 degrees of freedom (see Lemma 1), whose uncentred moment of integer order $r < \nu_0/2$ is

$$\begin{aligned} E\left[\frac{\zeta_t/N}{\xi_t/\nu_0}\right]^r &= \left(\frac{\nu_0}{N}\right)^r \frac{r-1+N/2}{-1+\nu_0/2}\frac{r-2+N/2}{-2+\nu_0/2}\cdots \\ &\times\frac{1+N/2}{-r+1+\nu_0/2}\frac{N/2}{-r+\nu_0/2} \end{aligned}$$

(Mood et al. 1973). Therefore, the restriction $\nu_0 > 8$ guarantees that the fourth moments of $\varsigma_t(\boldsymbol{\theta}_0)$ are bounded. Finally, the asymptotic distribution is obtained as a direct application of the Lindeberg-Levy central limit theorem for independent and identically distributed observations.

B SERIES EXPANSIONS IN TERMS OF η OF THE LOG-LIKELIHOOD FUNCTION

Let us start with the term:

$$\begin{aligned} c(\eta) &= \ln\left[\Gamma\left(\frac{N\eta+1}{2\eta}\right)\right] - \ln\left[\Gamma\left(\frac{1}{2\eta}\right)\right] \\ &-\frac{N}{2}\ln\left(\frac{1-2\eta}{\eta}\right) - \frac{N}{2}\ln\pi, \end{aligned}$$

whose first three derivatives are given by

$$\begin{aligned} \frac{\partial c(\eta)}{\partial\eta} &= \frac{N}{2\eta\,(1-2\eta)} - \frac{1}{2\eta^2}\left[\psi\left(\frac{N\eta+1}{2\eta}\right) - \psi\left(\frac{1}{2\eta}\right)\right], \\ \frac{\partial^2 c(\eta)}{\partial\eta^2} &= \frac{2N}{\eta\,(1-2\eta)^2} - \frac{1}{\eta^3}\left[\psi\left(\frac{1}{2\eta}\right) - \psi\left(\frac{N\eta+1}{2\eta}\right)\right] \\ &+\frac{1}{4\eta^4}\left[\psi'\left(\frac{N\eta+1}{2\eta}\right) - \psi'\left(\frac{1}{2\eta}\right)\right] - \frac{N}{2\eta^2(1-2\eta)^2}, \\ \frac{\partial^3 c(\eta)}{\partial\eta^3} &= \frac{3}{\eta^4}\left[\psi\left(\frac{1}{2\eta}\right) - \psi\left(\frac{N\eta+1}{2\eta}\right)\right] + \frac{12N}{\eta(1-2\eta)^3} \\ &-\frac{3}{2\eta^5}\left[\psi'\left(\frac{N\eta+1}{2\eta}\right) - \psi'\left(\frac{1}{2\eta}\right)\right] - \frac{6N}{\eta^2\,(1-2\eta)^3} \\ &-\frac{1}{8\eta^6}\left[\psi''\left(\frac{N\eta+1}{2\eta}\right) - \psi''\left(\frac{1}{2\eta}\right)\right] + \frac{N}{\eta^3\,(1-2\eta)^3}, \end{aligned}$$

where $\psi(x)$, $\psi'(x)$ and $\psi''(x)$ are the di-, tri-, and cuatri-gamma functions respectively. If we take limits as $\eta \to 0$ from above, we can show that

$$\begin{aligned}
\lim_{\eta\to0^+} c(\eta) &= -\frac{N}{2}\ln 2\pi, \\
\lim_{\eta\to0^+} \frac{\partial c(\eta)}{\partial\eta} &= \frac{N(N+2)}{4}, \\
\lim_{\eta\to0^+} \frac{\partial^2 c(\eta)}{\partial\eta^2} &= -\frac{N(N+2)(N-5)}{6}, \\
\lim_{\eta\to0^+} \frac{\partial^3 c(\eta)}{\partial\eta^3} &= \frac{N(N+2)(N^2-6N+16)}{4},
\end{aligned}$$

so that we finally obtain

$$\begin{aligned}
c(\eta) &= -\frac{N}{2}\ln 2\pi + \frac{N(N+2)}{4}\eta - \frac{1}{2}\frac{N(N+2)(N-5)}{6}\eta^2 \\
&\quad +\frac{1}{6}\frac{N(N+2)(N^2-6N+16)}{4}\eta^3 + O(\eta^4).
\end{aligned}$$

Our experience with $N = 1$ suggests that $c(\eta)$ and its derivatives can be accurately computed by their original expressions when $\eta > 8 * 10^{-4}$, but that the Taylor expansions are more reliable for smaller values.

Similarly,

$$\begin{aligned}
g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right] = -\frac{1}{2}\varsigma_t(\boldsymbol{\theta}) &+ \left[-\frac{N+2}{2}\varsigma_t(\boldsymbol{\theta}) + \frac{1}{4}\varsigma_t^2(\boldsymbol{\theta})\right]\eta \\
&+\frac{1}{2}\left[-2(N+2)\varsigma_t(\boldsymbol{\phi}) + \frac{N+4}{2}\varsigma_t^2(\boldsymbol{\phi}) - \frac{1}{3}\varsigma_t^3(\boldsymbol{\phi})\right]\eta^2 \\
&+\frac{1}{6}\left[\begin{array}{c} -12(2+N)\varsigma_t(\boldsymbol{\theta}) + 6(N+3)\varsigma_t^2(\boldsymbol{\theta}) \\ -(6+N)\,\varsigma_t^3(\boldsymbol{\theta}) + \frac{1}{8}\varsigma_t^4(\boldsymbol{\theta}) \end{array}\right]\eta^3 \\
&+\frac{1}{24}\left[\begin{array}{c} -96\,(N+2)\,\varsigma_t(\boldsymbol{\theta}) + 24\,(8+3\,N)\,\varsigma_t^2(\boldsymbol{\theta}) \\ -24\,(N+4)\,\varsigma_t^3(\boldsymbol{\theta}) + 3\,(N+8)\,\varsigma_t^4(\boldsymbol{\theta}) - \frac{12}{5}\varsigma_t^5(\boldsymbol{\theta}) \end{array}\right]\eta^4 \\
&+\frac{1}{120}\left[\begin{array}{c} -960\,(N+2)\,\varsigma_t(\boldsymbol{\theta}) + 600\,(2N+5)\,\varsigma_t^2(\boldsymbol{\theta}) \\ -1440\,(3N+10)\,\varsigma_t^3(\boldsymbol{\theta}) + \\ 120\,(N+5)\,\varsigma_t^4(\boldsymbol{\theta}) \\ -12\,(N+10)\,\varsigma_t^5(\boldsymbol{\theta}) + 10\varsigma_t^6(\boldsymbol{\theta}) \end{array}\right]\eta^5 + O(\eta^6).
\end{aligned}$$

It is important to mention that the above expression is only guaranteed to provide a good approximation if in addition $\varsigma_t(\boldsymbol{\phi})$ is not excessively large. In our experience, $g\left[\varsigma_t(\boldsymbol{\theta}),\eta\right]$ and its derivatives can be accurately evaluated with the analytical expressions in Section 2 when $\eta > .03$ or $\eta\varsigma_t(\boldsymbol{\theta}) >.001$, but otherwise the Taylor expansions are more reliable.

References

Abramowitz, M., and Stegun, I.A. (1964), *Handbook of Mathematic Functions*, AMS 55, National Bureau of Standards.

Andrews, D.W.K. (1999), "Estimation When a Parameter in on the Boundary", *Econometrica*, 67, 1341-1383.

—— (2001), "Testing When a Parameter is on the Boundary of the Maintained Hypothesis", *Econometrica*, 69, 683-734.

Balestra, P., and Holly, A. (1990), "A General Kronecker Formula for the Moments of the Multivariate Normal Distribution", DEEP Cahier 9002, University of Lausanne.

Barndorf-Nielsen, O. E., and Shephard, N. (2001), *Lévy-based Dynamic Models for Financial Economics*, in preparation, Nuffield College, Oxford.

Bauwens, L., and Laurent, S. (2002), "A New Class of Multivariate Skew Densities, with Application to GARCH Models", mimeo, CORE, Université Catholique de Louvain.

Bera, A.K., and McKenzie, C.R. (1987), "Additivity and Separability of Lagrange Multiplier, Likelihood Ratio and Wald Tests", *Journal of Quantitative Economics*, 3, 53-63.

Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *Review of Economics and Statistics*, 69, 542-547.

Bollerslev, T., Chou, R.Y., and Kroner, K.F. (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 52, 5-59.

Bollerslev, T., and J. M. Wooldridge (1992), "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances", *Econometric Reviews*, 11, 143-172.

Calzolari, G., Fiorentini, G., and Sentana, E. (2003), "Constrained Indirect Inference Estimation", mimeo, CEMFI, Madrid.

Crowder, M.J. (1976), "Maximum Likelihood Estimation for Dependent Observations", *Journal of the Royal Statistical Society*, Ser. B, 38, 45-53.

Davidson R., and MacKinnon, J.G. (1983), "Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test", *Economics Letters,* 12, 269-275.

—— (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

—— (1998), "Graphical Methods for Investigating the Size and Power of Tests Statistics", *The Manchester School*, 66, 1-26.

Demos, A., and Sentana, E. (1998), "Testing for GARCH Effects: A One-sided Approach", *Journal of Econometrics*, 86, 97-127.

Diebold, F.X., Gunther, A.G., and Tay, A.S. (1998), "Evaluating Density Forecasts, With Applications to Financial Risk Management" *International Economic Review* 39, 863-883.

Doornik, J.A., and Hansen, H. (1994), "An Omnibus Test for Univariate and Multivariate Normality", Working Paper W4&91, Nuffield College, Oxford.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1007.

Fang, K.-T., Kotz, S., and Ng, K.-W. (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall.

Fiorentini, G., Calzolari, G., and Panattoni, L. (1996), "Analytical Derivatives and the Computation of GARCH Models", *Journal of Applied Econometrics,* 11, 399-417.

Fiorentini, G., Sentana, E., and Calzolari, G. (2003), "On the Validity of the Jarque-Bera Normality Test in Conditionally Heteroskedastic Dynamic Regression Models", mimeo, CEMFI.

Gallant, A.R., and Tauchen, G. (1996), "Which Moments to Match?" *Econometric Theory*, 12, 657-681.

Gourieroux C., Holly A., and Monfort A. (1980), "Kuhn-Tucker, Likelihood Ratio and Wald Tests for Nonlinear Models with Inequality Constraints on the Parameters", Harvard Institute of Economic Research Discussion Paper 770.

Hall, A. (1990), "Lagrange Multiplier Tests for Normality Against Seminonparametric Alternatives", *Journal of Business and Economic Statistics*, 8, 417-426.

Harvey, A., Ruiz, E., and Sentana, E. (1992), "Unobservable Component Time Series Models with ARCH Disturbances", *Journal of Econometrics*, 52, 129-158.

Hentschel, L. (1995), "All in the Family: Nesting Symmetric and Asymmetric GARCH Models", *Journal of Financial Economics*, 39, 71-104.

Johnson, N.L. (1949), "Systems of Curves Generated by Methods of Translation", *Biometrika*, 36, 149-176.

Kiefer, N.M., and Salmon, M. (1983), "Testing Normality in Econometric Models", *Economic Letters*, 11, 123-127.

Kilian, L., and Demiroglu, U. (2000), "Residual-based Test for Normality in Autoregressions: Asymptotic Theory and Simulation Evidence", *Journal of Business and Economic Statistics*, 18, 40-50.

Jarque, C.M., and Bera, A.K. (1980), "Efficient Tests for Normality, Heteroskedasticity, and Serial Independence of Regression Residuals", *Economic Letters*, 6, 255-259.

Lange, K.L., Little, R.J.A., and Taylor, J.M.G. (1989), "Robust Satistical Modeling Using the t Distribution", *Journal of the American Statistical Association*, 84, 881-896.

Lee, J.H.H., and King, M.L. (1993), "A Locally Most Mean Powerful Based Score Test for ARCH and GARCH Regression Disturbances", *Journal of Business and Economic Statistics*, 11, 17-27.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, (2nd ed.), Springer-Verlag.

Magnus, J.R., and Neudecker, H. (1988), *Matrix Differential Calculus With Applications in Statistics and Econometrics*, Wiley, Chichester.

Mardia, K.V. (1970), "Measures of Multivariate Skewness and Kurtosis With Applications", *Biometrika* 57, 519-530.

McCullough, B. D., and Vinod, H.D. (1999), "The Numerical Reliability of Econometric Software", *Journal of Economic Literature*, 37, 633-665.

Mood, A.M., Graybill, F.A., and Boes, D.C. (1974), *Introduction to the Theory of Statistics*, (3rd ed.), McGraw Hill.

Numerical Algorithms Group (2001), *NAG Fortran 77 Library Mark 19 Reference Manual*, Author.

Newey, W.K., and Steigerwald, D.G. (1997), "Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models", *Econometrica*, 65, 587-99.

Pesaran, M.H., and Pesaran, B. (1997), *Working With Microfit 4.0: Interactive Econometric Analysis*, Oxford University Press.

Sentana, E. (1991), "Quadratic ARCH Models: A Potential Re-interpretation of ARCH Models", LSE Financial Markets Group Discussion Paper 122, London School of Eocnomics.

—— (1995), "Quadratic ARCH Models", *Review of Economic Studies*, 62, 639-661.

—— (in press), "Factor Representing Portfolios in Large Asset Markets", *Journal of Econometrics*.

Table 1

Estimates of a Conditionally Heteroskedastic Single Factor Model for 26 U.K. Sectorial Indices
Monthly Excess Returns 1971:2-1990:10 (237 obs.)
Estimates of Dynamic Variance Parameters and Degrees of Freedom

$$\lambda_t = (1-\alpha-\beta)(1-\varrho^2)$$
$$+\alpha\left[\left(f_{t-1|t-1} - \sqrt{(1-\alpha-\beta)/\alpha\varrho}\right)^2 + \omega_{t-1|t-1}\right] + \beta\lambda_{t-1}$$
$$0 \le \beta \le 1-\alpha \le 1, \quad -1 \le \varrho \le 1$$

Parameter	Gaussian		Student t	
		s.e.		s.e.
α	.111	.075	.053	.026
β	.670	.258	.675	.120
ϱ	.951	.629	1.0	–
η	*0*		.103	.012
Log-likelihood	-4,471.216		-4,221.162	

Notes: $f_{t|t}$ denotes the Kalman filter estimate of the latent factor, and $\omega_{t|t}$ the associated conditional mean square error (Harvey et al. 1992). Standard errors (s.e.) are computed using analytical derivatives based on the expressions in Bollerslev and Wooldridge (1992) in the Gaussian case, and Proposition 1 in the case of the t.

FIGURE 1: Derivative of $c(\eta)$ with respect to η and Taylor expansion

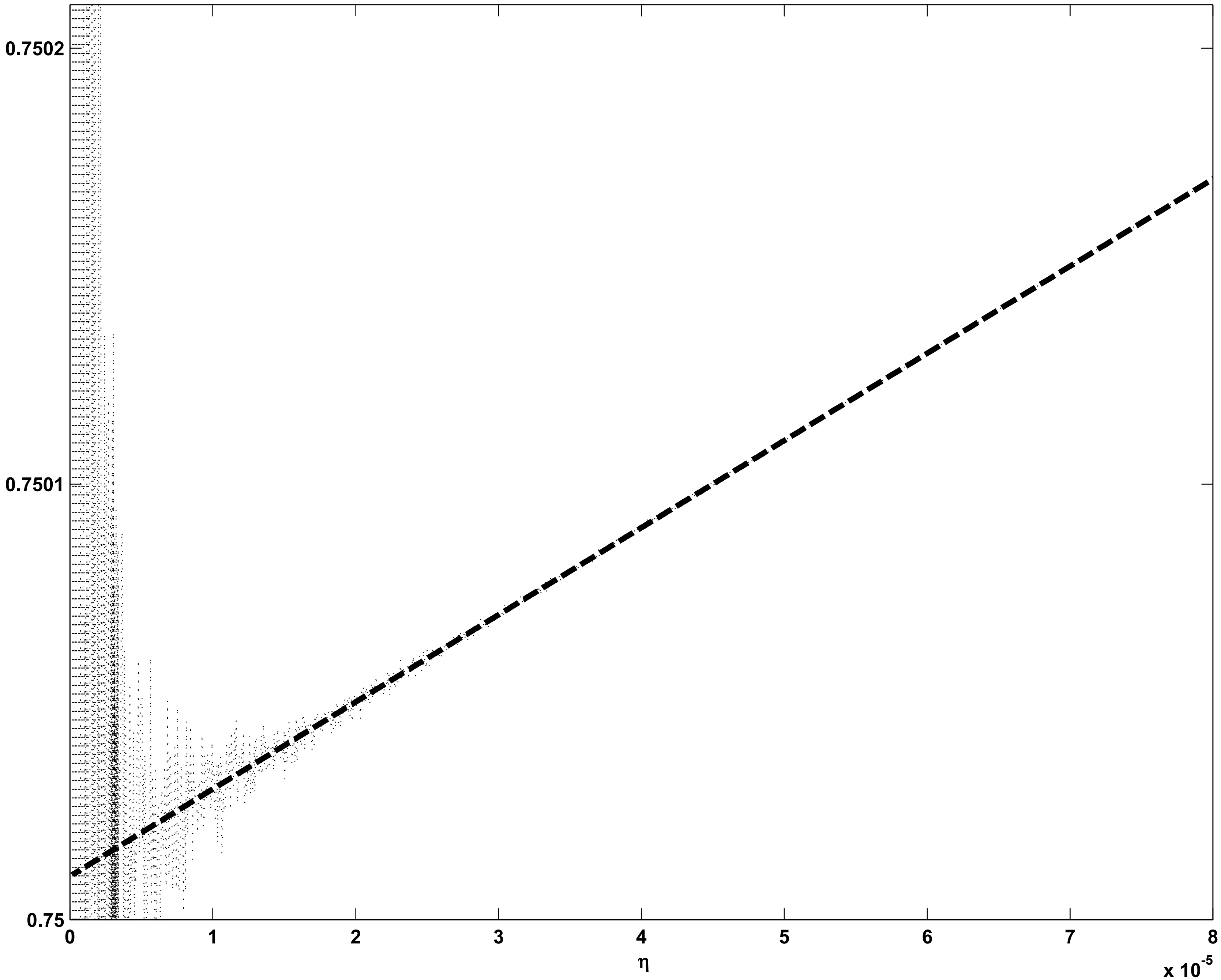


FIGURE 2-A: Power of the LM test (T=100, α=5%).

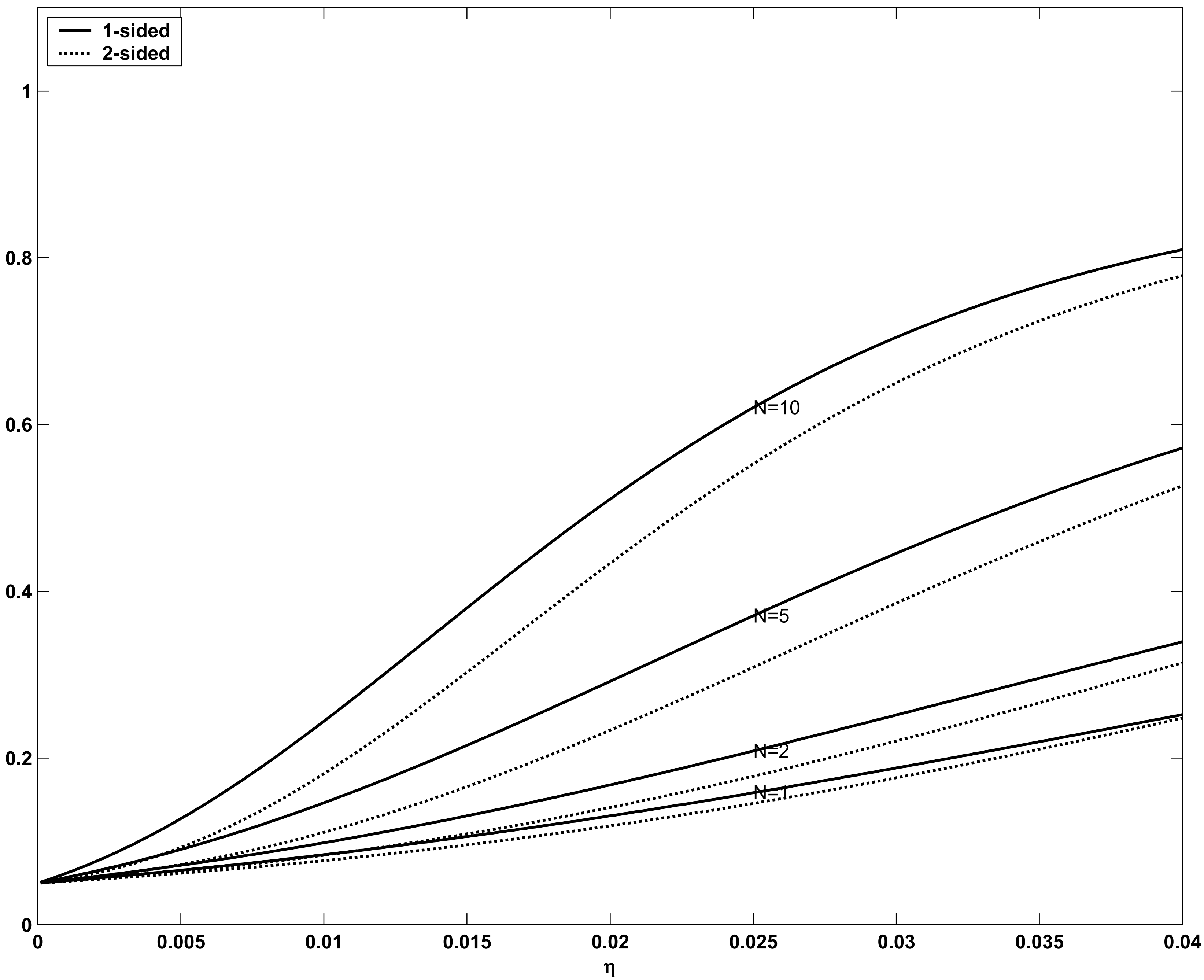


FIGURE 2-B: Power of the LM test (T=250, α=5%).

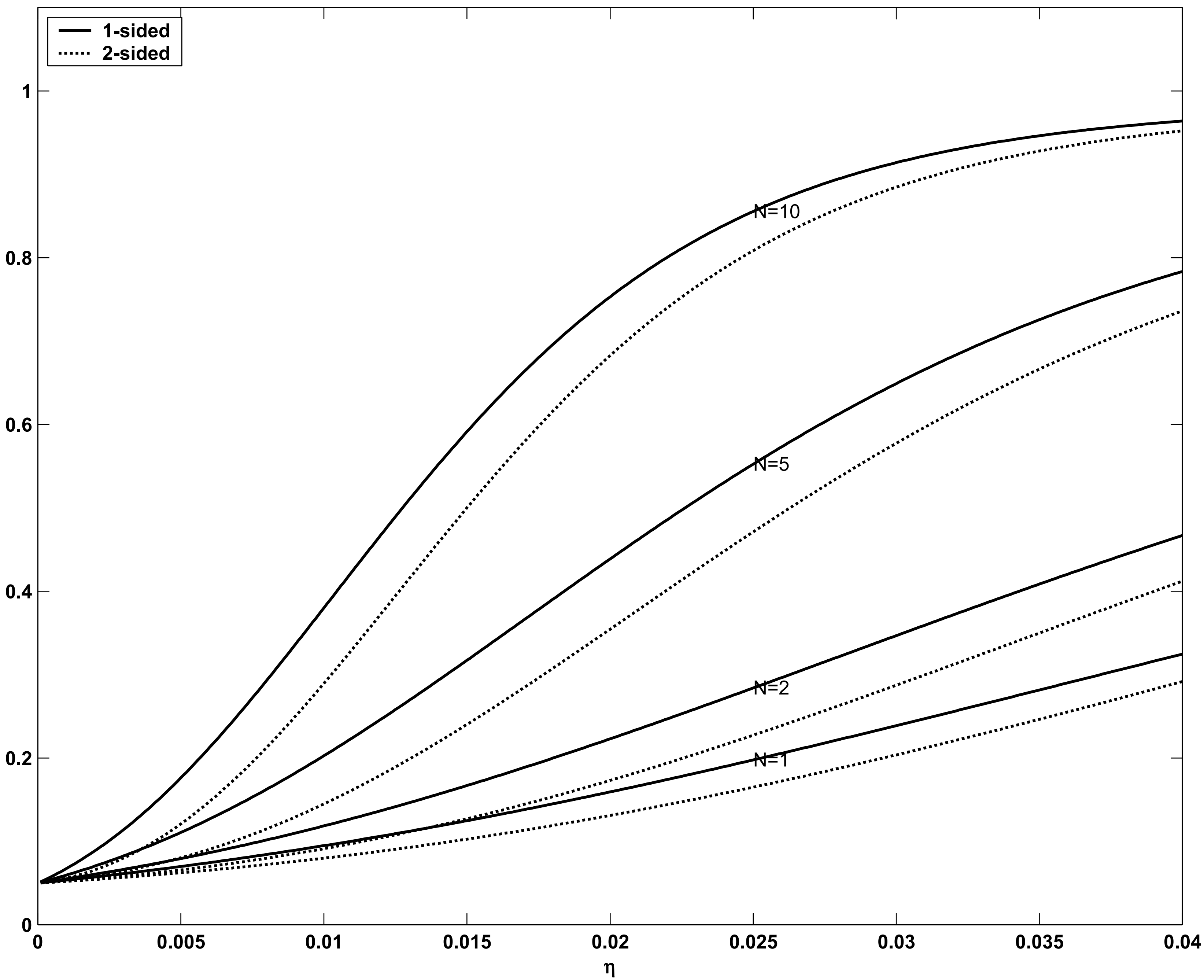


FIGURE 2-C: Power of the LM test (T=500, α=5%).

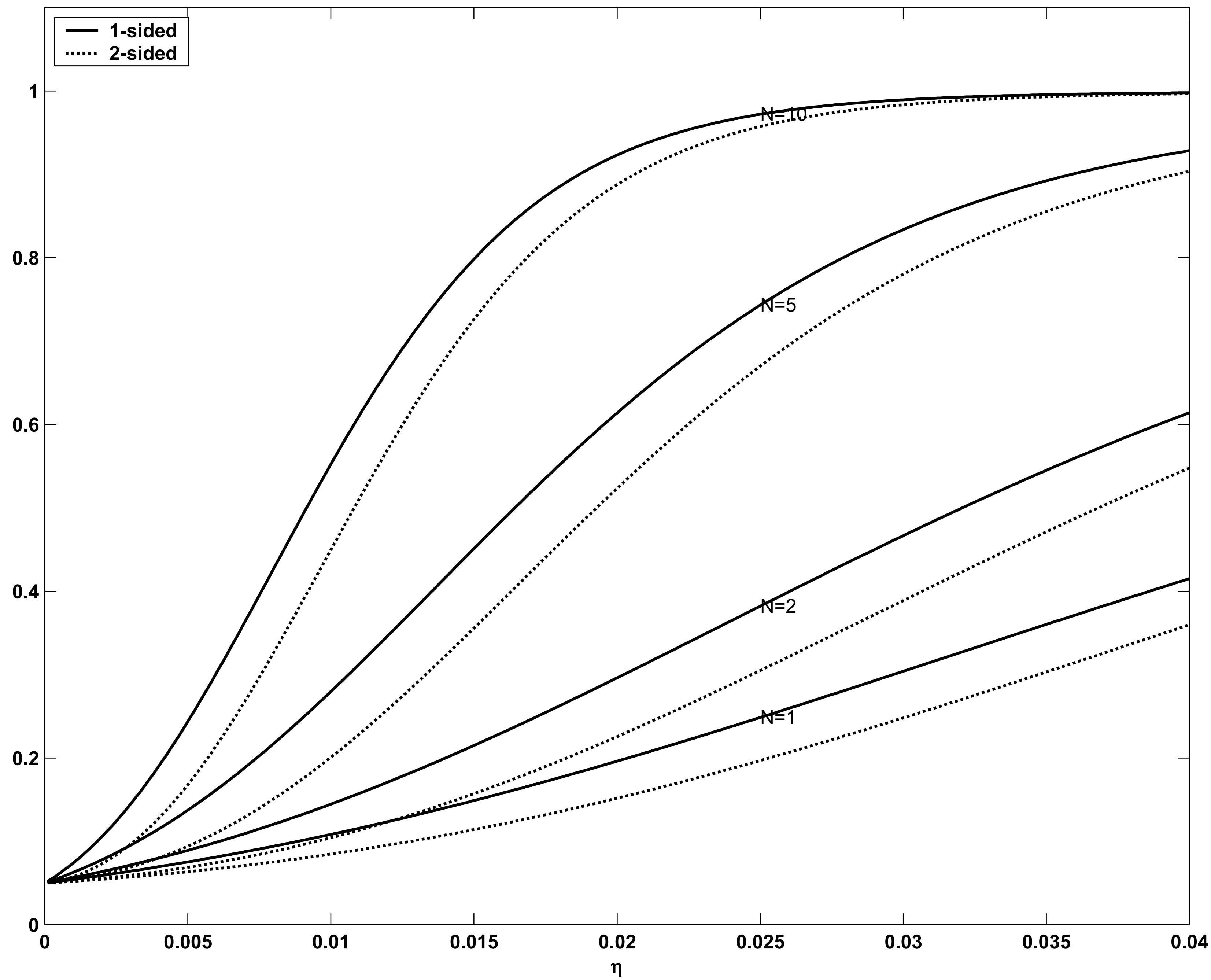
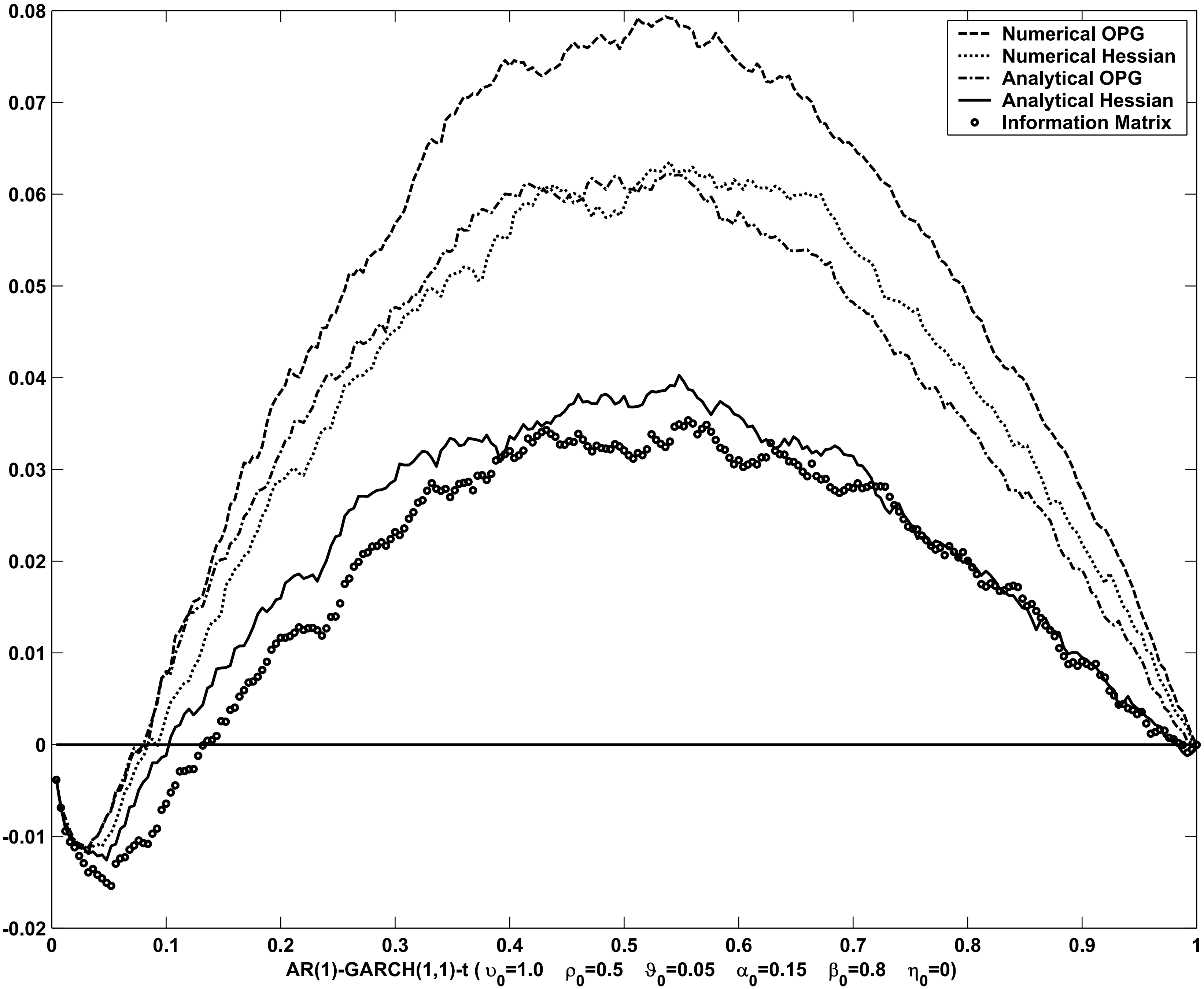


FIGURE 3-A: P-value discrepancy plot for Wald-test $\phi=\phi_0$ T=1000 Rep.=10000



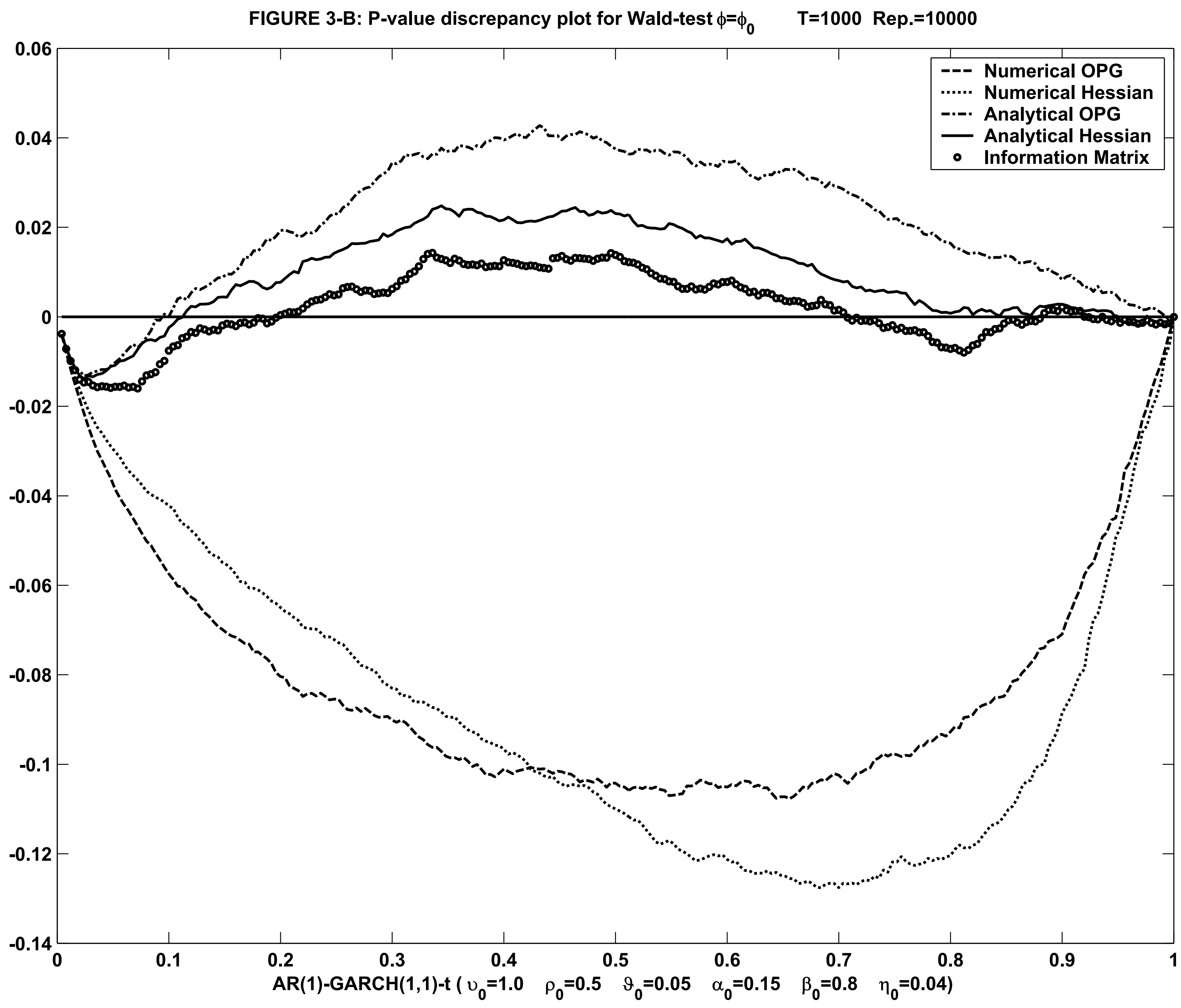


FIGURE 3-B: P-value discrepancy plot for Wald-test φ=φ0 T=1000 Rep.=10000
Numerical OPG
Numerical Hessian
Analytical OPG
Analytical Hessian
Information Matrix
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
-0.14
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
AR(1)-GARCH(1,1)-t (υ0=1.0 ρ0=0.5 ϑ0=0.05 α0=0.15 β0=0.8 η0=0.04)

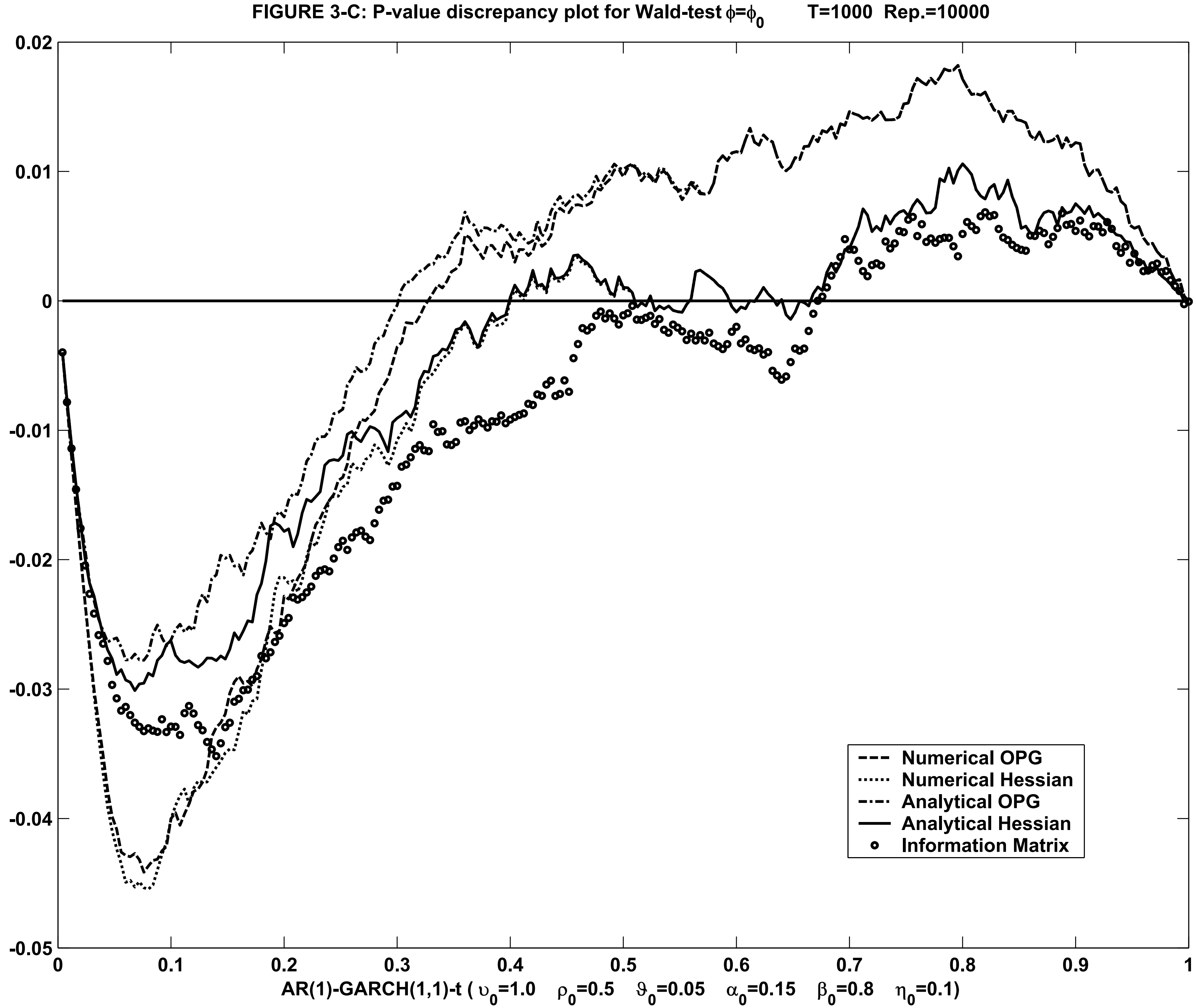
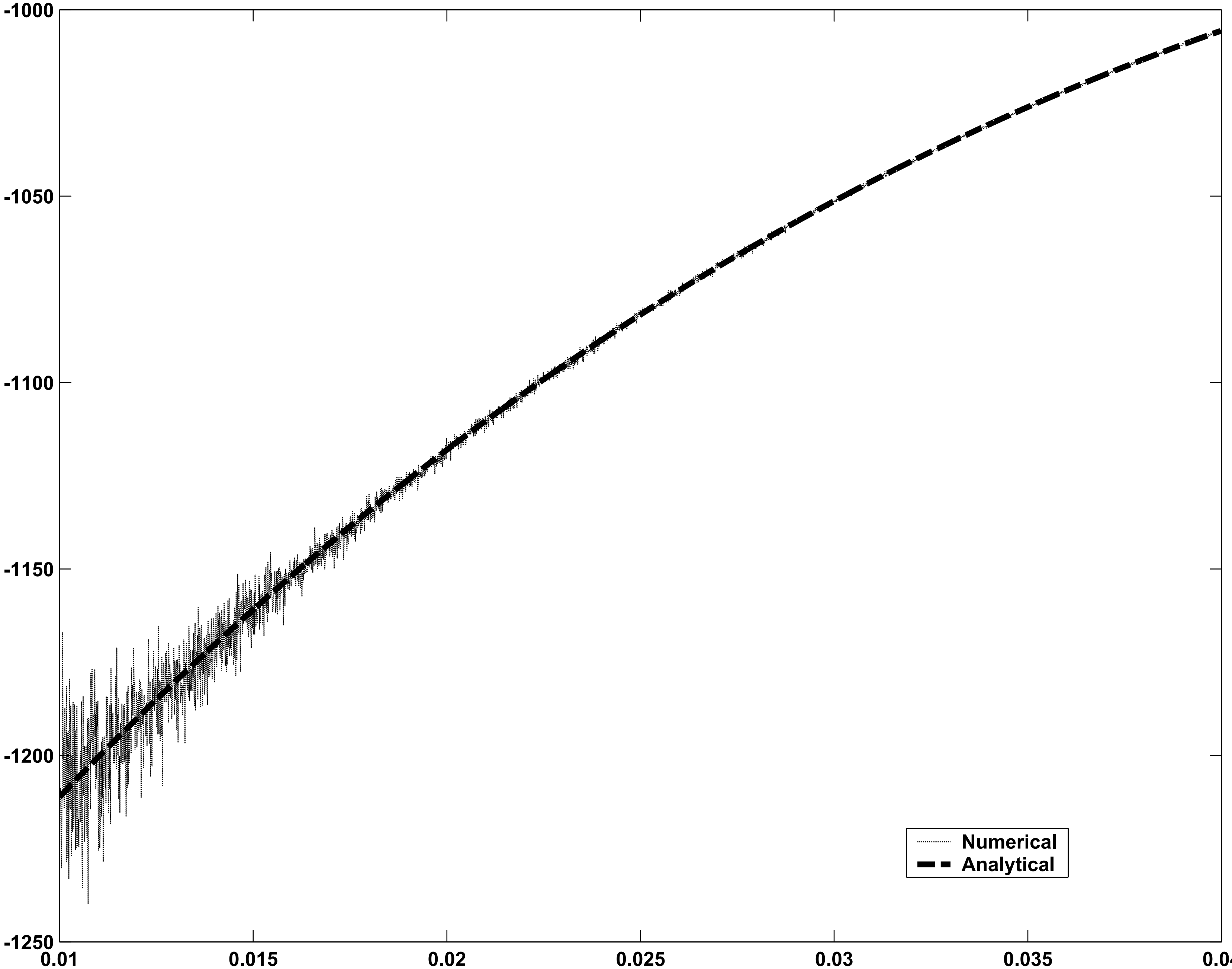


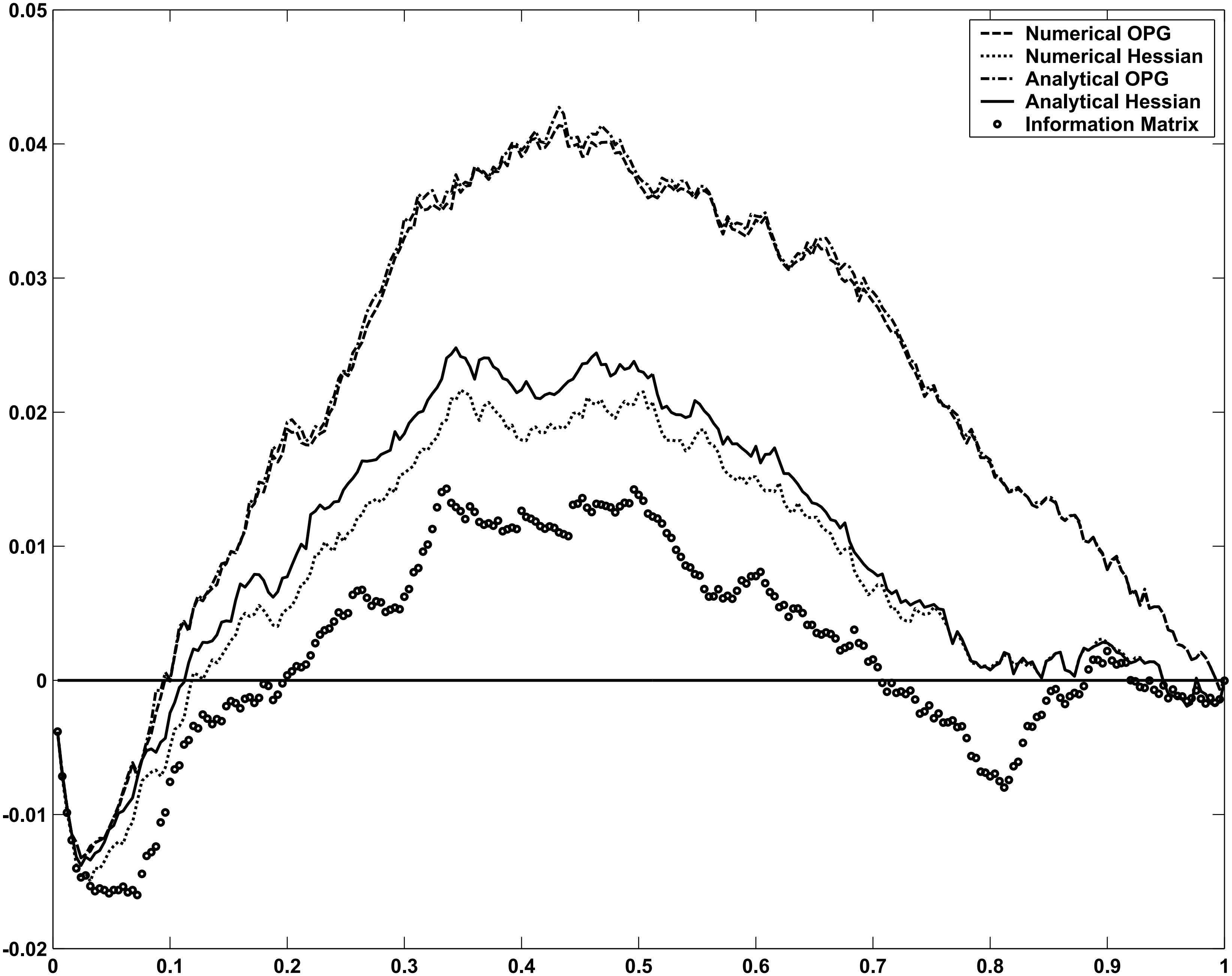
FIGURE 3-C: P-value discrepancy plot for Wald-test $\phi=\phi_0$ T=1000 Rep.=10000
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.05
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Numerical OPG
Numerical Hessian
Analytical OPG
Analytical Hessian
Information Matrix
AR(1)-GARCH(1,1)-t (υ_0=1.0 ρ_0=0.5 ϑ_0=0.05 α_0=0.15 β_0=0.8 η_0=0.1)

FIGURE 4: Second derivative of log-likelihood w.r.t. η T=1000



AR(1)-GARCH(1,1)-t (υ_0=1.0 ρ_0=0.5 ϑ_0=0.05 α_0=0.15 β_0=0.8 η_0=0.04)

FIGURE 5: P-value discrepancy plot for Wald-test $\phi=\phi_0$ T=1000 Rep.=10000



AR(1)-GARCH(1,1)-t (υ_0=1.0 ρ_0=0.5 ϑ_0=0.05 α_0=0.15 β_0=0.8 η_0=0.04)

FIGURE 6: Estimated conditional standard deviation of equally weighted portfolio

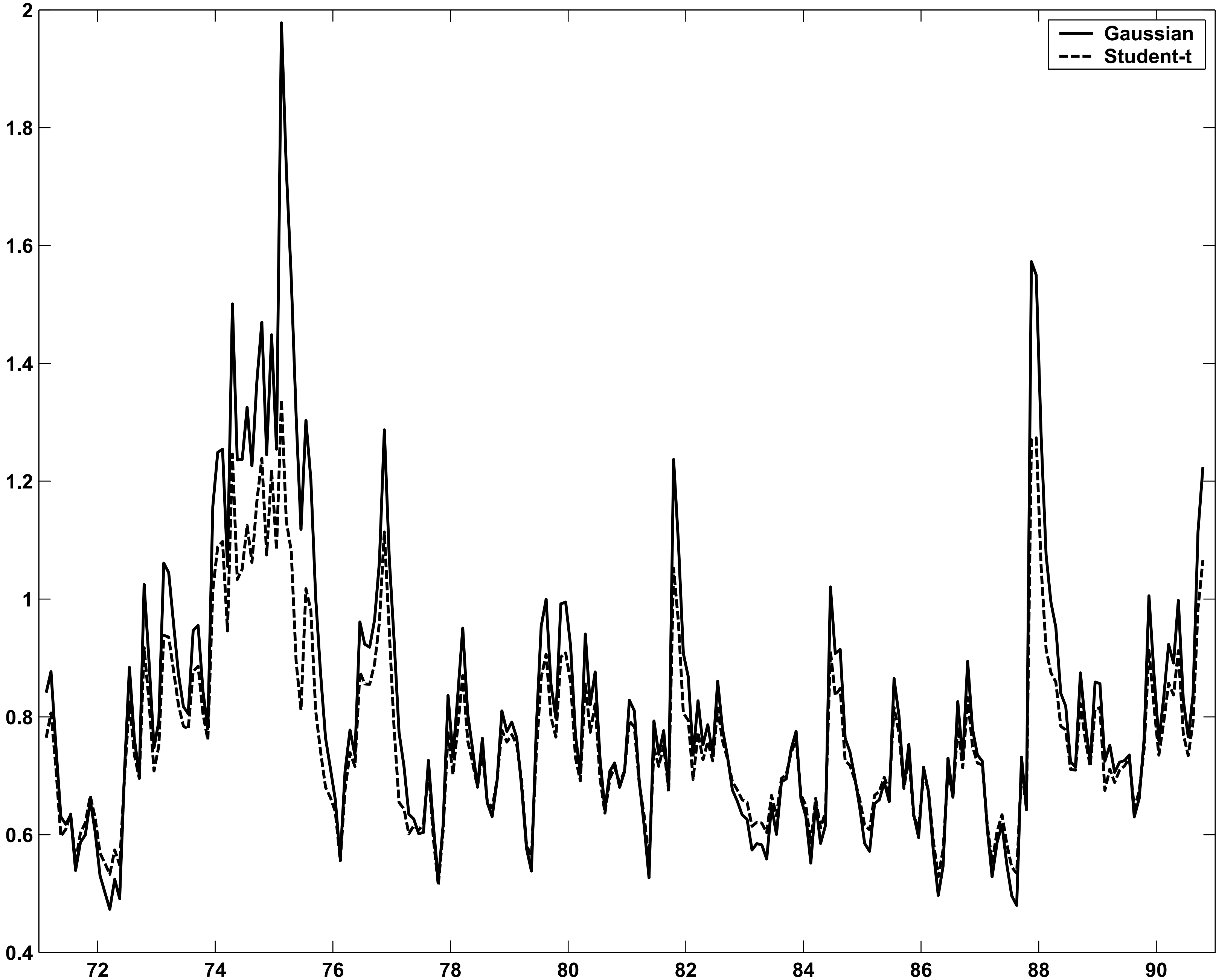


FIGURE 7: Discrepancy plot of the empirical c.d.f. of probability integral transform of returns on equally weighted portfolio