cemfi

working paper

1814

New Testing Approaches for Mean-Variance Predictability

Gabriele Fiorentini
Enrique Sentana

December 2018

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 1814
December 2018

New Testing Approaches for Mean-Variance Predictability

Abstract

We propose tests for smooth but persistent serial correlation in risk premia and volatilities that exploit the non-normality of financial returns. Our parametric tests are robust to distributional misspecification, while our semiparametric tests are as powerful as if we knew the true return distribution. Local power analyses confirm their gains over existing methods, while Monte Carlo exercises assess their finite sample reliability. We apply our tests to quarterly returns on the five Fama-French factors for international stocks, whose distributions are mostly symmetric and fat-tailed. Our results highlight noticeable differences across regions and factors and confirm the fragility of Gaussian tests.

JEL Codes: C12, C22, G17.

Keywords: Financial forecasting, moment tests, misspecification, robustness, volatility.

Gabriele Fiorentini
Università degli Studi di Firenze
gabriele.fiorentini@unifi.it

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

We would like to thank Dante Amengual, Christian Bontemps, Nour Meddahi, Javier Mencía and João Santos Silva, as well as audiences at Alicante, Bologna, Budapest School for Central Bank Studies, CEMFI, CREST, Crete, Erasmus, Essex, Università de la Svissera Italiana, the Toulouse Financial Econometrics Conference (May 2009), ASSET (October 2010), SAEe (December 2010) and SoFiE (June 2012) for helpful comments, discussions and suggestions. Of course, the usual caveat applies. Financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 and the Santander Research Chair at CEMFI (Sentana) is gratefully acknowledged.

1 Introduction

Most of the empirical literature assessing the predictability of the levels of financial returns has focused on the predictor variables. But despite hundreds of papers over several decades, the evidence remains controversial (see for example Spiegel (2008) and the references therein). The solomonic decision on the part of the 2013 Economic Sciences Nobel Prize Committee provides a case in point. While Fama, Hansen and Shiller agree on the relevance of the return predictability question, they do not necessarily agree on the answer.

There is of course much stronger evidence on time variation in volatilities at daily frequencies, but at the same time there is a widespread belief that those effects are irrelevant at monthly and quarterly frequencies. Nevertheless, theoretical and empirical considerations suggest that the movements in the first two moments of excess returns on financial assets, assuming that those movements are real, should be smooth and persistent.

Finally, many empirical studies indicate that regardless of the frequency of observation, the distribution of asset returns is rather leptokurtic, and possibly somewhat asymmetric. Still, most existing tests for predictability of the mean and volatility of asset returns ignore this fact by implicitly relying on normality.

In this context, we propose new testing approaches for mean-variance predictability that explicitly account for all those empirical regularities. Specifically, we propose tests for smooth but persistent serial correlation in asset risk premia and volatilities that exploit the non-normality of returns. In this sense, we consider both parametric tests that assume flexible non-normal distributions, and semiparametric tests.

For a given predictor variable, our tests differ from standard tests in that we effectively change the *regressand* in a manner that is reminiscent of the robust estimation literature. Thereby, we achieve two important improvements over the usual Gaussian tests: increases in their local power and reductions in their sensitivity to influential observations. Furthermore, we also transform the *regressor* to exploit the persistence of conditional means and variances.

Although we focus our discussion on Lagrange Multiplier (or score) tests, our results apply to Likelihood ratio and Wald tests, which are asymptotically equivalent under the null and sequences of local alternatives, and therefore share their optimality properties.

From the theoretical point of view, our most novel contribution is to show that our parametric tests remain valid regardless of whether or not the assumed distribution is correct, which puts them on par with the Gaussian pseudo-maximum likelihood (PML) testing procedures advocated by White (1982) and Bollerslev and Wooldridge (1992) among many others. We also show that our semiparametric tests should be as efficient as if we knew the true distribution of the data.

We present local power analyses that confirm the gains that our new testing approaches deliver over existing methods. We complement our theoretical results with detailed Monte Carlo exercises that study their finite sample reliability, as well as the ability of a straightforward non-parametric bootstrap procedure to improve it. Finally, we also illustrate our methods with an application to the five Fama and French (2015) (FF) factors for international stocks, which confirms the empirical relevance of our proposals.

The rest of the paper is organised as follows. We introduce our mean and variance predictability tests in sections 2 and 3, respectively, and study the power gains that they offer against local alternatives. Next, we study joint tests in section 4. A Monte Carlo evaluation of our proposed procedures can be found in section 5, followed by our empirical application in section 6. Finally, we present our conclusions in section 7. Proofs and auxiliary results are gathered in appendices.

2 Tests for predictability in mean

2.1 First order serial correlation tests

Although we can consider any predictor variable available at time $t-1$, for pedagogical reasons we initially develop tests of first order serial correlation under the maintained assumption that the conditional variance is constant. More specifically, the model under the alternative is

$$\left.\begin{array}{c} y_t = \pi(1-\rho) + \rho y_{t-1} + \sqrt{\omega}\varepsilon_t^*, \\ \varepsilon_t^* | I_{t-1}; \pi, \omega, \rho, \boldsymbol{\eta} \sim i.i.d.\ D(0,1,\boldsymbol{\eta}) \\ \text{with density function } f(.;\boldsymbol{\eta}) \end{array}\right\}, \tag{1}$$

where the parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, $\boldsymbol{\theta}' = (\boldsymbol{\theta}_s', \rho)'$ and $\boldsymbol{\theta}_s = (\pi, \omega)'$. In this context, the null hypothesis is $H_0 : \rho = 0$. Regardless of the specific parametric distribution, testing the null of white noise against first order serial correlation is extremely easy:

Proposition 1 *Let*

$$\bar{G}_m(l) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}); \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \epsilon_{t-l}(\boldsymbol{\theta}_{s0})$$

denote the sample cross moment of $\epsilon_{t-l}(\boldsymbol{\theta}_{s0})$ and the derivative of the conditional log density of ε_t^* with respect to its argument evaluated at $\epsilon_t(\boldsymbol{\theta}_{s0})$, where $\epsilon_t(\boldsymbol{\theta}_s) = \omega^{-1/2}(y_t - \pi)$ and $\boldsymbol{\theta}_{s0}$ the true parameter value.

1. *Under the null hypothesis $H_0 : \rho = 0$, the score test statistic*

$$LM_{AR(1)} = T \cdot \frac{\bar{G}_m^2(1)}{\mathcal{I}_{\rho\rho}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0)} \tag{2}$$

will be distributed as a χ^2 with 1 degree of freedom as T goes to infinity, where

$$\mathcal{I}_{\rho\rho}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}) = V[\epsilon_t(\boldsymbol{\theta}_s) | \boldsymbol{\theta}_s, 0, \boldsymbol{\eta}] \cdot M_{ll}(\boldsymbol{\eta})$$

and

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = V\left[\left.\frac{\partial \ln f(\varepsilon_t^*;\boldsymbol{\eta})}{\partial \varepsilon^*}\right| I_{t-1};\boldsymbol{\eta}\right]. \tag{3}$$

2. *This asymptotic distribution is unaffected if we replace the true values of the parameters* $\boldsymbol{\theta}_{s0}$ *or* $\boldsymbol{\eta}_0$ *by their maximum likelihood estimators under the null.*

The exact expression for $\bar{G}_m(1)$ depends on the assumed distribution. For example, in the standardised Student t case with η^{-1} degrees of freedom,

$$\frac{\partial \ln f(\varepsilon^*;\eta)}{\partial \varepsilon^*} = -\frac{\eta+1}{1-2\eta+\eta\varepsilon^{*2}}\varepsilon^*, \tag{4}$$

which reduces to $(-)\ \varepsilon^*$ under normality ($\eta = 0$). Similarly, for a standardised Kotz distribution with excess kurtosis parameter $\varkappa$, which also nest the normal when $\varkappa = 0$, it becomes

$$\frac{\partial \ln f(\varepsilon^*;\eta)}{\partial \varepsilon^*} = -\frac{1}{3\varkappa+2}\left(\frac{3\varkappa}{\varepsilon^*}+2\varepsilon^*\right),$$

which is a linear combination of the standardised residual and its reciprocal. On the other hand, a standardised Laplace (or double exponential) distribution, which does not depend on any additional parameters, yields

$$\frac{\partial \ln f(\varepsilon^*)}{\partial \varepsilon^*} = -\sqrt{2}sign(\varepsilon^*), \tag{5}$$

which means that (2) effectively becomes a directional prediction test, as in Christoffersen and Diebold (2006).

Similarly, in the case of a standardised two component mixture of normals with density function:

$$f(\varepsilon^*;\boldsymbol{\eta}) = \frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon^*-\mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right] + \frac{1-\lambda}{\sigma_2^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon^*-\mu_2^*(\boldsymbol{\eta})}{\sigma_2^*(\boldsymbol{\eta})}\right],$$

where $\phi(.)$ is the standard normal density, $\boldsymbol{\eta} = (\delta, \upsilon, \lambda)'$ are shape parameters, and $\mu_1^*(\boldsymbol{\eta})$, $\mu_2^*(\boldsymbol{\eta})$, $\sigma_1^{*2}(\boldsymbol{\eta})$ and $\sigma_2^{*2}(\boldsymbol{\eta})$ are defined in appendix C.1, the relevant regressand becomes

$$\frac{\partial \ln f(\varepsilon^*;\boldsymbol{\eta})}{\partial \varepsilon^*} = \frac{1}{\sigma_1^*(\boldsymbol{\eta})}\left[\frac{\varepsilon^*-\mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right] w(\varepsilon^*;\boldsymbol{\eta}) + \frac{1}{\sigma_2^*(\boldsymbol{\eta})}\left[\frac{\varepsilon^*-\mu_2^*(\boldsymbol{\eta})}{\sigma_2^*(\boldsymbol{\eta})}\right][1-w(\varepsilon^*;\boldsymbol{\eta})],$$

with

$$w(\varepsilon^*;\boldsymbol{\eta}) = \frac{\frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon^*-\mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right]}{\frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon^*-\mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right] + \frac{1-\lambda}{\sigma_2^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon^*-\mu_2^*(\boldsymbol{\eta})}{\sigma_2^*(\boldsymbol{\eta})}\right]},$$

so that it can be understood as the average of the standardised residuals for each component weighted by the posterior probabilities that the observation belongs to each of those components.

As for $\mathcal{M}_{ll}$, we can either use its theoretical expression (for instance $(1+\eta)(1-2\eta)^{-1}$ $\times(1+3\eta)^{-1}$ in the case of the Student t, or 1 under normality), compute the sample analogue

of (3), or exploit the information matrix equality and calculate it as the sample average of $-\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right] / \partial \varepsilon^* \partial \varepsilon^*$. As shown by Davidson and MacKinnon (1983) and many others, this choice will affect the finite sample properties of the tests, as well as their validity under distributional misspecification.

Intuitively, we can interpret the above score test as a moment test based on the following orthogonality condition:

$$E\left\{ \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}); \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \epsilon_{t-1}(\boldsymbol{\theta}_{s0}) \middle| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0 \right\} = 0, \tag{6}$$

which is related to the moment conditions used by Bontemps and Meddahi (2012) in their distributional tests.[1] Given that the score with respect to π under the null is proportional to

$$\frac{1}{T} \sum\nolimits_{t=1}^{T} \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s); \boldsymbol{\eta}]}{\partial \varepsilon^*},$$

the sample second moment will numerically coincide with the sample covariance if we evaluate the standardised residuals at the ML estimators. Hence, an asymptotically equivalent test under the null and sequences of local alternatives would be obtained as $T \cdot R^2$ in the regression of $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}] / \partial \varepsilon^*$ on a constant and $\epsilon_{t-1}(\boldsymbol{\theta}_s)$.[2]

Our regressand can also be regarded as $\epsilon_t(\boldsymbol{\theta}_s)$ times a damping factor that accounts for skewness and kurtosis, as in the robust estimation literature. Figure 1C illustrates the regressand as a function of $\epsilon_t(\boldsymbol{\theta}_s)$ for four common distributions: normal, Laplace (whose kurtosis is 6), Student t with 6 degrees of freedom (and therefore the same kurtosis as the Laplace), and a two component mixture of normals with skewness coefficient -.5 and kurtosis coefficient 6. As a reference, we also plot the corresponding standardised densities in Figure 1A, and (minus) log-densities in Figure 1B, which can be understood as estimation loss functions.

These damping factors are closely related to those used in the robust estimation literature (see e.g. Maronna et al (2006)). In fact, the Student t and Laplace distributions are common choices for robust influence functions.[3] In this sense, the Student t factor $(\eta+1)/(1-2\eta+\eta\varepsilon^{*2})$ clearly downweights big observations because it is a decreasing function of ε^{*2} for fixed $\eta > 0$, the more so the higher η is. As a result, the ML estimators of π and ω can be regarded as M-estimators, which are typically less sensitive to outliers than the sample mean and variance.

[1] See Arellano (1991), Newey (1985), Newey and McFadden (1994) and Tauchen (1985) for a thorough discussion of moment tests.

[2] It is straightforward to show that (6) coincides with the moment condition one would obtain if the model under the alternative was $y_t = \pi + \sqrt{\omega}(\varepsilon_t^* + \varphi \varepsilon_{t-1}^*)$, thereby encompassing the well known result that MA(1) and AR(1) processes provide locally equivalent alternatives in univariate Gaussian tests for serial correlation (see e.g. Godfrey (1988)).

[3] Other well-known choices not directly related to parametric densities are Tukey's biweight function, which behaves like a quadratic loss function for small values of $\epsilon_t(\boldsymbol{\theta}_s)$ but then tapers off, and the so-called windorising approach, whose loss function is also initially quadratic in $\epsilon_t(\boldsymbol{\theta}_s)$ but eventually becomes linear.

A notable exception is a discrete mixture of normals, since we prove in appendix D.7 that the ML estimators of π and ω coincide with the Gaussian ones.[4]

Despite the theoretical advantages and numerical robustness of our proposed tests, in practice most researchers will test for first order serial correlation in y_t by checking whether its first order sample autocorrelation lies in the 95% confidence interval $(-1.96/\sqrt{T},\ 1.96/\sqrt{T})$. Such a test, though, is nothing other than the test in (1) under the assumption that the conditional distribution of the standardised innovations is *i.i.d.* $N(0,1)$. Apart from tradition, the main justification for using a Gaussian test is the following (see e.g. Breusch and Pagan (1980) or Godfrey (1988)):

Proposition 2 *If in model (1) we assume that the conditional distribution of ε_t^* is i.i.d. $N(0,1)$, when in fact it is i.i.d. $D(0,1,\boldsymbol{\varrho}_0)$ with bounded second moments, then $T \cdot \bar{G}_m^2(1)$ will continue to be distributed as a χ^2 with 1 degree of freedom as T goes to infinity under the null hypothesis of $H_0 : \rho = 0$.*

Nevertheless, it is important to emphasise that the orthogonality condition (6) underlying our proposed mean predictability test also remains valid under the null regardless of whether or not the assumed parametric distribution is correct. More precisely, if we fixed π, ω and $\boldsymbol{\eta}$ to some arbitrary values, $T \cdot R^2$ in the regression of $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial \varepsilon^*$ on a constant and $\epsilon_{t-1}(\boldsymbol{\theta}_s)$ would continue to be asymptotically distributed as a χ_1^2 under the null. In practice, though, researchers will typically replace $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$ by their ML estimators obtained on the basis of the assumed misspecified distribution, $\hat{\boldsymbol{\theta}}_s$ and $\hat{\boldsymbol{\eta}}$, say, and then apply our tests. In principle, one would need to take into account the sampling uncertainty in those PML estimators of $\boldsymbol{\theta}_\infty$ and $\boldsymbol{\eta}_\infty$. However, it is not really necessary to robustify our proposed AR test to distributional misspecification:

Proposition 3 *If in model (1) we assume that the conditional distribution of ε_t^* is i.i.d. with density function $f(.;\boldsymbol{\eta})$, when in fact it is i.i.d. $D(0,1,\boldsymbol{\varrho}_0)$, then $T \cdot R^2$ in the regression of $\partial \ln f[\epsilon_t(\hat{\boldsymbol{\theta}}_s), \hat{\boldsymbol{\eta}}]/\partial \varepsilon^*$ on a constant and $\epsilon_{t-1}(\hat{\boldsymbol{\theta}}_s)$ will continue to be distributed as a χ^2 with 1 degree of freedom as T goes to infinity under the null hypothesis $H_0 : \rho = 0$.*

In this sense, Proposition 2 can be regarded as a corollary to Proposition 3. Unlike in the Gaussian case, however, this proposition holds despite the fact that the (pseudo) maximum likelihood estimators of π and ω will generally be inconsistent under distributional misspecification, with substantial asymptotic biases, as illustrated in Figures 2A-B of Fiorentini and Sentana (2018). Intuitively, the reason is that both the expected value of the Hessian and the variance of the score of the misspecified log-likelihood are block diagonal between ρ and $\boldsymbol{\theta}_s$ under the null.

[4] We also show in Appendix D.5 that the Kotz-based ML estimator of ω coincides with the Gaussian one for a fixed value of π, but the ML estimator of π differs from the sample mean in that it sets to 0 some combination of the artihmetic and harmonic means of the standardised residuals $\epsilon_t(\boldsymbol{\theta}_s)$.

Importantly, a moment test based on (6) will have non-trivial power even when it is no longer a proper LM test. In fact, in section 2.3 we show that our proposed tests are more powerful than the usual regression-based tests in Proposition 2 even though the parametric distribution is misspecified.

The test proposed in Proposition 1, though, requires the specification of a parametric distribution. Given that some researchers might be reluctant to do so, we next consider semiparametric tests that do not make any specific assumptions about the conditional distribution of the standardised innovations ε_t^*, as in González-Rivera and Ullah (2001) and Bera and Ng (2002). There are two possibilities: unrestricted non-parametric density estimates (SP) and non-parametric density estimates that impose symmetry (SSP), which is the univariate version of the procedure used by Hodgson and Vorkink (2003) and Hafner and Rombouts (2007) among others to reduce the curse of dimensionality in multivariate contexts by assuming sphericity. It turns out that not only the asymptotic null distribution of our proposed serial correlation test remains valid if we replace $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial \varepsilon^*$ by one of those non-parametric estimators, but also that the resulting tests are as powerful as if we knew the distribution of ε_t^*, including the true values of the shape parameters:

Proposition 4 *1. The asymptotic distribution of the test in Proposition 3 under the null hypothesis $H_0 : \rho = 0$ is unaffected if we replace $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]/\partial \varepsilon^*$ by a non-parametric estimator, and π_0 and ω_0 by their efficient semiparametric estimators under the null,*

$$\bar{\pi} = \frac{1}{T} \sum\nolimits_{t=1}^{T} y_t \tag{7}$$

and

$$\bar{\omega} = \frac{1}{T} \sum\nolimits_{t=1}^{T} (y_t - \bar{\pi})^2, \tag{8}$$

which coincide with the sample mean and variance of y_t.

2. The resulting test is adaptive, in the sense of having the same non-centrality parameter against sequences of local alternatives of the form $H_l : \rho_T = \bar{\rho}/\sqrt{T}$ as the parametric tests in Proposition 1 with full knowledge of the distribution of ε_t^*.

3. If the true distribution of ε_t^* is symmetric, then the previous two results are valid if we replace $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]/\partial \varepsilon^*$ by a non-parametric estimator that imposes symmetry, and π_0 and ω_0 by their efficient symmetric semiparametric estimators under the null,

$$\dot{\pi} = \bar{\pi} + \frac{1}{\sqrt{\omega}} \left\{ \sum\nolimits_{t=1}^{T} \left[\frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \right]^2 \right\}^{-1} \left[\sum\nolimits_{t=1}^{T} \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \right] \tag{9}$$

and

$$\dot{\omega} = \frac{1}{T} \sum\nolimits_{t=1}^{T} (y_t - \dot{\pi})^2. \tag{10}$$

The adaptivity of the semiparametric tests is a direct consequence of the fact that ρ is partially adaptive, in the sense that after partialling out the effect of estimating $\boldsymbol{\theta}_s$, it can be estimated as efficiently as if we knew the true distribution.

Proposition 4 might suggest that one should never use parametric tests because at best (i.e. under correct specification) they will be as powerful as the semiparametric ones. In finite samples, though, the power of these semiparametric procedures may not be well approximated by the first-order asymptotic theory underlying this proposition. In that sense, a parametric test based on a flexible non-Gaussian distribution might provide a good practical compromise.

Finally, it is worth mentioning that our tests of $H_0 : \rho = 0$ are different from the tests that one would obtain in the so-called Generalised Autoregressive Score models (GAS), also known as Dynamic Conditional Score models (see Creal, Koopman and Lucas (2013) and Harvey (2013)). Those models implicitly change the specification of the conditional mean or variance, which become a function of the lagged value of the log-likelihood score $\partial \ln f(\varepsilon^*; \eta)/\partial \varepsilon^*$. For example, the conditional mean in (1) is a function of $sign(\varepsilon^*_{t-1})$ for the Laplace distribution, while it is affine in $(\eta + 1)\varepsilon^*_{t-1}/(1 - 2\eta + \eta\varepsilon^{*2}_{t-1})$ in the Student t case. In that context, we can use straightforward algebra to show that an LM test of lack of predictability in mean in a Student t-based GAS model would check the significance of the first sample autocorrelation of (4), while it would look at the first-order autocorrelation of $sign(\varepsilon^*_t)$ in the Laplace case, as in the popular Henriksson - Merton (1981) market timing test.[5]

2.2 Exploiting the persistence of expected returns

Let us now consider a situation in which

$$y_t = \pi(1 - \sum\nolimits_{l=1}^{h} \rho_l) + \sum\nolimits_{l=1}^{h} \rho_l y_{t-l} + \sqrt{\omega}\varepsilon^*_t,$$

with $h > 1$ but finite, so that the null hypothesis of lack of predictability becomes $H_0 : \rho_1 = \ldots = \rho_h = 0$. In view of our previous discussion, it is not difficult to see that under this maintained assumption the score test of $\rho_l = 0$ will be based on the orthogonality condition

$$E\left\{ \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \epsilon_{t-l}(\boldsymbol{\theta}_s) | \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0 \right\} = 0.$$

In this context, it is straightforward to show that the joint test for AR(h) dynamics will be given by the sum of h terms of the form

$$T \cdot \frac{\bar{G}^2_m(l)}{\mathcal{I}_{\rho\rho}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0)}$$

for $l = 1, \ldots, h$, whose asymptotic distribution would be a χ^2_h under the null.

Such a test, though, does not impose any prior knowledge on the nature of the expected return process, other than its lag length is h. Nevertheless, there are theoretical and empirical reasons which suggest that time-varying expected returns should be smooth processes.

[5] Tests that transform both regressand and regressor to make them robust to outliers have also been discussed by Amengual and Sentana (2018) in a copula context with arbitrary margins, and Camponovo, Scaillet and Trojani (2013) in a more general non-likelihood setting.

A rather interesting example of persistent expected returns is an autoregressive model in which $\rho_l = \rho$ for all l. In this case, we can use the results in Fiorentini and Sentana (1998) to show that the process for expected returns will be given by the following not strictly invertible $\text{ARMA}(h, h-1)$ process:

$$\mu_{t+1} = \pi(1-h\rho) + \sum\nolimits_{j=1}^{h} \rho\mu_{t+1-j} + \rho\left[\varepsilon_t + \sum\nolimits_{j=1}^{h-1} \varepsilon_{t-j}\right]. \tag{11}$$

As long as the covariance stationarity condition $h\rho < 1$ is satisfied, the autocorrelations of the expected return process can be easily obtained from its autocovariance generating function

$$\psi_{\mu\mu}(z) = \frac{\left(1 + \sum_{j=1}^{h-1} z^j\right)\left(1 + \sum_{j=1}^{h-1} z^{-j}\right)}{\left(1 - \rho\sum_{j=1}^{h} z^j\right)\left(1 - \rho\sum_{j=1}^{h} z^{-j}\right)}, \tag{12}$$

which contrasts with the autocovariance generating function of the observed process

$$\psi_{yy}(z) = \frac{1}{\left(1 - \rho\sum_{j=1}^{h} z^j\right)\left(1 - \rho\sum_{j=1}^{h} z^{-j}\right)}.$$

In this context, we can easily find examples in which the autocorrelations of the observed return process are very small while the autocorrelations of the expected return process are much higher, and decline slowly towards 0. For example, Figure 2 presents the correlograms of y_t and μ_{t+1} on the same vertical scale for $h = 24$ and $\rho = .015$.[6]

This differential behaviour suggests that a test against first order correlation will have little power to detect such departures from white noise, the optimal test being one against an $\text{AR}(h)$ process with common coefficients. We shall formally analyse this issue in the next section.

From the econometric point of view, the assumption that $\rho_l = \rho$ for all l does not pose any additional problems. Specifically, it is easy to prove that the relevant orthogonality condition will become

$$E\left\{\frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \sum\nolimits_{l=1}^{h} \epsilon_{t-l}(\boldsymbol{\theta}_s) \middle| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0\right\} = 0, \tag{13}$$

with $h\mathcal{I}_{\rho\rho}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta})$ being the corresponding asymptotic variance.

This moment condition is analogous to the one proposed by Jegadeesh (1989) to test for long run predictability of individual asset returns without introducing overlapping regressands. Cochrane (1991) and Hodrick (1992) discussed related suggestions. The intuition is that if returns contain a persistent but mean reverting predictable component, using a persistent right hand side variable such as an overlapping h-period return may help to pick it up. Not surprisingly, the asymptotic variance is analogous to the so-called Hodrick (1992) standard errors used in tests for long run predictability in univariate OLS regressions with overlapping regressands.

[6] Expression (12) implies the correlograms of μ_{t+1} and an overlapping sum of h consecutive returns coincide.

More recently, the Gaussian version of (13) has also been tested by Moskowitz, Ooi and Pedersen (2012) in their empirical analysis of time series momentum. These authors provide both behavioural and rational justifications for the forecasting ability of lagged compound returns.

It is important to mention that the regressor $\sum_{l=1}^{h} \epsilon_{t-l}(\boldsymbol{\theta}_s)$ will be quite persistent even if returns are serially uncorrelated because of the data overlap. Specifically, the first-order autocorrelation coefficient will be $1 - 1/h$ in the absence of return predictability. Nevertheless, since the correlation between the innovation to the regressor at time $t+1$ and the innovations $\epsilon_t(\boldsymbol{\theta}_s)$ is $1/\sqrt{h}$ under the null, the size problems that plague predictive regressions should not affect much our test (see Campbell and Yogo (2006)).

2.3 The relative power of mean predictability tests

Let us begin by assessing the power gains obtained by exploiting the persistence of expected returns. For simplicity we consider Gaussian tests only, and evaluate asymptotic power against *compatible* sequences of local alternatives of the form $\rho_{0T} = \bar{\rho}/\sqrt{T}$. As we show in appendix B, when the true model is (11), the non-centrality parameter of the Gaussian score test for first order serial correlation is $\bar{\rho}^2$ regardless of h, while the non-centrality parameter of the test that exploits the persistence of the conditional mean will be $h\bar{\rho}^2$. Hence, Pitman's asymptotic relative efficiency of the two tests is precisely h. Figure 3A shows that those differences in non-centrality parameters result in substantive power gains. However, the asymptotic relative efficiency would be exactly reversed in the unlikely event that the true model were an AR(1) but we tested for it by using the moment condition (13) (see appendix B). Not surprisingly, this would result in substantial power losses, which are also illustrated in Figure 3A.

Let us now turn to study the improvements obtained by considering distributions other than the normal. The following result gives us the necessary ingredients.

Lemma 1 *If the true DGP corresponds to (1) with $\rho_0 = 0$, then the feasible ML estimator of ρ is as efficient as the infeasible ML estimator, which require knowledge of $\boldsymbol{\eta}_0$. In contrast, the inefficiency ratio of the Gaussian PML estimator of ρ is $\mathcal{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$, with $\mathcal{M}_{ll}(\boldsymbol{\eta}_0)$ defined in (3).*

This means that Pitman's asymptotic relative efficiency of those serial correlation tests that exploit the non-normality of y_t will be $\mathcal{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$. Figure 3B assesses the power gains against local AR(1) alternatives under the assumption that the true conditional distribution of ε_t^* is a Student t with either 6 or 4.5 degrees of freedom. This figure confirms that the power gains that accrue to our proposed serial correlation tests by exploiting the leptokurtosis of the t distribution are noticeable, the more so the higher the kurtosis of y_t. Similarly, Figure 3C repeats the same exercise for two normal mixtures whose kurtosis coefficients are both 6, and whose skewness coefficients are -.5 and -1.219, respectively. Once again, we can see that there are significant

power gains. In this sense, it is worth remembering that since our semiparametric tests are adaptive, they should achieve these gains, at least asymptotically.

For the parametric tests, however, the results in those figures are based on the assumption that the non-Gaussian distribution is correctly specified. Given that we have proved in Proposition 3 that those tests are robust, an obvious question is what their relative power is under distributional misspecification. For the sake of concreteness, we answer this question for the Student t tests when the degrees of freedom parameter is estimated and the true distribution is either a fourth-order Gram-Charlier (GC) expansion of the normal or an unrestricted location-scale mixture of two normals. In this regard, our results complement those in Amengual and Sentana (2010), who show that mean tests based on the Student t distribution dominate the corresponding Gaussian tests when the true distribution is either Kotz or a symmetric scale mixture of two normals. Since the non-centrality parameters do not depend on the true values of π and ω, we assume they are equal to 0 and 1 without loss of generality.

Figure 4A illustrates the ratio of the non-centrality parameters of the Gaussian and Student t tests of $H_0 : \rho = 0$ when the true distribution is an admissible fourth-order GC expansion of the standard normal as a function of the skewness and kurtosis coefficients compatible with a non-negative density everywhere (see Jondeau and Rockinger (2001) and appendix C.2.2 for a characterisation of the set of skewness and kurtosis values that give rise to a non-negative density for the fourth-order expansion).[7] The results clearly show that a misspecified Student t systematically leads to more powerful tests of $H_0 : \rho = 0$ than the Gaussian ones.

Similarly, Figure 4B repeats the same calculations, but this time assuming that the true distribution is a mixture of two normals in which the probability of the first component is 5%. To facilitate its comparison with Figure 4A, the axes are again the skewness and kurtosis coefficients of the mixture as we vary the shape parameters δ and υ (see appendix C.1.2 for a characterisation of the set of skewness and kurtosis values that are compatible with a mixture of two normals when the mixing probability is fixed).[8] Once again, a misspecified Student t systematically leads to more powerful tests.

These results raise the question of whether the observed advantage of the Student t is universal. Obviously, it crucially depends on the (reciprocal) degrees of freedom parameter η being estimated, for otherwise it is easy to see that fixing η to a positive value when the true value is 0 necessarily means that the Gaussian test will dominate. To answer this question, we have considered all possible discrete mixtures of two normals because this family of distribution covers

[7] Since the non-centrality parameters do not depend on the sign of the skewness coefficient, we only show the positive side of the admissible region.

[8] As in Figure 4A, we only show the positive skewness side of the admissible region because the non-centrality parameters do not depend on the sign of the skewness coefficient.

the entire admissible range of skewness-kurtosis coefficients that any distribution can achieve, which is characterised by the bound $E(\varepsilon_t^{*4}) \geq 1 + E^2(\varepsilon_t^{*3})$ (see Stuart and Ord (1977) and appendix C.1.2). In effect, this requires repeating the calculations underlying Figure 4B for every value of λ between 0 and 1. Although we have not been able to find any counterexample, at least in a triple grid for λ, υ and δ over (0,1)×(0,1)×(-4,4), we cannot rule out that it might exist for some other family of true distributions.

3 Tests for predictability in variance

3.1 First-order ARCH tests

Although we can consider any variance predictor variable available at time $t-1$, for pedagogical reasons we initially develop tests of first order ARCH effects under the maintained assumption that the conditional mean is constant. More specifically, the model under the alternative is

$$\left.\begin{array}{c} y_t = \pi_0 + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha(y_{t-1} - \pi)^2, \\ \varepsilon_t^* | I_{t-1}; \pi, \omega, \alpha, \boldsymbol{\eta} \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}), \\ \text{with density function } f(., \boldsymbol{\eta}) \end{array}\right\}, \tag{14}$$

where the parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, with $\boldsymbol{\theta}' = (\boldsymbol{\theta}_s', \alpha)'$. In this context, the null hypothesis is $H_0 : \alpha = 0$. Regardless of the specific parametric distribution, testing the null of conditional homoskedasticity against first order ARCH is extremely easy:

Proposition 5 *Let*

$$\bar{G}_s(j) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \left\{ 1 + \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \epsilon_t(\boldsymbol{\theta}_{s0}) \right\} \epsilon_{t-j}^2(\boldsymbol{\theta}_{s0})$$

denote the sample cross moment of $\epsilon_{t-j}^2(\boldsymbol{\theta}_{s0})$ and 1 plus the derivative of the conditional log density of ε_t^* with respect to its argument evaluated at $\epsilon_t(\boldsymbol{\theta}_{s0})$ times $\epsilon_t(\boldsymbol{\theta}_{s0})$.

1. *Under the null hypothesis $H_0 : \alpha = 0$, the score test statistic*

$$LM_{ARCH(1)} = \frac{T}{4} \cdot \frac{\bar{G}_s^2(1)}{\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0)} \tag{15}$$

will be distributed as a χ^2 with 1 degree of freedom as T goes to infinity, where

$$\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}) = V[\tfrac{1}{2}\epsilon_t^2(\boldsymbol{\theta}_s) | \boldsymbol{\theta}_s, 0, \boldsymbol{\eta}] \cdot M_{ss}(\boldsymbol{\eta})$$

and

$$M_{ss}(\boldsymbol{\eta}) = V\left[\frac{\partial \ln f(\varepsilon_t^*, \boldsymbol{\eta})}{\partial \varepsilon^*} \varepsilon_t^* \middle| I_{t-1}; \boldsymbol{\eta} \right]. \tag{16}$$

2. *This asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_{s0}$ or $\boldsymbol{\eta}_0$ by their maximum likelihood estimators.*

As in the case of the mean predictability tests discussed in the previous section, the exact expression for $\bar{G}_s(1)$ depends on the assumed distribution. As for $M_{ss}(\boldsymbol{\eta})$, we can either use its theoretical expression (for instance $2(1+3\eta)^{-1}$ in the case of the Student t, which reduces to 2 under normality), compute the sample analogue of (16), or exploit the information matrix equality and calculate it as twice the sample average of $1 - \partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right] / \partial \varepsilon^* \partial \varepsilon^* \cdot \epsilon_t^2(\boldsymbol{\theta}_s)$. Once again, this choice will affect the finite sample properties of the tests (see Davidson and MacKinnon (1983)), as well as their validity under distributional misspecification.

Intuitively, we can interpret the above score test a moment test based on the following orthogonality condition:

$$E\left[\left\{1 + \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*}\epsilon_t(\boldsymbol{\theta}_s)\right\}\epsilon_{t-1}^2(\boldsymbol{\theta}_s)\middle|\, \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0\right] = 0, \tag{17}$$

which is also related to the conditions used by Bontemps and Meddahi (2012). In fact, given that the score with respect to ω under the null is proportional to

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\left\{1 + \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]}{\partial \varepsilon^*}\epsilon_t(\boldsymbol{\theta}_s)\right\},$$

the sample second moment will numerically coincide with the sample covariance if we evaluate the standardised residuals at the ML estimators. As a result, an asymptotically equivalent test under the null and sequences of local alternatives would be obtained as $T \cdot R^2$ in the regression of $1 + \epsilon_t(\boldsymbol{\theta}_s)\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial \varepsilon^*$ on a constant and $\epsilon_{t-1}^2(\boldsymbol{\theta}_s)$.

Importantly, the numerical invariance of LM tests to non-linear transformations of the restrictions when the asymptotic variance under the null is computed using either the information matrix or the sample variance of the score (see section 17.4 of Ruud (2000)) implies that the test for $H_0 : \gamma = 0$, where $\gamma = \alpha/\omega$, will be numerically identical in finite samples to the test above, and the same is obviously true to its size and power properties.

Godfrey (1988) re-interprets Glejser (1969) heteroskedasticity test, which regresses the absolute value of the residuals on several predictor variables, as an ML test based on the Laplace distribution. More generally, our regressand can be regarded as $\epsilon_t^2(\boldsymbol{\theta}_s)$ times a damping factor that accounts for skewness and kurtosis.[9] Figure 5 illustrates the transformation of the regressands for the same four standardised distributions depicted in Figure 1: normal, Laplace distribution, Student t with 6 degrees of freedom (and therefore the same kurtosis as the Laplace), and a discrete mixture of normals with skewness coefficient -.5 and kurtosis coefficient 6. A comparison with Figure 1C indicates that the regressands of the mean and variance predictability

[9] This factor also plays an important role in the beta-t-ARCH models proposed by Harvey and Chakravarty (2008), although if one derived an LM test for conditional homoskedasticity against their models, the damping factor $(\eta+1)/(1-2\eta+\eta\varepsilon^{*2})$ would appear not only in the regressand but also in the regressor, as we discussed in the case of the serial correlation tests at the end of section 2.1.

tests can behave rather differently for a given distribution.[10]

Despite the theoretical advantages and numerical robustness of our proposed tests, in practice, most researchers will test for first order ARCH effects in y_t by checking whether the first order sample autocorrelation of $\epsilon_t^2(\boldsymbol{\theta}_s)$ lies in the 95% confidence interval $(-1.96/\sqrt{T}, 1.96/\sqrt{T})$. Such a test, though, is nothing other than the test in (5) under the assumption that the conditional distribution of the standardised innovations is *i.i.d.* $N(0,1)$. Apart from tradition, the main justification for using a Gaussian test is the following (see e.g. Demos and Sentana (1998)):

Proposition 6 *If in model (14) we assume that the conditional distribution of ε_t^* is i.i.d. $N(0,1)$, when in fact it is i.i.d. $D(0,1,\boldsymbol{\varrho}_0)$ with bounded fourth moments, then (15) will still be distributed as a χ^2 with 1 degrees of freedom as T goes to infinity under the null hypothesis of $H_0 : \alpha = 0$ as long as we replace the Gaussian expression for $M_{ss}(\boldsymbol{\eta})$ with $V[\epsilon_t^2(\boldsymbol{\theta}_s)\big|\,\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0]$.*

Notice that in this case we have to use Koenker's (1981) version of the usual heteroskedasticity test because the information matrix version of Engle's (1982) test, which assumes that $V[\epsilon_t^2(\boldsymbol{\theta}_s)\big|\,\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0] = 2$, will be incorrectly sized.

But again, it is important to emphasise that the orthogonality condition (17) underlying our proposed ARCH test also remains valid under the null regardless of whether or not the assumed parametric distribution is correct. Specifically, if we fixed π, ω and $\boldsymbol{\eta}$ to some arbitrary values, $T \cdot R^2$ in the regression of $1 + \epsilon_t(\boldsymbol{\theta}_s)\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial\varepsilon^*$ on a constant and $\epsilon_{t-1}^2(\boldsymbol{\theta}_s)$ would continue to be asymptotically distributed as a χ_1^2 under the null. In practice, though, researchers will typically replace $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$ by their ML estimators obtained on the basis of the assumed distribution, $\hat{\boldsymbol{\theta}}_s$ and $\hat{\boldsymbol{\eta}}$, say, and then apply our tests. In principle, one would have to take into account the sampling uncertainty in those PML estimators of $\boldsymbol{\theta}_\infty$ and $\boldsymbol{\eta}_\infty$. However, it is not really necessary to robustify our proposed ARCH test to distributional misspecification:

Proposition 7 *If in model (14) we assume that the conditional distribution of ε_t^* is i.i.d. with density function $f(.;\boldsymbol{\eta})$, when in fact it is i.i.d. $D(0,1,\boldsymbol{\varrho}_0)$, then $T \cdot R^2$ in the regression of $1 + \epsilon_t(\hat{\boldsymbol{\theta}}_s)\partial \ln f[\epsilon_t(\hat{\boldsymbol{\theta}}_s), \hat{\boldsymbol{\eta}}]/\partial\varepsilon^*$ on a constant and $\epsilon_{t-1}^2(\hat{\boldsymbol{\theta}}_s)$ will continue to be distributed as a χ^2 with 1 degree of freedom as T goes to infinity under the null hypothesis $H_0 : \gamma = 0$.*

In this sense, the result in Proposition 6 can be regarded as a corollary to Proposition 7 in the Gaussian case. Similarly, the suggestion made in Proposition 2 of Machado and Santos Silva (2000) to robustify Glejser's heteroskedasticity test, which in our case would involve replacing π by the sample median of y_t, can also be regarded as a corollary to this Proposition

[10] Another interesting example is given by the Kotz distribution, whose variance regresand is proportional to $\epsilon_t^2(\boldsymbol{\theta}_s)$, with a factor of proportionality that depends on $\varkappa$, while its mean regressand is a linear combination of $\epsilon_t(\boldsymbol{\theta}_s)$ and $\epsilon_t^{-1}(\boldsymbol{\theta}_s)$. Although the ML estimator of π is not the sample mean, if we knew that $\pi_0 = 0$ *and* we correctly imposed this restriction in estimation, $\hat{\omega}^2$ would coincide with the second sample moment, which is also the Gaussian PMLE. As a result, the Gaussian and Kotz ARCH(1) tests would be numerically identical too. Proposition 9.1 in Francq and Zakoïan (2010) confirms that the numerical equality will apply to the corresponding ML estimators of α under the alternative.

in the Laplace case. Once again, though, Proposition 6 holds despite the fact that the (pseudo) maximum likelihood estimators of π and ω will generally be inconsistent under distributional misspecification, with substantial asymptotic biases, as illustrated in Figures 2A-B of Fiorentini and Sentana (2018). Once more, the intuitive reason is that both the expected value of the Hessian and the variance of the score of the misspecified log-likelihood are block diagonal between γ and $\boldsymbol{\theta}_s$ under the null.

Importantly, a moment test based on (17) will continue to have non-trivial power even though it will no longer be an LM test. In fact, in section 3.3 we show that our proposed tests are generally more powerful than the usual regression-based tests in Proposition 6 even though the parametric distribution is misspecified.

The test proposed in Proposition 5, however, requires to specify a parametric distribution. Since some researchers might be reluctant to do so, we next consider semiparametric tests that do not make any specific assumptions about the conditional distribution of the innovations ε_t^*, as in Linton and Steigerwald (2000) and Bera and Ng (2002). Once again, there are two possibilities: unrestricted non-parametric density estimates (SP) and non-parametric density estimates that impose symmetry (SSP). Unsurprisingly, it turns out that not only the asymptotic null distribution of our proposed serial correlation test remains valid if we replace $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial \varepsilon^*$ by one of those non-parametric estimators, but also that the resulting tests are as powerful as if we knew the distribution of ε_t^*, including the true values of the shape parameters:

Proposition 8 *1. The asymptotic distribution of the test in Proposition 7 is unaffected if we replace $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]/\partial \varepsilon^*$ by a non-parametric estimator and π_0 and ω_0 by their efficient semiparametric estimators under the null defined in (7) and (8), which coincide with the sample mean and variance of y_t.*

2. The resulting test is adaptive, in the sense of having the same non-centrality parameter against sequences of local alternatives of the form $H_l : \gamma_T = \bar{\gamma}/\sqrt{T}$ with $\gamma = \alpha/\omega$, as the parametric tests in Proposition 5 with full knowledge of the distribution of ε_t^*.

3. If the true conditional distribution is symmetric, then the previous two results are valid if we replace $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]/\partial \varepsilon^*$ by a non-parametric estimator that imposes symmetry, and π_0 and ω_0 by their efficient symmetric semiparametric estimators under the null, which are defined in (9) and (10).

The adaptivity of the semiparametric tests is a direct consequence of the fact that γ is partially adaptive, in the sense that after partialling out the effect of estimating $\boldsymbol{\theta}_s$, it can be estimated as efficiently as if we knew the true distribution.

Once again, Proposition 8 might suggest that one should never use parametric tests because at best (i.e. under correct specification) their local power coincides with those of the semiparametric ones. As before, though, the power of these semiparametric procedures in finite samples

may not be well approximated by the first-order asymptotic theory that justifies their adaptivity, so a parametric test based on a flexible but parsimoniously parametrised non-Gaussian distribution might provide a good practical compromise.

3.2 Exploiting the persistence of volatilities

Let us now consider a situation in which

$$\sigma_t^2(\boldsymbol{\theta}) = \omega(1 - \sum_{j=1}^{q} \alpha_j) + \sum_{j=1}^{q} \alpha_j (y_{t-j} - \pi)^2,$$

with $q > 1$ but finite, so that the null hypothesis of conditional homoskedasticity becomes $H_0 : \alpha_1 = \ldots = \alpha_q = 0$. In view of our previous discussion, it is not difficult to see that under this maintained assumption the score test of $\alpha_j = 0$ will be based on the orthogonality condition

$$E\left[\left\{1 + \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \epsilon_t(\boldsymbol{\theta}_s)\right\} \epsilon_{t-1}^2(\boldsymbol{\theta}_s) | \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0\right] = 0.$$

In this context, it is straightforward to show that the joint test for ARCH(q) dynamics will be given by the sum of q terms of the form

$$\frac{T}{4} \cdot \frac{\bar{G}_s^2(j)}{\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0)}$$

for $l = 1, \ldots, q$, whose asymptotic distribution would be a χ_q^2 under the null.

But since the inequality constraints $\alpha_1 \geq 0, \ldots, \alpha_q \geq 0$ must be satisfied to guarantee nonnegative conditional variances of an ARCH(q) model, an even more powerful test can be obtained if we test $H_0 : \alpha_1 = 0, \ldots, \alpha_q = 0$ versus $H_1 : \alpha_1 \geq 0, \ldots, \alpha_q \geq 0$, with at least one strict inequality. An argument analogous to the one in Demos and Sentana (1998) shows that a version of the Kuhn-Tucker multiplier test of Gouriéroux, Holly and Monfort (1980) can be simply computed as the sum of the square t-ratios associated with the positive estimated coefficients in the regression of $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial \varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s)$ on a constant and the first q lags of $\epsilon_t^2(\boldsymbol{\theta}_s)$. The asymptotic distribution of such a test will be given by $\sum_{i=0}^{q} \binom{q}{i} 2^{-q} \chi_i^2$, which is a mixture of $q+1$ independent $\chi^{2\prime}$s whose critical values can be found in Table 1 in that paper.

Nevertheless, there is a lot of evidence which suggests that volatilities are rather persistent processes. In this sense, the obvious model that we shall use to capture such an effect is a GARCH(1,1) process in which q is in fact unbounded, and $\alpha_j = \alpha\beta^{j-1}$ for $j = 1, 2, \ldots$

From the econometric point of view, this model introduces some additional complications because the parameter β becomes underidentified when $\alpha = 0$ (see Bollerslev (1986)). Note, however, that since α has to be positive under the alternative to guarantee that $\sigma_t^2(\boldsymbol{\theta}) = \omega(1-\beta)^{-1} + \alpha \sum_{j=0}^{t-2} \beta^j \varepsilon_{t-j-1}^2(\boldsymbol{\theta})$ is nonnegative everywhere, we should still test $H_0 : \alpha = 0$ vs. $H_1 : \alpha \geq 0$ even if we knew β. One solution to testing situations such as this involves

computing the test statistic for many values of β in the range [0,1), which are then combined to construct an overall statistic, as initially suggested by Davies (1977, 1987). Andrews (2001) discusses ways of obtaining critical values for such tests by regarding the different LM statistics as continuous stochastic processes indexed with respect to the parameter β. An alternative solution involves choosing an arbitrary value of β, $\bar{\beta}$ say, to carry out a one-sided LM test as $T \cdot R^2$ from the regression of $\{1 + \partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]/\partial \varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s)\}$ on a constant and the distributed lag $\sum_{j=0}^{t-2} \bar{\beta}^j \epsilon^2_{t-j-1}(\boldsymbol{\theta}_s)$ (see Demos and Sentana (1998)). The one-sided versions of such tests are asymptotically distributed as a $50:50$ mixture of χ^2_0 and χ^2_1 irrespective of the value of $\bar{\beta}$. Obviously, the chosen value of $\bar{\beta}$ influences the small sample power of the test, an issue to which we shall return in the next section, but the advantage is that the resulting test has a standard distribution under H_0. An attractive possibility is to choose $\bar{\beta}$ equal to the decay factor recommended by RiskMetrics (1996) to obtain their widely used exponentially weighted average volatility estimates (e.g. $\bar{\beta} = .94$ for daily observations). In this respect, note that since the RiskMetrics volatility measure is proportional to $\sum_{j=0}^{t-2} \bar{\beta}^j \epsilon^2_{t-j-1}(\boldsymbol{\theta}_s)$, in effect our proposed GARCH(1,1) tests differ from the ARCH(q) tests discussed before in that the q lags of the squared residuals are replaced by the RiskMetrics volatility estimate in the auxiliary regressions. Straightforward algebra shows that the asymptotic variance of this statistic would be $(1-\bar{\beta}^2)^{-1}$ times the ARCH(1) expression under the null of conditional homoskedasticity.

3.3 The relative power of variance predictability tests

Let us begin by assessing the power gains obtained by exploiting the persistence of conditional variances. For simplicity, we first compare the Gaussian versions of the ARCH(1) and fixed-$\bar{\beta}$ GARCH(1,1) tests, and evaluate asymptotic power against *compatible* sequences of local alternatives of the form $\alpha_{0T} = \bar{\alpha}/\sqrt{T}$. Given that the sample variance is consistent for ω, exactly the same results will be obtained if we worked with the transformed sequence $\gamma_{0T} = (\bar{\alpha}\omega_0^{-1})/\sqrt{T} = \bar{\gamma}/\sqrt{T}$.

As we show in appendix B, when the true model is (11), the non-centrality parameter of the Gaussian pseudo-score test based on the first order serial correlation coefficient of $\epsilon^2_t(\boldsymbol{\theta}_s)$ is $\bar{\alpha}^2$ regardless of the true value of β. In contrast, the non-centrality parameter of the fixed-$\bar{\beta}$ GARCH(1,1) test is $\bar{\alpha}^2(1-\bar{\beta}^2)/(1-\bar{\beta}\beta_0)^2$. Hence, the asymptotic relative efficiency of the two tests is $(1-\bar{\beta}^2)/(1-\bar{\beta}\beta_0)^2$, which is not surprisingly maximised when $\bar{\beta} = \beta_0$. Figure 6A shows that for a realistic value of β_0 these efficiency gains yield substantive power gains when we set $\bar{\beta}$ to its RiskMetrics value of .94

Let us now study the power gains obtained by considering distributions other than the normal. The following proposition gives us the necessary ingredients:

Lemma 2 *If the true DGP corresponds to (14) with $\alpha_0 = 0$, then the feasible ML estimator of α is as efficient as the infeasible ML estimator, which require knowledge of $\boldsymbol{\eta}_0$. In contrast, the inefficiency ratio of the Gaussian PML estimator of α is $4/[(\kappa_0 - 1)\mathcal{M}_{ss}(\boldsymbol{\eta}_0)]$, where $\mathcal{M}_{ss}(\boldsymbol{\eta}_0)$ is defined in (16).*

Lemma 2 then implies that the local non-centrality parameter of the Gaussian test for ARCH is α^2, while the non-centrality parameter of the parametric test for ARCH is $\frac{1}{4}[(\kappa_0-1)\mathcal{M}_{ss}(\boldsymbol{\eta}_0)]\alpha^2$. Figure 6B assesses the power gains under the assumption that the true conditional distribution of ε_t^* is a Student t with either 6 or 4.5 degrees of freedom. This figure confirms that the power gains that accrue to our proposed ARCH tests by exploiting the leptokurtosis of the t distribution are in fact more pronounced than the corresponding gains in the mean predictability tests. Similarly, Figure 6C repeats the same exercise for two discrete location scale mixture of normals whose kurtosis coefficients are both 6, and whose skewness coefficients are either -.5 or -1.219. In this case, our tests also yield significant power gains. In this sense, it is worth remembering that since our semiparametric tests are adaptive, they should achieve these gains, at least asymptotically.

For the parametric tests, however, the results in those figures are based on the assumption that the non-Gaussian distribution is correctly specified. Given that we have proved in Proposition 7 that those tests are robust, an obvious question is what their relative power is under distributional misspecification. As in section 2.3, we answer this question for the Student t-based tests when the degrees of freedom parameter is estimated and the true distribution is either a fourth-order GC expansion of the normal or a mixture of two normals. Once again, we set π and ω to 0 and 1 without loss of generality.

Figure 7A depicts the ratio of the non-centrality parameters of the Gaussian and Student t tests of $H_0 : \gamma = 0$ when the true distribution is an admissible fourth-order GC expansion of the standard normal as a function of the skewness and kurtosis coefficients. As can be seen, the results clearly show that the misspecified Student t systematically leads to more powerful tests than the Gaussian ones. Figure 7B repeats the same calculations, but this time assuming a mixture of two normals with mixing probability $\lambda = .05$. Once more, the misspecified Student t systematically leads to more powerful tests.

These results raise the question of whether the advantage of the Student t are pervasive, at least when the (reciprocal) degrees of freedom parameter η is estimated. Unlike in the case of the serial correlation tests discussed in section 2.3, we have been able to find a somewhat contrived counterexample in which the Student t test is marginally less powerful than the Gaussian one. Specifically, if the true distribution is a scale mixture of two normals with $\delta = 0$, $\varkappa = .03$ and $\lambda = .97$, so that the relative variance of the less likely component is very small - the so-called *inlier case* in Amengual and Sentana (2011) - then the Gaussian test is marginally more powerful

than the misspecified Student t test, with an asymptotic inefficiency ratio of merely 1.0035.

4 Joint tests for mean-variance predictability

In this section we shall consider joint tests of AR and ARCH effects. Specifically, our alternative in the first-order case will be

$$\left.\begin{array}{c} y_t = \mu_t(\pi_0, \rho_0) + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \mu_t(\pi, \rho) = \pi(1-\rho) + \rho y_{t-j}, \\ \sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha_j[y_{t-1} - \mu_{t-1}(\pi, \rho)]^2, \\ \varepsilon_t^* | I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}_0) \end{array}\right\}, \tag{18}$$

where the parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, with $\boldsymbol{\theta}' = (\boldsymbol{\theta}_s', \rho, \alpha)'$. When the conditional variance $\sigma_t^2(\boldsymbol{\theta})$ is constant ($\alpha = 0$), the above formulation reduces to (1). Similarly, when the levels of the observed variable are unpredictable ($\rho = 0$), the above model simplifies to (14). Finally, the joint null hypothesis of lack of predictability in mean and variance corresponds to $\rho = 0$ and $\alpha = 0$.

In this context, the double length artificial regression of Davidson and MacKinnon (1988) might seem natural. However, there are two potential problems. First, in general the mean and variance regressands, namely $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial\varepsilon^*$ and $1 + \epsilon_t(\boldsymbol{\theta}_s)\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial\varepsilon^*$, have different variances, which introduces heteroskedasticity. More seriously, those two regressands will be correlated unless the true distribution is symmetric. The solution is a system of seemingly unrelated regression equations (SURE) in which one simultaneously regresses each of those regressands on the corresponding regressors, $\epsilon_{t-1}(\boldsymbol{\theta}_s)$ and $\epsilon_{t-1}^2(\boldsymbol{\theta}_s)$, respectively, and jointly tests the significance of both slope coefficients. In effect, this is a joint moment test of (6) and (17). Under the null, the covariance matrix of those moment conditions is

$$V \left[\begin{array}{c} \epsilon_{t-1}(\boldsymbol{\theta}_{s0}) \\ \frac{1}{2}\epsilon_{t-1}^2(\boldsymbol{\theta}_{s0}) \end{array} \right] \odot \left[\begin{array}{cc} M_{ll}(\boldsymbol{\eta}_0) & M_{ls}(\boldsymbol{\eta}_0) \\ M_{ls}(\boldsymbol{\eta}_0) & M_{ss}(\boldsymbol{\eta}_0) \end{array} \right],$$

where $\odot$ denotes the Hadamard (or element-by-element) product of two matrices and

$$M_{ls}(\boldsymbol{\eta}_0) = cov\left[\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}), \boldsymbol{\eta}_0]/\partial\varepsilon^*, 1 + \epsilon_t(\boldsymbol{\theta}_{s0})\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}), \boldsymbol{\eta}_0]/\partial\varepsilon^*\right],$$

which reduces to

$$\left[\begin{array}{cc} 1 & \frac{1}{2}\phi_0^2 \\ \frac{1}{2}\phi_0^2 & \frac{1}{4}(\kappa_0 - 1)^2 \end{array} \right]$$

when the assumed distribution is Gaussian but the true one has skewness and kurtosis coefficients ϕ_0 and κ_0, respectively.

Nevertheless, if the true distribution of ε_t^* is symmetric, then it turns out that the joint tests of AR(1)-ARCH(1) in Propositions 1 and 5 is simply the sum of the separate tests:

Proposition 9 *If ε_t^* is symmetrically distributed, then*

1. *Under the joint null hypothesis* $H_0 : \rho = 0$ *and* $\alpha = 0$ *the score test statistic*

$$LM_{AR(1)-ARCH(1)}(\boldsymbol{\eta}_0) = LM_{AR(1)}(\boldsymbol{\eta}_0) + LM_{ARCH(1)}(\boldsymbol{\eta}_0),$$

will be distributed as a χ^2 *with* 2 *degrees of freedom as* T *goes to infinity. This asymptotic null distribution is unaffected if we replace* $\boldsymbol{\theta}_s$ *and* $\boldsymbol{\eta}_0$ *by their joint maximum likelihood estimators.*

2. *It also remains valid if we replace* $\boldsymbol{\theta}_{s0}$ *by its symmetric semiparametric estimator*

3. *Under the same null hypothesis*

$$LM_{AR(1)-ARCH(1)}(\mathbf{0}) = LM_{AR(1)}(\mathbf{0}) + LM_{ARCH(1)}(\mathbf{0})$$

will also be distributed as a χ^2 *with* 2 *degrees of freedom as* T *goes to infinity irrespective of whether the conditional distribution is normal. This result continues to hold if we replace* $\boldsymbol{\theta}_{s0}$ *by its Gaussian pseudo maximum likelihood estimator* $\bar{\boldsymbol{\theta}}_s$

Intuitively, the serial correlation orthogonality condition (6) is asymptotically orthogonal to the ARCH orthogonality condition (17) because all odd order moments of symmetric distributions are 0, which means that the joint test is simply the sum of its two components.

Obviously, all previous results continue to hold *mutatis mutandi* for the alternative persistent regressors that we have discussed in sections 2.2 and 3.2.

5 Monte Carlo analysis

5.1 Design and computational details

In this section, we assess the finite sample performance of the different testing procedures discussed above by means of an extensive Monte Carlo exercise adapted to the behaviour of the market portfolios in the empirical application in section 6. Specifically, we consider the following univariate, covariance stationary ARMA(1,1)-GARCH(1,1) model:

$$\begin{gathered} y_t = \mu_t(\pi_0, \rho_0, \varphi_0) + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \mu_t(\pi, \rho, \varphi) = \pi + \rho y_{t-1} + \varphi[y_{t-1} - \mu_{t-1}(\pi, \rho, \varphi)], \\ \sigma_t^2(\boldsymbol{\theta}) = \omega + \alpha[y_{t-1} - \mu_{t-1}(\pi, \rho, \varphi)]^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}), \\ \varepsilon_t^* | I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}_0). \end{gathered}$$

We set $\pi = .5$ and $\omega = 18$. Although these values are inconsequential for the simulation results because our tests are numerically invariant to location-scale affine transformations of the observations, in annualised terms they imply a realistic risk premia of 6%, a standard deviation 14.7%, and a Sharpe ratio .41 under the null.

For the sake of brevity, we focus on the results for $T = 100$ observations (plus another 112 for initialisation), equivalent to 25 years of quarterly data, roughly the same as in our empirical analysis. The ARMA(1,1) specification corresponds to the reduced form of the model with smooth, persistent autoregressive expected returns observed subject to negatively correlated

noise that underlies the mean reversion literature (see e.g. Fiorentini and Sentana (1998) and the references therein). When $\varphi = 0$, this mean specification reduces to the AR(1) process that we considered in section 2.1. In general, though, it nests neither this process nor the restricted AR(h) process in section 2.2. As a result, we can use it to assess which of the tests we developed in those sections has more power to detect predictability in a realistic context.

We systematically rely on 20,000 replications, so the 95% confidence interval for nominal sizes of 1, 5 and 10% would be $(0.86, 1.14)$, $(4.70, 5.30)$ and $(9.58, 10.42)$%, respectively.

As for $\boldsymbol{\eta}_0$, we consider five different standardised distributions: Gaussian, Student t_6, a mixture of two normals with the same kurtosis (=6) but negative skewness (=-.5), and two 4^{th}-order Gram Charlier expansions: GC(0,3.0) - symmetric - and GC(-0.8, 3.0) - asymmetric. Finally, we also consider a sixth distribution to assess the sensitivity of the different tests to the presence of unusual values. Specifically, we replace five observations of the GC(-0.8, 3.0) random variable with additive outliers four standard deviations away from the mean, two of which are consecutive. In principle, influential observations may have three possibly compensating effects on the different testing procedures. First, they can affect the estimates of the mean and variance of the distribution, and thereby the distribution of the standardised residuals. Second, even if we could observe the true standardised innovations, the presence of two consecutive unusual values can generate large test statistics even though the null is true. And finally, in as much as large observations are a reflection of a non-Gaussian distribution, they will tend to increase the power of the non-Gaussian tests relative to the Gaussian ones.

We use the same underlying pseudo-random numbers in all designs to minimise experimental error. In particular, we make sure that the underlying Gaussian random variables are the same for all five distributions. Given that the usual routines for simulating gamma random variables involve some degree of rejection, which unfortunately can change for different values of the shape parameters, we use the slower but smooth inversion method based on the NAG G01FFF gamma quantile function so that we can keep the underlying uniform variates fixed across simulations. Those uniform random variables are also recycled to generate the normal mixture.

For each Monte Carlo sample thus generated, our ML estimation procedure employs the following numerical strategy. First, we estimate the static mean and variance parameters $\boldsymbol{\theta}_s$ under normality using (7) and (8). Then, we compute the sample coefficient of kurtosis κ, on the basis of which we obtain the sequential Method of Moments estimator of the shape parameter of the t distribution suggested by Fiorentini, Sentana and Calzolari (2003), which exploits the theoretical relationship $\eta = \max[0, (\kappa - 3)/(4\kappa - 6)]$. Next, we use this estimator as initial value for a univariate optimisation procedure that uses the E04ABF routine to obtain a sequential

ML estimator of η, keeping π and ω fixed at their Gaussian PML estimators. The resulting estimates of η, together with the PMLE of $\boldsymbol{\theta}_s$, become the initial values for the t-based ML estimators. Following Fiorentini, Sentana and Calzolari (2003), the final stage of our 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 analytical expressions for the score and information matrix $\mathcal{I}(\boldsymbol{\phi}_0)$ derived in section 2 of that paper. We rule out numerically problematic solutions by imposing the inequality constraints $0 \leq \eta \leq .499$. As for the discrete mixture of normals, we use the EM algorithm described in appendix D.7 to obtain good initial values, and then we numerically maximise the log-likelihood function of y_t in terms of the shape parameters $\boldsymbol{\eta} = (\delta, \upsilon, \lambda)'$ keeping $\boldsymbol{\theta}_s$ fixed at their Gaussian ML estimates. To reduce the chances that the mixture ML estimator corresponds to a singular configuration in which the relative variance parameter υ is either 0 or infinity, we repeat the EM optimisation using 100 different starting values.

Computational details for the symmetric and general semiparametric procedures can be found in appendix B of Fiorentini and Sentana (2018). Given that a proper cross-validation procedure is extremely costly to implement in a Monte Carlo exercise, we have chosen the "optimal" bandwidth in Silverman (1986).[11]

For each Monte Carlo sample we compute the predictability tests based on six different scores: Gaussian, Student t, discrete location-scale mixture of two normals (DLSMN), Laplace, the efficient semiparametric score and an efficient semiparametric score that imposes symmetry of the innovation distribution.

5.2 Finite sample size

We report the size properties of the different predictability tests under the null in Table 1A, which displays the empirical rejection rates of the different tests at the conventional 10, 5 and 1% significance levels. The small sample size of the first-order serial correlation tests are rather accurate for every outlier-free distribution. In contrast, the restricted 12^{th}-order serial correlation tests discussed in section 2.2 and the conditional heteroskedasticity tests in section 3 show some moderate size distortions. Those distortions are more pronounced for the ARCH(1) tests, which tend to under-reject the null, and less evident for the GARCH(1,1) tests calculated with the

[11] Nevertheless, the optimality of this bandwidth for density estimation purposes does not necessarily extend to the estimation of the efficient score, as illustrated in Robinson (2010). See e.g. Prakasa Rao (1983) for a formal statistical treatment of kernel density estimation, including regularity conditions on the tail behaviour of the distributions which are estimated.

discount factor $\bar{\beta} = .84$.[12] In this case, the under-rejection is attenuated and sometimes even reversed except for the Gaussian-based test, which still suffers from a noticeable small sample size distortion when the true distribution is not Gaussian. Nevertheless, in simulation exercises with $T = 500$ and $T = 1000$ (available upon request), all size distortions quickly disappear as the sample length increases.

But when the distribution is contaminated with additive outliers, the size distortions reported in the bottom panel of Table 1A become remarkable for all tests but the restricted serial correlation tests, which benefit from the compound regressor averaging the outliers out. As one might have expected, though, the size distortions of the Student and mixture-based tests are acceptable, so these tests turn out to be rather robust in this case. In contrast, the other tests and particularly the Gaussian tests display a rather wild behaviour. For example, the Gaussian test against GARCH(1,1) almost never rejects even at a the 10% significance level, while the Gaussian AR(1) and ARCH(1) tests massively reject their nulls.

As is well known, the bootstrap often manages to partly correct the finite sample size distortions evidenced in simulation exercises. Unfortunately, given that the semiparametric tests are computationally intensive, it would be incredibly time consuming to carry out standard versions of the bootstrap with a large number of samples B for each of the 20,000 Monte Carlo simulations. For that reason, we adapt the so-called warp-speed bootstrap procedure of Giacomini, Politis and White (2013) to our testing framework. Specifically, we set $B = 1$, so that effectively there is a single bootstrap sample for each of the 20,000 simulated samples, but then pool those 20,000 bootstrap samples to obtain the relevant critical values. The fact that the (pseudo) maximum likelihood estimators of π and ω based on the affine transformation of the observations $a + by_t$ will be precisely $a + b\hat{\pi}_T$ and $b^2\hat{\omega}_T^2$ regardless of the correct specification of the distribution, means that a non-parametric bootstrap procedure applied to the original observations y_t or the estimated standardised residuals $\epsilon_t(\hat{\pi}_T, \hat{\omega}_T^2)$ yield identical test statistics in each bootstrap sample.

The results in Table 1B show that these bootstrap critical values render all tests very accurate even in samples of size $T = 100$ for every outlier-free distribution. For this reason, the finite sample power results in the next section will use bootstrap critical values. In contrast, the Gaussian tests continue to behave wildly in the outlier case because the bootstrap procedure destroys the artificial dependence introduced by the two consecutive outliers.

[12] The conventional discount factor $\bar{\beta} = .94$ suggested in RiskMetrics (1999) for daily data seems inappropriate for the quarterly data used in the empirical application.

5.3 Finite sample power

In order to gauge the power of the serial correlation tests we look at a design in which $\rho = 2/\sqrt{T}$ but $\alpha = 0$. The evidence at the 10, 5 and 1% significance level is presented in first two groups of columns of Table 1C, which includes rejection rates based on bootstrap critical values rather than nominal ones. The bootstrap procedure adopted under the null remains valid under the alternative too because the resampling scheme destroys the dynamic dependence in y_t arising from the presence of serial correlation or conditional heteroskedasticity.

As expected from the theoretical analysis in section 2.3, our proposed non-Gaussian tests show some power gains over standard Gaussian procedures when the true distribution is non-normal, with the parametric tests performing on par with the semiparametric ones even in those situations in which the assumed distribution is misspecified. Nevertheless, the mixture-based serial correlation test is not the most powerful when this distribution is correctly specified. This surprising result reflects the poor performance of the ML estimators of the mixture shape parameters when $T = 100$, which implies that the estimated distribution is in practice misspecified. We will return to this point at the end of this section. Finally, the restricted 12^{th}-order serial correlation tests show very little power under the AR(1) alternative, thereby confirming the results depicted in Figure 3A.

We also look at a persistent ARMA(1,1) design with $\rho = 0.98$ and $\varphi = -0.92$, which implies a restricted infinite-order autoregressive process with $\rho_i = 0.06 \times 0.92^{i-1}$. The results of these alternative experiments are displayed in the two rightmost groups of columns of Table 1C. Our tests against a restricted AR(12) process show substantial power gains over the first-order serial correlation test in this context despite the fact that the true alternative differs from the ones for which those two tests are optimal. Moreover, our proposed non-Gaussian tests also show clear power gains over standard (i.e. Gaussian) tests in the presence of non-normal distributions, including when the distribution is misspecified.

Turning to the variance predictability tests in Table 4D, we consider a design with $\rho = 0$ but $\alpha = 2.5/\sqrt{T}$ and $\beta = 0$ to assess the power of the ARCH(1) tests. We find again that the usual Gaussian tests are usually worse than our flexible parametric tests. We also find that when the true DGP follows an ARCH(1) process, the GARCH(1,1) tests have small but non-negligible power, a situation which gets reversed when we simulate with $\beta = 0.88$ and $\alpha = 1/\sqrt{T}$. In this case, the GARCH(1,1) tests have substantially more power than the ARCH(1) statistics even though the assumed dumping factor $\bar{\beta} = .84$ differs from the true one.

Once more, the poor performance of the mixture ML estimators in small sample sizes implies that the mixture-based test is not the most powerful under correct distributional specification.

To confirm that this is a small sample problem, we consider a simulation exercise with $T = 1,000$, a common sample size in empirical finance applications even with low frequency data. For example, the original three FF factor portfolios for the US comprise $1,074$ monthly observations as of December 2018. Importantly, though, by dividing the parameter values by the square root of the sample size, the alternative hypotheses we analyse in this larger sample exercise are comparable with those in the simulations with $T = 100$.

The results presented in Table 1E show that the mixture-based tests are indeed the most powerful under correct specification, which confirms the relevance in larger samples of the theoretical results reported in Figure 3C. In addition, the Gaussian tests are systematically dominated by all our flexible proposals, with the Student t achieving almost full efficiency despite being distributionally misspecified. In contrast, the semiparametric tests are only slightly better than the Gaussian ones, so once again they fail to achieve maximum power even though the sample size is ten times as large. Finally, the Laplace conditional mean tests turn out to be worse than the Gaussian ones, which probably reflects the fact that it is not a flexible distribution either.

6 Empirical application

We apply the procedures studied previously to the five FF factors for international stocks, which we have obtained from Ken French's Data Library (see Fama and French (1993, 2012, 2015) and <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html> for further details). Those five factors are constructed using 6 value-weight portfolios formed on size and book-to-market, 6 value-weight portfolios formed on size and operating profitability, and 6 value-weight portfolios formed on size and investment. The exact factor definition is as follows:

1. MK is the return on a region's value-weight market portfolio minus the relevant safe rate.
2. SMB contains the returns on small cap firms in excess of the returns on large cap firms.
3. HML are the returns on value firms in excess of the returns on growth firms.
4. RMW contains the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios.
5. CMA is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios

We consider portfolios for four world regions: North America (US and Canada), Europe (Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden and Switzerland), Japan, and "Asia Pacific ex Japan" (Australia, Hong Kong, New Zealand and Singapore). All returns are in US dollars, include dividends and capital gains, and are not continuously compounded. We construct quarterly

data from July 1991 to September 2018, with the US 3-month Tbill rate as the safe asset. All in all, our sample contains 113 observations because we reserve the first quarters for pre-sample values (see appendix C.4 for further details).

Figure 8 displays the temporal evolution of the returns on those portfolios over the sample. We have highlighted two periods in grey: the years 1999 and 2000, which correspond to the dotcom bubble, and the global financial crisis, which goes from the last quarter of 2007 to the first quarter of 2009. Each subplot also contains lower and upper confidence bands at the first quartile minus 1.5 times the interquartile range (IQR) and the third quartile plus 1.5·IQR. If the return distributions were normal, those bands would be centred around the mean with width ± 2.7 standard deviations, so they would cover 99.3% of the observations. For that reason, they are often used in the robust statistical literature as a simple visual device to detect "unusual" observations, with the IQR being preferred to the sample standard deviation because of its lower sensitivity to outliers. As can be seen, in practice the number of "influential" observations is much larger than what one would expect from a Gaussian distribution.

We present more formal statistics in Table 2. As expected from the plots in Figure 8, the most striking result is the extent to which many of the FF factor returns have fat tails at the quarterly frequency. Specifically, if we use the one-sided test of the null of mesokurtosis against the alternative of leptokurtosis in Fiorentini, Sentana and Calzolari (2003), we reject at the 5% significance level in 16 out of 20 cases, with the Value and Investment factors systematically showing excess kurtosis.

In turn, we test for symmetry by means of a moment test of the null hypothesis H_0 : $E[H_3(y; m, o)] = 0$, which we compute as the t-ratio of the sample mean of the third Hermite polynomial of the standardised residuals evaluated at the sample mean and standard deviation of the returns (see appendix C.3.1). As can be seen, we cannot reject the null hypothesis that the quarterly returns on all the different portfolios are symmetric, with the exception of the North American stock market, which shows statistically significant negative skewness. Although it might be the case that the moment test that we use has low power to detect asymmetries, the usual skewness component of the Jarque-Bera (1980) test, which assumes normal returns, shows massive size distortions for symmetric but leptokurtic distributions.

In any event, the results in Table 2 clearly motivate the use of our proposed tests for serial dependence. In this regard, Tables 3A-3D report the results of the different mean and variance predictability tests across the four regions. The first column of each table displays the results of the first-order serial correlation test. Similarly, the second and third columns show the results of our tests against restricted AR(4) and AR(12) alternatives, respectively. As we mentioned before,

these tests are effectively testing the presence of 1- and 3-year time series momentum in quarterly returns, as in Moskowitz, Ooi and Pedersen (2012). In turn, the fourth and fifth columns display the results of our one-sided tests for ARCH and GARCH effects. Testing for conditional heteroskedasticity in quarterly returns might seem unjustified because many academics and financial markets participants believe that high frequency movements in volatility mean revert at a simple, non-negative exponential rate so that they wash out at such a low frequency. However, the persistent movements in volatility indices such as the VIX, whose central tendency seems to fluctuate over a long-run mean for several years (see e.g. Mencía and Sentana (2013)), suggest that volatility changes might still be relevant at low frequencies.

As can be seen, we find substantial differences across regions. In North America, we do not reject the null hypothesis of mean predictability, with the possible exception of the profitability factor, for which some of the tests find 3-year momentum effects. Interestingly, the Gaussian test for first-order serial correlation applied to the value factor rejects its null in marked contrast to all the other tests. This rejection seems to be driven by the presence of some unusually large observations of the same sign during the dotcom bubble and its aftermath. In contrast, we systematically find evidence for conditional heteroskedasticity.

On the other hand, we find that the quarterly returns on the European value factor are predictable, both in the short run and in the long run. Once again, the Gaussian test for first-order serial correlation in the investment factor is the only one that rejects the null. As for the conditionally heteroskedasticity tests, we find evidence for ARCH effects for the same two factors and the market, as well as more persistent changes in volatility for the size factor.

The evidence for Japan is also different. Aside from finding ARCH(1) effects in the market portfolio, and GARCH effects in all the other factors except investment, we do not find any evidence of predictability in levels, except if we rely on the Gaussian tests, which once again are the only ones that reject the null for the value and investment factors.

Finally, in Asia Pacific we find predictability for the value factor using one-year momentum, and first-order serial correlation for the profitability factor. There is also evidence against conditional homoskedasticity for all factors except size, which is stronger for persistent changes in volatility than for short-run movements.

7 Conclusions

We propose more powerful score tests of predictability in the levels and squares of financial returns by exploiting the non-normality of their distributions. For our purposes the conditional distribution of returns can be either parametrically or non-parametrically specified.

We show that our score tests are equivalent to standard orthogonality tests of predictability in which the regressand has been multiplied by a damping factor that reflects the skewness and kurtosis of the data, as in the robust estimation literature. Thereby, we achieve two important improvements over the usual Gaussian tests advocated by White (1982) and Bollerslev and Wooldridge (1992) among many others: increases in their local power and reductions in their sensitivity to influential observations. Both these improvements are very useful from a practical point of view because they will allow researchers to go beyond the binary question of the presence or absence of mean-variance predictability, helping them understand better which predictors are really relevant. In this regard, we also explain how to transform the regressor to exploit the persistence of expected returns and volatilities.

Importantly, we prove that our parametric tests remain valid regardless of whether or not the assumed distribution is correct, which puts them on par with the Gaussian testing procedures. We also show that our semiparametric tests should be (locally) as powerful as if we knew the true distribution of the data.

We present local power analyses which confirm that irrespective of whether the parametric distribution is correctly specified, there are clear power gains from exploiting the non-normality of financial returns, as well as the persistent behaviour of risk premia and volatility. We complement our theoretical results with detailed Monte Carlo exercises that assess the reliability of our predictability tests in finite samples. We also show that straightforward non-parametric bootstrap procedures correct the observed small size distortions. In addition, we verify that our parametric tests offer clear power gains over the usual Gaussian procedures even in those situations in which the assumed distribution is misspecified. Finally, we also observe that the finite sample power of the semiparametric procedures is not well approximated by the first-order asymptotic theory that justifies their adaptivity, not even in samples of 1,000 observations.

Finally, we apply our methods to quarterly stock returns on the five FF factors for international stocks, which in most cases have fat tailed symmetric distributions. Our results highlight noticeable differences across regions and factors. While we find no evidence in favour of either short-run serial correlation or long-run momentum for the different market portfolios, we find persistent components in the European and Asian-Pacific value factors, as well as a few of the profitability and investment factors. We also find stronger evidence for persistent serial correlation in the volatility of many series, but certainly not all. Importantly, the inability of the extant Gaussian tests to deal with unusually large observations sometimes results in rejections that are not supported by the robust tests.

Multivariate extensions of our testing procedures are simple in theory but difficult in practice

because of the curse of dimensionality. For that reason, in Fiorentini and Sentana (2015) we imposed a parsimonious factor structure both under the null and under the different alternatives.

We could study the effect of replacing the kernel-based non-parametric density estimators that we have considered by either positive Hermite expansions of the normal density (see e.g. León, Mencía and Sentana (2009)), or discrete normal mixture models with multiple underlying components. In this sense, it is worth mentioning that the robustness of the parametric dynamic specification tests that we have highlighted holds for those flexible distributions for any finite number of terms. In addition, one would expect that the larger the number of components, the closer one would get to achieving the adaptivity of the semiparametric tests. In this regard, an interesting question is the effect of overparametrising the parametric distribution. For example, imagine that the true distribution is Gaussian but we estimate the model under the null using a fourth-order GC expansion. The block-diagonality between conditional mean and variance parameters on the one hand, and shape parameters on the other in Proposition 3 of Fiorentini and Sentana (2007) suggests that there should be no efficiency loss in conducting the tests using the maximum likelihood estimates of the overparametrised GC distribution.

Another interesting extension would be to consider non-parametric alternatives, in which the lag length is implicitly determined by the choice of bandwidth parameter in a kernel-based estimator of a spectral density matrix (see e.g. Hong (1996) and Hong and Shehadeh (1999)). In addition, we could test for the effect of exogenous regressors in either the conditional mean or the conditional variance. Finally, it would be interesting to extend our mean-variance predictability tests so that they apply to the residuals of models which are already dynamic under the null. We are currently exploring some of these interesting research avenues.

Appendix

A Proofs

Proposition 1

Given the discussion in appendix D, to find the score function and conditional information matrix all we need is the matrix $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s)$, which in turn requires the Jacobian of the conditional mean and covariance functions. In view of (1), we will have that

$$\partial \mu_t(\pi, 0, \omega)/\partial \boldsymbol{\theta}' = \begin{pmatrix} 1 & y_{t-1} - \pi & 0 \end{pmatrix}$$

and

$$\partial \sigma_t^2(\pi, 0, \omega)/\partial \boldsymbol{\theta}' = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix},$$

whence

$$\mathbf{Z}_{dt}(\pi, 0, \omega) = \begin{bmatrix} \omega^{-1/2} & 0 \\ \epsilon_{t-1}(\boldsymbol{\theta}_s) & 0 \\ 0 & \frac{1}{2}\omega^{-1} \end{bmatrix}, \tag{A1}$$

so that

$$\mathbf{Z}_{d}(\pi_0, 0, \omega_0, \boldsymbol{\eta}_0) = \begin{bmatrix} \omega_0^{-1/2} & 0 \\ 0 & 0 \\ 0 & \frac{1}{2}\omega_0^{-1} \end{bmatrix}. \tag{A2}$$

As a result, the score under the null will be

$$\begin{bmatrix} s_{\boldsymbol{\pi}t}(\pi, 0, \omega, \boldsymbol{\eta}) \\ s_{\rho t}(\pi, 0, \omega, \boldsymbol{\eta}) \\ s_{\omega t}(\pi, 0, \omega, \boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} -\omega^{-1/2}\partial f\left[\epsilon_t(\boldsymbol{\theta}_s), \eta\right]/\partial \varepsilon^* \\ -\partial f\left[\epsilon_t(\boldsymbol{\theta}_s), \eta\right]/\partial \varepsilon^* \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s) \\ -\frac{1}{2}\omega^{-1}[\partial f\left[\epsilon_t(\boldsymbol{\theta}_s), \eta\right]/\partial \varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + 1] \end{bmatrix}.$$

Similarly, the conditional information matrix will be

$$\begin{bmatrix} \omega^{-1/2} & 0 & \mathbf{0} \\ \epsilon_{t-1}(\boldsymbol{\theta}_s) & 0 & \mathbf{0} \\ 0 & \frac{1}{2}\omega^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathcal{M}'_{lr}(\boldsymbol{\eta}) & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix} \begin{bmatrix} \omega^{-1/2} & \epsilon_{t-1}(\boldsymbol{\theta}_s) & 0 & \mathbf{0} \\ 0 & 0 & \frac{1}{2}\omega^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix}$$

$$= \begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{M}_{ll}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & \omega^{-1/2}\mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{M}_{ll}(\boldsymbol{\eta}) & \epsilon_{t-1}^2(\boldsymbol{\theta}_s)\mathcal{M}_{ll}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{M}_{ls}(\boldsymbol{\eta}) & \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{M}_{ls}(\boldsymbol{\eta}) & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \omega^{-1}\mathcal{M}'_{lr}(\boldsymbol{\eta}) & \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{M}'_{lr}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{bmatrix},$$

while the unconditional one becomes

$$\begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & \omega^{-1/2}\mathcal{M}_{lr}(\boldsymbol{\eta}) \\ 0 & \mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & 0 \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & 0 & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \omega^{-1/2}\mathcal{M}'_{lr}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-1}\mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{bmatrix}.$$

This result confirms the expression for $\mathcal{I}_{\rho\rho}(\boldsymbol{\phi})$, as well as the fact that the sampling uncertainty in the ML estimators of π, ω and $\boldsymbol{\eta}$ is inconsequential for the asymptotic distribution of the test, at least up to first order. □

Proposition 2

As discussed in appendix D.2, the asymptotic distribution of the Gaussian Pseudo ML estimators and tests will depend on

$$\begin{aligned}\mathcal{A}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0) &= E[\mathcal{A}_{\boldsymbol{\theta\theta}t}\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0)|\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0],\\ \mathcal{A}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\theta},\boldsymbol{\varrho}) &= -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(0,0)\mathbf{Z}'_{dt}(\boldsymbol{\theta})\end{aligned}$$

and

$$\mathcal{B}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\theta},\boldsymbol{\varrho}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}]] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\varphi,\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),$$

where

$$\mathcal{K}\,(\varphi,\kappa) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t,I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}]] = \begin{bmatrix} 1 & \varphi(\boldsymbol{\varrho}) \\ \varphi(\boldsymbol{\varrho}) & \kappa(\boldsymbol{\varrho})-1 \end{bmatrix}$$

and $\boldsymbol{\varrho}$ are the shape parameters of the true distribution of ε_t^*.

But given the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ in (A1) and the consistency of the Gaussian PML estimators of π and ω, which implies that $E[\epsilon_t(\boldsymbol{\theta}_{s0})|\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0]=0$, it is clear that $\mathcal{A}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0)$ will be block diagonal between ρ and $\boldsymbol{\theta}_s$ irrespective of the true distribution of y_t. In addition, $\mathcal{A}_{\rho\rho}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0)$ will coincide with $\mathcal{I}_{\rho\rho}(\boldsymbol{\theta}_s,0,\boldsymbol{\varrho}_0)$. A closely related argument shows that $\mathcal{B}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\theta},\boldsymbol{\varrho})$ will also be block diagonal between ρ and $\boldsymbol{\theta}_s$, and that $\mathcal{B}_{\rho\rho}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0) = \mathcal{A}_{\rho\rho}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0)$. As a result, the Gaussian-based LM test for $H_0:\rho=0$ remains valid irrespective of the true distribution of y_t. □

Proposition 3

We can use standard arguments (see e.g. Newey and McFadden (1994)) to show that

$$\begin{aligned}\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} s_{\rho t}(\hat{\boldsymbol{\phi}}_s,0) &= \frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} s_{\rho t}(\boldsymbol{\phi}_{s\infty},0) + \frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbf{h}_{\rho\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)\sqrt{T}(\hat{\boldsymbol{\phi}}_s-\boldsymbol{\phi}_{s\infty}) + o_p(1)\\ &= \frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} s_{\rho t}(\boldsymbol{\phi}_{s\infty},0) - \frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbf{h}_{\rho\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)\left[\frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbf{h}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)\right]^{-1}\\ &\quad\times\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T}\mathbf{s}_{\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0) + o_p(1),\end{aligned}$$

where $\boldsymbol{\phi}_s = (\boldsymbol{\theta}_s',\boldsymbol{\eta}')'$. Hence, the asymptotic variance of $\frac{\sqrt{T}}{T}\sum_{t=1}^{T} s_{\rho t}(\hat{\boldsymbol{\phi}}_s,0)$ will be given by $\mathcal{F}_{\rho\rho}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_\infty;\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0)$, where

$$\mathcal{F}_{\rho\rho} = \mathcal{B}_{\rho\rho} - 2\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{B}'_{\rho\boldsymbol{\phi}_s} + \mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}'_{\rho\boldsymbol{\phi}_s},$$

and $\mathcal{B}_{\rho\rho}$, $\mathcal{A}_{\rho\boldsymbol{\phi}_s}$, etc. are the relevant elements of

$$\begin{aligned}\mathcal{B}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0) &= V[s_{\boldsymbol{\phi}t}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta})|\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0),\\ \mathcal{A}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0) &= -E[h_{\boldsymbol{\phi\phi}t}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta})|\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0).\end{aligned}$$

Tedious but straightforward algebra shows that at $\rho = 0$:

$$\begin{aligned}
h_{\pi\pi t}(\boldsymbol{\phi}) &= \omega^{-1}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \\
h_{\pi\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-3/2}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\} \\
\mathbf{h}_{\pi\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= -\omega^{-1/2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}' \\
h_{\omega\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-2}\{1 + \tfrac{3}{2}\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + \tfrac{1}{2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t^2(\boldsymbol{\theta}_s)\} \\
\mathbf{h}_{\omega\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= -\tfrac{1}{2}\omega^{-2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}' \cdot \epsilon_t(\boldsymbol{\theta}_s) \\
\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= \partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'
\end{aligned}$$

Similarly, we can show that at $\rho = 0$

$$\begin{aligned}
h_{\rho\pi t}(\boldsymbol{\phi}) &= \omega^{-1/2}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\} \\
h_{\rho\rho t}(\boldsymbol{\phi}) &= \partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_{t-1}^2(\boldsymbol{\theta}_s) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^* \cdot \epsilon_{t-2}(\boldsymbol{\theta}_s) \\
h_{\rho\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-1}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\} \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s) \\
\mathbf{h}_{\rho\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= -\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta} \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s)
\end{aligned}$$

Given that the pseudo-true values of π, ω and $\boldsymbol{\eta}$ are implicitly defined in such a way that

$$\begin{aligned}
E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}), \boldsymbol{\eta}_\infty\right]/\partial\varepsilon^* | \boldsymbol{\varphi}_0\} &= 0, \\
E\{1 + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}), \boldsymbol{\eta}_\infty\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_{s\infty}) | \boldsymbol{\varphi}_0\} &= 0, \\
E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}), \boldsymbol{\eta}_\infty\right]/\partial\boldsymbol{\eta} | \boldsymbol{\varphi}_0\} &= \mathbf{0},
\end{aligned}$$

the law of iterated expectations implies that

$$\begin{aligned}
E[h_{\pi\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \omega_\infty^{-1}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[h_{\pi\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-3/2}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[\mathbf{h}_{\pi\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= -\omega_\infty^{-1/2}\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) - 1] \\
E[\mathbf{h}_{\omega\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \mathcal{H}_{rr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)
\end{aligned}$$

and

$$\begin{aligned}
E[h_{\rho\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \omega_\infty^{-1/2}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}(\boldsymbol{\theta}_{s\infty}) \\
E[h_{\rho\rho t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty}) \\
E[h_{\rho\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{ls}(\boldsymbol{\phi}; \boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}(\boldsymbol{\theta}_{s\infty}) \\
E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= -\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^* | I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) | I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}' | I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t^2(\boldsymbol{\theta}_s) | I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}' \cdot \epsilon_t(\boldsymbol{\theta}_s) | I_{t-1};\boldsymbol{\varphi}_0]
\end{aligned}$$

and $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}'_{s0}, 0, \boldsymbol{\varrho}'_0)'$.

Consequently,

$$\begin{aligned}
E[h_{\rho\pi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \omega_\infty^{-1/2}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] \\
E[h_{\rho\rho t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] \\
E[h_{\rho\omega t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] \\
E[\mathbf{h}_{\rho\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]
\end{aligned}$$

where

$$E[\epsilon_t(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0] = E[\omega^{-1/2}(y_t-\pi)|\boldsymbol{\varphi}_0] = E[\omega^{-1/2}(\pi_0+\omega_0^{1/2}\varepsilon_t^*-\pi)|\boldsymbol{\varphi}_0] = \omega^{-1/2}(\pi_0-\pi)$$

and

$$E[\epsilon_t^2(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0] = E[\omega^{-1}(y_t-\pi)^2|\boldsymbol{\varphi}_0] = E[\omega^{-1}(\pi_0+\omega_0^{1/2}\varepsilon_t^*-\pi)^2|\boldsymbol{\varphi}_0] = \omega^{-1}[(\pi_0-\pi)^2+\omega_0],$$

so that

$$V[\epsilon_t(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0] = \omega^{-1}\omega_0. \tag{A3}$$

Given that $\mathcal{A}_{\rho\boldsymbol{\phi}_s}$ is proportional to the first column of $\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}$, we can immediately show that

$$\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}^{-1} = \begin{pmatrix} E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\sqrt{\omega_\infty} & 0 & \mathbf{0}' \end{pmatrix} = E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1' \tag{A4}$$

if we evaluate these expressions at the pseudo true values. Therefore, the only elements of $\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$ that we need are the ones corresponding to π and ρ. But since

$$\begin{aligned}
\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) &= E[\mathcal{B}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)|\boldsymbol{\varphi}_\infty], \\
\mathcal{B}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) &= V[\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_\infty)|\, I_{t-1};\boldsymbol{\varphi}_\infty] = \mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{K}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathbf{Z}_t'(\boldsymbol{\theta}_\infty), \\
\mathcal{K}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= V\left[\left.\begin{pmatrix} e_{lt}(\boldsymbol{\phi}) \\ e_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi}) \end{pmatrix}\right| \boldsymbol{\varphi}\right] = \begin{bmatrix} \mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{lr}'(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}'(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{bmatrix}
\end{aligned}$$

we will have that under the null of $H_0 : \rho = 0$,

$$\left[\begin{array}{cc} \mathcal{B}_{\pi\pi}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{B}_{\pi\rho}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{B}_{\pi\rho}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{B}_{\rho\rho}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \end{array} \right] = \mathcal{K}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \left[\begin{array}{cc} \omega_\infty^{-1} & \omega_\infty^{-1/2} E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] \\ \omega_\infty^{-1/2} E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] & E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] \end{array} \right].$$

Finally we obtain

$$\mathcal{F}_{\rho\rho}(\boldsymbol{\theta}_{s\infty}, 0, \boldsymbol{\eta}_\infty; \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0) = \mathcal{K}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) V[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0],$$

which is precisely the denominator of the R^2 in the regression of $\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}), \boldsymbol{\eta}\right] / \partial \varepsilon^*$ on a constant and $\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})$.

We can also use these expressions to derive the asymptotic variance of the pseudo ML estimator of ρ under the null. Specifically, straightforward algebra shows that the "$\rho\rho$" element of the matrix

$$\mathcal{C}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$$

will be given by

$$\frac{\mathcal{F}_{\rho\rho}(\boldsymbol{\theta}_{s\infty}, 0, \boldsymbol{\eta}_\infty; \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0)}{\mathcal{G}_{\rho\rho}^2(\boldsymbol{\theta}_{s\infty}, 0, \boldsymbol{\eta}_\infty; \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0)},$$

where

$$\mathcal{G}_{\rho\rho} = \mathcal{A}_{\rho\rho} - \mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}^{-1}\mathcal{A}_{\rho\boldsymbol{\phi}_s}'.$$

But (A4) immediate implies that

$$\begin{aligned} \mathcal{G}_{\rho\rho}(\boldsymbol{\theta}_{s\infty}, 0, \boldsymbol{\eta}_\infty; \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0) &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\left\{E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0] - E^2[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\right\} \\ &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) V[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0], \end{aligned}$$

whence

$$\sqrt{T}\hat{\rho}_T \to N\left[0, \frac{\mathcal{K}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)}{\mathcal{H}_{ll}^2(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)} \frac{\omega_\infty}{\omega_0}\right]$$

in view of (A3). Not surprisingly, this expression nests both the Gaussian PML expression in Proposition 2, as well as the true ML expression in Proposition 1.

Let us now find the remaining elements of $\mathcal{C}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$.

We need to find out an expression for $\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$, which is given by the unconditional

expected value of

$$\left[\begin{array}{ccc} \omega^{-1/2} & 0 & \mathbf{0} \\ \epsilon_{t-1}(\boldsymbol{\theta}_s) & 0 & \mathbf{0} \\ 0 & \frac{1}{2}\omega^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right] \left[\begin{array}{ccc} \mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}'_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array}\right] \left[\begin{array}{cccc} \omega^{-1/2} & \epsilon_{t-1}(\boldsymbol{\theta}_s) & 0 & \mathbf{0} \\ 0 & 0 & \frac{1}{2}\omega^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right]$$

$$\left[\begin{array}{ccc} \omega^{-1/2}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-1}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}'_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array}\right] \left[\begin{array}{cccc} \omega^{-1/2} & \epsilon_{t-1}(\boldsymbol{\theta}_s) & 0 & \mathbf{0} \\ 0 & 0 & \frac{1}{2}\omega^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right]$$

$$\left[\begin{array}{cccc} \omega^{-1}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon^2_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{4}\omega^{-2}\mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}'_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array}\right]$$

As for $\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$, we can use the partitioned inverse formula to write

$$\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \left(\begin{array}{cc} \mathcal{A}^{-1}_{\phi_s\phi_s} + \mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{A}'_{\rho\phi_s}\mathcal{G}^{-1}_{\rho\rho}\mathcal{A}_{\rho\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & -\mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{A}'_{\rho\phi_s}\mathcal{G}^{-1}_{\rho\rho} \\ -\mathcal{G}^{-1}_{\rho\rho}\mathcal{A}_{\rho\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & \mathcal{G}^{-1}_{\rho\rho} \end{array}\right).$$

But if we use the expression for $\mathcal{A}_{\rho\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s}$, we will get

$$\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \left(\begin{array}{cc} \mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \\ 0 & 0 \end{array}\right) + \mathcal{G}^{-1}_{\rho\rho}\left(\begin{array}{c} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}_1 \\ 1 \end{array}\right)\left(\begin{array}{cc} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}'_1 & 1 \end{array}\right)$$

Hence,

$$\begin{aligned} &\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \left(\begin{array}{cc} \mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{B}_{\phi_s\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \\ 0 & 0 \end{array}\right) \\ &+\mathcal{G}^{-1}_{\rho\rho}\left(\begin{array}{c} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}_1 \\ 1 \end{array}\right)\left(\begin{array}{cc} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}'_1 & 1 \end{array}\right)\left(\begin{array}{cc} \mathcal{B}_{\phi_s\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \\ \mathcal{B}_{\rho\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \end{array}\right) \\ &+\mathcal{G}^{-1}_{\rho\rho}\left(\begin{array}{cc} \mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{B}_{\phi_s\phi_s} & \mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{B}'_{\rho\phi_s} \\ 0 & 0 \end{array}\right)\left(\begin{array}{c} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}_1 \\ 1 \end{array}\right)\left(\begin{array}{cc} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}'_1 & 1 \end{array}\right) \\ &+\mathcal{F}_{\rho\rho}\mathcal{G}^{-2}_{\rho\rho}\left(\begin{array}{c} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}_1 \\ 1 \end{array}\right)\left(\begin{array}{cc} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}'_1 & 1 \end{array}\right). \end{aligned}$$

But

$$\left(\begin{array}{cc} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty\mathbf{e}'_1 & 1 \end{array}\right)\left(\begin{array}{cc} \mathcal{B}_{\phi_s\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \\ \mathcal{B}_{\rho\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \end{array}\right) = \mathbf{0}$$

because

$$\mathcal{B}_{\phi_s\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{A}'_{\rho\phi_s} = \left[\begin{array}{c} \omega^{-1}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array}\right] E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty$$

and

$$\mathcal{B}'_{\rho\phi_s} - \mathcal{B}_{\phi_s\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{A}'_{\rho\phi_s} = \left[\begin{array}{c} \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \epsilon_{t-1}(\boldsymbol{\theta}_s)\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array}\right] - \left[\begin{array}{c} \omega^{-1}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array}\right] E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega^{1/2}_\infty = \mathbf{0}.$$

As a result,

$$\mathcal{C}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$$
$$= \begin{pmatrix} \mathcal{A}^{-1}_{\phi_s\phi_s}\mathcal{B}_{\phi_s\phi_s}\mathcal{A}^{-1}_{\phi_s\phi_s} & 0 \\ 0 & 0 \end{pmatrix} + \mathcal{F}_{\rho\rho}\mathcal{G}^{-2}_{\rho\rho} \begin{pmatrix} E^2[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega_\infty \mathbf{e}_1\mathbf{e}_1' & -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1 \\ -E[\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1' & 1 \end{pmatrix},$$

which means that the PML estimator of ρ will be asymptotically orthogonal to the PML estimators of ω and $\boldsymbol{\eta}$, but not to the PML estimator of π.

It is also worth deriving the previous expressions for a fixed value of $\boldsymbol{\eta}$, so that we can say what would happen for a restricted pseudo ML estimator that fixes the shape parameters to some arbitrary value $\bar{\boldsymbol{\eta}}$. In this case, all the previous expressions remain valid after eliminating the rows and columns corresponding to $\boldsymbol{\eta}$, and replacing $\boldsymbol{\theta}_\infty$ by $\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}}) = [\pi_\infty(\bar{\boldsymbol{\eta}}), \omega_\infty(\bar{\boldsymbol{\eta}})]$, which are the values that solve the system of equations

$$\begin{aligned} E[\partial \ln f\{\epsilon_t[\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}})], \bar{\boldsymbol{\eta}}\}/\partial\varepsilon^*|\boldsymbol{\varphi}_0] &= 0, \\ E[1 + \partial \ln f\{\epsilon_t[\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}})], \bar{\boldsymbol{\eta}}\}/\partial\varepsilon^* \cdot \epsilon_t[\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}})]|\boldsymbol{\varphi}_0] &= 0. \end{aligned}$$

In fact, we would obtain exactly the same expressions even if fixed both ω and $\boldsymbol{\eta}$ to some arbitrary values $\bar{\omega}$ and $\bar{\boldsymbol{\eta}}$, as long as we replaced π_∞ by $\pi_\infty(\bar{\omega}, \bar{\boldsymbol{\eta}})$, which would be the value that solves

$$E[\partial \ln f\{\bar{\omega}^{-1/2}[y_t - \pi_\infty(\bar{\omega}, \bar{\boldsymbol{\eta}})], \bar{\boldsymbol{\eta}}\}/\partial\varepsilon^*|\boldsymbol{\varphi}_0] = 0.$$

□

Proposition 4

Given that

$$\mathbf{W}_d'(\pi_0, 0, \omega_0, \boldsymbol{\eta}_0) = \begin{pmatrix} 0 & \frac{1}{2}\omega_0^{-1} & 0 \end{pmatrix},$$

it is easy to see that the symmetric semiparametric efficient score and bound are given by:

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\left\{-\left[\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]/\partial\varepsilon^*\epsilon_t^2(\boldsymbol{\theta}_{s0}) + 1\right] - \frac{2}{\kappa - 1}\left[\epsilon_t^2(\boldsymbol{\theta}_0) - 1\right]\right\}$$

and

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \begin{bmatrix} \frac{1}{\omega}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & 0 \\ 0 & \mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 \\ 0 & 0 & \frac{1}{\omega^2(\kappa-1)} \end{bmatrix}.$$

Since this matrix is block diagonal and the efficiency bound for ρ coincides with the corresponding element of the information matrix under correct specification of the conditional

distribution, the asymptotic variance of the SSP estimator of this parameter coincides with that of the infeasible ML estimator which uses knowledge of the shape parameters $\boldsymbol{\eta}_0$. As a result, the non-centrality parameters will also be the same.

Similarly, we can use the expression for (A2) to show that the semiparametric efficient score will be given by:

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^{-1}(\varphi, \kappa)\,\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})\right],$$

while the semiparametric efficiency bound is

$$\begin{aligned}
\mathcal{S}(\phi_0) &= \begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) \\ 0 & \mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & 0 & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} - \\
&\quad \begin{pmatrix} \omega_0^{-1/2} & 0 \\ \mathbf{0} & \mathbf{0} \\ 0 & \frac{1}{2}\omega_0^{-1} \end{pmatrix} \left\{ \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} \right. \\
&\quad \left. - \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & \varphi \\ \varphi & \kappa - 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right\} \begin{pmatrix} \omega_0^{-1/2} & 0 & 0 \\ 0 & 0 & \frac{1}{2}\omega_0^{-1} \end{pmatrix} \\
&= \begin{bmatrix} \omega^{-1} & 0 & \frac{1}{2}\omega^{-3/2}\varphi \\ 0 & \mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 \\ \frac{1}{2}\omega^{-3/2}\varphi & 0 & \frac{1}{4}\omega^{-2}(\kappa - 1) \end{bmatrix}.
\end{aligned}$$

Given that this matrix is block diagonal and the efficiency bound for ρ coincides with the corresponding element of the information matrix under correct specification of the conditional distribution, the asymptotic variance of the SP estimator of this parameter coincides with that of the infeasible ML estimator which uses knowledge of the shape parameters $\boldsymbol{\eta}_0$. Consequently, the non-centrality parameters will also be the same. □

Lemma 1

The proof is trivial if we combine several results that appear in the proofs of Propositions 1, 2 and 4. □

Proposition 5

As explained in appendix D, we must start once again by finding an expression for the matrix $\mathbf{Z}_{dt}$. Given (14), we will have that

$$\partial\mu_t(\boldsymbol{\theta}_s, 0)/\partial\boldsymbol{\theta}' = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

and

$$\partial\sigma_t^2(\boldsymbol{\theta}_s, 0)/\partial\boldsymbol{\theta}' = \begin{pmatrix} 0 & 1 & (y_{t-1} - \pi)^2 - \omega \end{pmatrix},$$

whence

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_s, 0) = \begin{bmatrix} \omega^{-1/2} & 0 \\ 0 & \frac{1}{2}\omega^{-1} \\ 0 & \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1] \end{bmatrix}, \tag{A5}$$

so that

$$\mathbf{Z}_{d}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}_0) = \begin{bmatrix} \omega_0^{-1/2} & 0 \\ 0 & \frac{1}{2}\omega_0^{-1} \\ 0 & 0 \end{bmatrix}. \tag{A6}$$

As a result, the score under the null will be

$$\begin{bmatrix} s_{\boldsymbol{\pi}t}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}) \\ s_{\omega t}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}) \\ s_{\alpha t}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} -\omega^{-1/2}\partial f\left[\epsilon_t(\boldsymbol{\theta}_s), \eta\right]/\partial\varepsilon^* \\ -\frac{1}{2}\omega^{-1}[\partial f\left[\epsilon_t(\boldsymbol{\theta}_s), \eta\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + 1] \\ -\frac{1}{2}[\partial f\left[\epsilon_t(\boldsymbol{\theta}_s), \eta\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + 1][\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1] \end{bmatrix}.$$

Similarly, the conditional information matrix will be

$$\begin{aligned} &\begin{bmatrix} \omega^{-1/2} & 0 & \mathbf{0} \\ 0 & \frac{1}{2}\omega^{-1} & \mathbf{0} \\ 0 & \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathcal{M}_{lr}'(\boldsymbol{\eta}) & \mathcal{M}_{sr}'(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix} \\ &\times \begin{bmatrix} \omega^{-1/2} & 0 & 0 & \mathbf{0} \\ 0 & \frac{1}{2}\omega^{-1} & \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix} \\ = &\begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) \\ \frac{1}{2}\omega^{-1/2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]\mathcal{M}_{ls}(\boldsymbol{\eta}) & \frac{1}{4}\omega^{-1}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]\mathcal{M}_{ss}(\boldsymbol{\eta}) \\ \omega^{-1/2}\mathcal{M}_{lr}'(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\mathcal{M}_{sr}'(\boldsymbol{\eta}) \end{bmatrix} \\ &\begin{bmatrix} \frac{1}{2}\omega^{-1/2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]\mathcal{M}_{ls}(\boldsymbol{\eta}) & \omega^{-1/2}\mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \frac{1}{4}\omega^{-1}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]\mathcal{M}_{ss}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \frac{1}{4}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]^2\mathcal{M}_{ss}(\boldsymbol{\eta}) & \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]\mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1]\mathcal{M}_{sr}'(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{bmatrix}, \end{aligned}$$

while the unconditional one becomes

$$\begin{bmatrix} \frac{1}{\omega}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-1/2}\mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-1}\mathcal{M}_{sr}(\boldsymbol{\eta}) \\ 0 & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta}) & 0 \\ \omega^{-1/2}\mathcal{M}_{lr}'(\boldsymbol{\eta}) & \frac{1}{2}\omega^{-1}\mathcal{M}_{sr}'(\boldsymbol{\eta}) & 0 & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{bmatrix}.$$

This result confirms the expression for $\mathcal{I}_{\alpha\alpha}(\boldsymbol{\phi})$, as well as the fact that the sampling uncertainty in the ML estimators of π, ω and $\boldsymbol{\eta}$ is inconsequential for the asymptotic distribution of the test, at least up to first order.

Proposition 6

Once again, the asymptotic distribution of the Gaussian Pseudo ML estimators and tests will depend on

$$\begin{aligned} \mathcal{A}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_s, 0, \mathbf{0}; \boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0) &= E[\mathcal{A}_{\boldsymbol{\theta\theta}t}\boldsymbol{\theta}_s, 0, \mathbf{0}; \boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0)|\boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0], \\ \mathcal{A}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}, \mathbf{0}; \boldsymbol{\theta}, \boldsymbol{\varrho}) &= -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\,\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(0,0)\mathbf{Z}_{dt}'(\boldsymbol{\theta}) \end{aligned}$$

and

$$\mathcal{B}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\theta},\boldsymbol{\varrho}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{z}_t, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}]] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\varphi,\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),$$

where

$$\mathcal{K}(\varphi,\kappa) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}] = \begin{bmatrix} 1 & \varphi(\boldsymbol{\varrho}) \\ \varphi(\boldsymbol{\varrho}) & \kappa(\boldsymbol{\varrho})-1 \end{bmatrix}$$

and $\boldsymbol{\varrho}$ are the shape parameters of the true distribution of ε_t^*.

But given the structure of $\mathbf{Z}_{dt}(\boldsymbol{\theta})$ in (A5) and the consistency of the Gaussian PML estimators of π and ω, which implies that $E[\epsilon_t^2(\boldsymbol{\theta}_{s0})|\boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0] = 1$, it is clear that $\mathcal{A}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_s, 0, \mathbf{0}; \boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0)$ will be block diagonal between α and $\boldsymbol{\theta}_s$ irrespective of the true distribution of y_t. In addition, $\mathcal{A}_{\alpha\alpha}(\boldsymbol{\theta}_s, 0, \mathbf{0}; \boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0)$ will coincide with $\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_s, 0, \boldsymbol{\varrho}_0)$ provided that we use the true value of $\kappa(\boldsymbol{\varrho}) - 1$ instead of its value under normality. A closely related argument shows that $\mathcal{B}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0};\boldsymbol{\theta},\boldsymbol{\varrho})$ will also be block diagonal between α and $\boldsymbol{\theta}_s$, and that $\mathcal{B}_{\alpha\alpha}(\boldsymbol{\theta}_s, 0, \mathbf{0}; \boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0) = \frac{1}{2}[\kappa(\boldsymbol{\varrho})-1]\mathcal{A}_{\alpha\alpha}(\boldsymbol{\theta}_s, 0, \mathbf{0}; \boldsymbol{\theta}_0, 0, \boldsymbol{\varrho}_0)$. As a result, the Gaussian-based LM test for $H_0 : \alpha = 0$ remains valid irrespective of the true distribution of y_t as long as we replace the 2 in the denominator by the variance of the score. □

Proposition 7

Consider the following model:

$$\left.\begin{array}{c} y_t = \pi_0 + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \sigma_t^2(\boldsymbol{\theta}) = \omega[1 + \gamma(y_{t-1} - \pi)^2], \\ \varepsilon_t^*|I_{t-1};\pi,\omega,\gamma,\boldsymbol{\eta} \sim i.i.d.\ D(0,1,\boldsymbol{\eta}), \\ \text{with density function } f(.,\boldsymbol{\eta}) \end{array}\right\},$$

where the parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, $\boldsymbol{\theta}' = (\boldsymbol{\theta}_s', \gamma)'$ and $\boldsymbol{\theta}_s = (\pi, \omega)'$. In this context, the null hypothesis is $H_0 : \gamma = 0$.

It is then easy to see that

$$\frac{\partial \mu_t}{\partial \theta'} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

while

$$\frac{\partial \sigma_t^2}{\partial \theta'} = \begin{pmatrix} -2\omega\gamma(x_{t-1} - \pi) & 1 + \gamma(x_{t-1} - \pi)^2 & \omega(x_{t-1} - \pi)^2 \end{pmatrix}.$$

As a result, the score vector will be

$$\begin{aligned}
s_{\pi t} &= -\frac{1}{\{\omega[1 + \gamma(x_{t-1} - \pi)^2]\}^{1/2}} \frac{\partial \ln f\,[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varepsilon^*} + \frac{\gamma(x_{t-1} - \pi)}{[1 + \gamma(x_{t-1} - \pi)^2]} \left\{1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\,[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varepsilon^*}\right\}, \\
s_{\omega t} &= -\frac{1}{2\omega} \left\{1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\,[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varepsilon^*}\right\}, \\
s_{\gamma t} &= -\frac{(x_{t-1} - \pi)^2}{2[1 + \gamma(x_{t-1} - \pi)^2]} \left\{1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\,[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \varepsilon^*}\right\}, \\
s_{\boldsymbol{\eta} t} &= \frac{\partial \ln f\,[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]}{\partial \boldsymbol{\eta}}
\end{aligned}$$

which under the null of $\gamma = 0$ reduces to

$$\begin{aligned}
s_{\pi t} &= -\frac{1}{\omega^{1/2}} \frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]}{\partial \varepsilon^*}, \\
s_{\omega t} &= -\frac{1}{2\omega} \left\{ 1 + \epsilon_t(\boldsymbol{\theta}_s) \cdot \frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\}, \\
s_{\gamma t} &= -\frac{\omega}{2} \epsilon_{t-1}^2(\boldsymbol{\theta}_s) \left\{ 1 + \epsilon_t(\boldsymbol{\theta}_s) \cdot \frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\}, \\
s_{\boldsymbol{\eta} t} &= \frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}\right]}{\partial \boldsymbol{\eta}}.
\end{aligned}$$

Note that we could have obtained the same expressions by using the chain rule for first derivatives since

$$\begin{aligned}
s_{\omega t} &= -\frac{1 - \gamma\omega}{2\omega[1 + \gamma(x_{t-1} - \pi)^2]} \left\{ 1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\} \\
s_{\alpha t} &= -\frac{(x_{t-1} - \pi)^2 - \frac{\omega}{1-\gamma\omega}}{2\omega[1 + \gamma(x_{t-1} - \pi)^2]} \left\{ 1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\}
\end{aligned}$$

and

$$\frac{\partial \begin{pmatrix} \omega \\ \alpha \end{pmatrix}}{\partial \begin{pmatrix} \omega & \gamma \end{pmatrix}} = \begin{pmatrix} (1 - \gamma\omega)^{-2} & \omega^2(1 - \gamma\omega)^{-2} \\ \gamma & \omega \end{pmatrix}.$$

Similarly,

$$\begin{aligned}
h_{\pi\pi t}(\boldsymbol{\phi}) &= \frac{1}{\sigma_t^2} \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} - \frac{\omega\gamma\left(x_{t-1} - \pi\right)}{\sigma_t^3} \left(\frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} + \varepsilon_t^*(\boldsymbol{\theta}) \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} \right) \\
&\quad + \gamma \frac{-1 + \gamma(x_{t-1} - \pi)^2}{[1 + \gamma(x_{t-1} - \pi)^2]^2} \left\{ 1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\} \\
&\quad + \frac{\gamma(x_{t-1} - \pi)}{[1 + \gamma(x_{t-1} - \pi)^2]} \left(-\frac{1}{\sigma_t} + \frac{\omega\gamma(x_{t-1} - \pi)}{\sigma_t^2} \varepsilon_t^*(\boldsymbol{\theta}) \right) \left[\frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} + \varepsilon_t^*(\boldsymbol{\theta}) \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} \right]
\end{aligned}$$

$$h_{\pi\omega t}(\boldsymbol{\phi}) = -\frac{1}{2\omega} \left(-\frac{1}{\sigma_t} + \frac{\omega\gamma(x_{t-1} - \pi)}{\sigma_t^2} \varepsilon_t^*(\boldsymbol{\theta}) \right) \left[\frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} + \varepsilon_t^*(\boldsymbol{\theta}) \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} \right]$$

$$\begin{aligned}
h_{\pi\gamma t}(\boldsymbol{\phi}) &= \frac{(x_{t-1} - \pi)}{[1 + \gamma(x_{t-1} - \pi)^2]^2} \left\{ 1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\} \\
&\quad - \frac{(x_{t-1} - \pi)^2}{2[1 + \gamma(x_{t-1} - \pi)^2]} \left(-\frac{1}{\sigma_t} + \frac{\omega\gamma(x_{t-1} - \pi)}{\sigma_t^2} \varepsilon_t^*(\boldsymbol{\theta}) \right) \left\{ \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} + \varepsilon_t^*(\boldsymbol{\theta}) \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} \right\}
\end{aligned}$$

$$h_{\pi\boldsymbol{\eta} t}(\boldsymbol{\phi}) = \left(-\frac{1}{\sigma_t} + \frac{\omega\gamma(x_{t-1} - \pi)}{\sigma_t^2} \varepsilon_t^*(\boldsymbol{\theta}) \right) \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon_t^*(\boldsymbol{\theta}) \partial \boldsymbol{\eta}'}$$

$$\begin{aligned}
h_{\omega\omega t}(\boldsymbol{\phi}) &= \frac{1}{2\omega^2} \left\{ 1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} \right\} \\
&\quad + \frac{1}{4\omega^2} \left[\varepsilon_t^*(\boldsymbol{\theta}) \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} + \varepsilon_t^{*2}(\boldsymbol{\theta}) \cdot \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} \right]
\end{aligned}$$

$$h_{\omega\gamma t}(\boldsymbol{\phi}) = \frac{(x_{t-1} - \pi)^2}{4\omega[1 + \gamma(x_{t-1} - \pi)^2]} \left[\varepsilon_t^*(\boldsymbol{\theta}) \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^*} + \varepsilon_t^{*2}(\boldsymbol{\theta}) \cdot \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon^* \partial \varepsilon^*} \right]$$

$$h_{\omega\boldsymbol{\eta}t}(\boldsymbol{\phi}) = -\frac{1}{2\omega}\varepsilon_t^*(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\boldsymbol{\eta}}$$

$$\begin{aligned} h_{\gamma\gamma t}(\boldsymbol{\phi}) &= \frac{(x_{t-1}-\pi)^4}{2[1+\gamma(x_{t-1}-\pi)^2]^2}\left\{1+\varepsilon_t^*(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\ &\quad +\frac{(x_{t-1}-\pi)^4}{4[1+\gamma(x_{t-1}-\pi)^2]^2}\left[\varepsilon_t^*(\boldsymbol{\theta})\frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\varepsilon_t^{*2}(\boldsymbol{\theta})\cdot\frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right] \end{aligned}$$

$$h_{\gamma\boldsymbol{\eta}t}(\boldsymbol{\phi}) = -\frac{(x_{t-1}-\pi)^2}{2[1+\gamma(x_{t-1}-\pi)^2]}\varepsilon_t^*(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon_t^*(\boldsymbol{\theta})\partial\boldsymbol{\eta}'}$$

and

$$h_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}.$$

Under the null of $\gamma = 0$ these expressions reduce to

$$\begin{aligned} h_{\pi\pi t}(\boldsymbol{\phi}) &= \frac{1}{\omega}\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*} \\ h_{\pi\omega t}(\boldsymbol{\phi}) &= \frac{1}{2\omega^{3/2}}\left[\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right] \\ h_{\pi\gamma t}(\boldsymbol{\phi}) &= \omega^{1/2}\epsilon_{t-1}(\boldsymbol{\theta})\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\ &\quad +\frac{\omega^{1/2}}{2}\epsilon_{t-1}^2(\boldsymbol{\theta})\left\{\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right\} \\ \mathbf{h}_{\pi\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= -\frac{1}{\omega^{1/2}}\frac{\partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon_t^*(\boldsymbol{\theta})\partial\boldsymbol{\eta}'} \end{aligned}$$

$$\begin{aligned} h_{\gamma\gamma t}(\boldsymbol{\phi}) &= \frac{1}{2}\omega^2\epsilon_{t-1}^4(\boldsymbol{\theta})\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\ &\quad +\frac{1}{4}\omega^2\epsilon_{t-1}^4(\boldsymbol{\theta})\left[\epsilon_t(\boldsymbol{\theta})\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t^2(\boldsymbol{\theta})\cdot\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right] \\ \mathbf{h}_{\gamma\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= -\frac{1}{2}\omega\epsilon_{t-1}^2(\boldsymbol{\theta})\cdot\epsilon_t(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon_t^*(\boldsymbol{\theta})\partial\boldsymbol{\eta}'} \end{aligned}$$

and

$$h_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}$$

Given that the pseudo-true values of π, ω and $\boldsymbol{\eta}$ are implicitly defined in such a way that

$$\begin{aligned} E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*|\boldsymbol{\varphi}_0\} &= 0, \\ E\{1+\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\cdot\epsilon_t(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0\} &= 0, \\ E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\boldsymbol{\eta}|\boldsymbol{\varphi}_0\} &= \mathbf{0}, \end{aligned}$$

the law of iterated expectations implies that

$$
\begin{aligned}
E[h_{\pi\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \omega_\infty^{-1}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[h_{\pi\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-3/2}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[\mathbf{h}_{\pi\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= -\omega_\infty^{-1/2}\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]\\
E[\mathbf{h}_{\omega\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \mathcal{H}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)
\end{aligned}
$$

and

$$
\begin{aligned}
E[h_{\pi\gamma t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1/2}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\\
E[h_{\omega\gamma t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{4}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]\cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0]\\
E[h_{\gamma\gamma t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^2[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]\cdot E[\epsilon_{t-1}^4(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0]\\
E[\mathbf{h}_{\gamma\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty})|\boldsymbol{\varphi}_0]\\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^*|I_{t-1};\boldsymbol{\varphi}_0]\\
\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\epsilon_t(\boldsymbol{\theta}_s)|I_{t-1};\boldsymbol{\varphi}_0]\\
\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}'|I_{t-1};\boldsymbol{\varphi}_0]\\
\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\epsilon_t^2(\boldsymbol{\theta}_s)|I_{t-1};\boldsymbol{\varphi}_0]\\
\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}'\cdot\epsilon_t(\boldsymbol{\theta}_s)|I_{t-1};\boldsymbol{\varphi}_0]
\end{aligned}
$$

and $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}_{s0}', 0, \boldsymbol{\varrho}_0')'$. Finally,

$$
E[\epsilon_t^2(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0] = E[\omega^{-1}(y_t-\pi)^2|\boldsymbol{\varphi}_0] = E[\omega^{-1}(\pi_0+\omega_0^{1/2}\varepsilon_t^*-\pi)^2|\boldsymbol{\varphi}_0] = \omega^{-1}[(\pi_0-\pi)^2+\omega_0]
$$

and

$$
\begin{aligned}
E\{[\epsilon_t^4(\boldsymbol{\theta}_s)|\boldsymbol{\varphi}_0\} &= E\{\omega^{-2}[(y_t-\pi)^4|\boldsymbol{\varphi}_0\} = \omega^{-2}E\{[(\pi_0-\pi)+\omega_0^{1/2}\varepsilon_t^*]^4|\boldsymbol{\varphi}_0\}\\
&= \omega^{-2}[(\pi_0-\pi)^4+6(\pi_0-\pi)^2\omega_0+4\omega_0^{3/2}(\pi_0-\pi)\varphi(\boldsymbol{\varrho}_0)+\omega_0^2\kappa(\boldsymbol{\varrho}_0)].
\end{aligned}
$$

where $\varphi(\boldsymbol{\varrho}_0) = E(\varepsilon_t^{*3}|\boldsymbol{\varrho}_0)$ and $\kappa(\boldsymbol{\varrho}_0) = E(\varepsilon_t^{*4}|\boldsymbol{\varrho}_0)$ are the skewness and kurtosis coefficients of the true distribution of ε_t^*.

Proposition 8

Given that

$$
\mathbf{W}_d'(\pi_0,0,\omega_0,\boldsymbol{\eta}_0) = \begin{pmatrix} 0 & \tfrac{1}{2}\omega_0^{-1} & 0 \end{pmatrix},
$$

it is easy to see that

$$\mathring{\mathcal{S}}(\phi_0) = \begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & 0 \\ 0 & \frac{1}{(\kappa-1)\omega^2} & 0 \\ 0 & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix}.$$

Since this matrix is block diagonal and the efficiency bound for γ coincides with the corresponding element of the information matrix under correct specification of the conditional distribution, the asymptotic variance of the SSP estimator of this parameter coincides with that of the infeasible ML estimator which uses knowledge of the shape parameters $\boldsymbol{\eta}_0$. As a result, the non-centrality parameters will also be the same.

Similarly, we can use the expression for (A2) to show that

$$\begin{aligned} \mathcal{S}(\phi_0) &= \begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) \\ 0 & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) & 0 \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} - \\ &\quad \begin{pmatrix} \omega_0^{-1/2} & 0 \\ 0 & \frac{1}{2}\omega_0^{-1} \\ 0 & 0 \end{pmatrix} \left\{ \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} \right. \\ &\quad \left. - \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & \varphi \\ \varphi & \kappa-1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right\} \begin{pmatrix} \omega_0^{-1/2} & 0 & 0 \\ 0 & \frac{1}{2}\omega_0^{-1} & 0 \end{pmatrix} \\ &= \begin{bmatrix} \omega^{-1} & \frac{1}{2}\omega^{-3/2}\varphi & 0 \\ \frac{1}{2}\omega^{-3/2}\varphi & \frac{1}{4}\omega^{-2}(\kappa-1) & 0 \\ 0 & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix}. \end{aligned}$$

Given that this matrix is block diagonal and the efficiency bound for γ coincides with the corresponding element of the information matrix under correct specification of the conditional distribution, the asymptotic variance of the SP estimator of this parameter coincides with that of the infeasible ML estimator which uses knowledge of the shape parameters $\boldsymbol{\eta}_0$. Consequently, the non-centrality parameters will also be the same. □

Lemma 2

The proof is trivial if we combine several results that appear in the proofs of Propositions 5, 6 and 8. □

Proposition 9

The proof of the three statements is trivial if we combine several results that appear in the proofs of Propositions 1 and 5, 2 and 6, and 4 and 8, respectively, with the fact that the corresponding efficiency bounds are block diagonal between $\boldsymbol{\theta}_s$, ρ and γ when the true distribution of ε_t^* is symmetric.

References

Abramowitz, M. and Stegun, I.A. (1964): *Handbook of mathematic functions*, AMS 55, National Bureau of Standards.

Amengual, D. and Sentana, E. (2010): "A comparison of mean-variance efficiency tests", *Journal of Econometrics* 154, 16-34.

Amengual, D. and Sentana, E. (2011): "Inference in multivariate dynamic models with elliptical innovations", mimeo CEMFI.

Amengual, D. and Sentana, E. (2018): "Is a normal copula the right copula?", forthcoming in the *Journal of Business and Economic Statistics*.

Andrews, D.W.K. (2001): "Testing when a parameter is on the boundary of the maintained hypothesis", *Econometrica* 69, 683-734.

Arellano, M. (1991): "Moment testing with non-ML estimators", mimeo, CEMFI.

Bera, A. and Ng, P. (2002): "Robust tests for heteroskedasticity and autocorrelation using score function", *Journal of the Indian Society of Probability and Statistics* 6, 78-96.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* 31, 307-327.

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.

Bontemps, C. and Meddahi, N. (2012): "Testing distributional assumptions: a GMM approach", *Journal of Applied Econometrics* 27, 978-1012.

Breusch, T. S. and Pagan, A.R. (1980): "The Lagrange multiplier test and its applications to model specification in econometrics", *Review of Economic Studies* 47, 239-253.

Campbell, J.Y. and Yogo, M. (2006): "Efficient tests of stock return predictability", *Journal of Financial Economics* 81, 27-60.

Camponovo, L., Scaillet, O. and Trojani, F. (2012): "Predictive regression and robust hypothesis testing: predictability hidden by anomalous observations", Swiss Finance Institute DP 2013.05.

Christoffersen, P.F. and Diebold, F.X. (2006): "Financial asset returns, direction-of-change forecasting, and volatility dynamics", *Management Science* 52, 1273-1287.

Cochrane, J.H. (1991): "Volatility tests and efficient markets: a review essay", *Journal of Monetary Economics* 27, 661-676.

Creal, D.D., Koopman, S.J. and Lucas, A. (2013): "Generalized autoregressive score models with applications", *Journal of Applied Econometrics* 28, 777-795.

Crowder, M.J. (1976): "Maximum likelihood estimation for dependent observations", *Jour-*

nal of the Royal Statistical Society 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.

Davidson R., and MacKinnon, J.G. (1988): "Double-length artificial regressions", *Oxford Bulletin of Economics and Statistics* 50, 203–217.

Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika* 64, 247-254.

Davies, R. B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika* 74, 33-43.

Demos, A. and Sentana, E. (1998): "Testing for GARCH effects: a one-sided approach", *Journal of Econometrics*, 86, 97-127.

Dempster, A., Laird, N., and Rubin, D. (1977)**:** "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B* 39, 1-38.

Engle, R.F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* 50, 987-1007.

Fama, E.F., and French, K.R. (1993): "Common risk factors in the returns on stock and bonds", *Journal of Financial Economics* 33, 3-56.

Fama, E.F. and French, K.R. (2012) "Size, value, and momentum in international stock returns", *Journal of Financial Economics* 105, 457–472.

Fama, E.F., and French, K.R. (2015): "A five-factor asset pricing model", *Journal of Financial Economics* 116, 1-22.

Fiorentini, G. and Sentana, E. (1998): "Conditional means of time series processes and time series processes for conditional means", *International Economic Review* 39, 1101-1118.

Fiorentini, G. and Sentana, E. (2007): "On the efficiency and consistency of likelihood estimation in multivariate conditionally heteroskedastic dynamic regression models", CEMFI Working Paper 0713.

Fiorentini, G. and Sentana, E. (2015): "Tests for serial dependence in static, non-Gaussian factor models", in S.J. Koopman and N. Shephard (eds.) *Unobserved components and time series econometrics* 118-189, Oxford University Press.

Fiorentini, G. and Sentana, E. (2018): "Consistent non-Gaussian pseudo-log likelihood estimators", CEMFI Working Paper 1802.

Fiorentini, G., Sentana, E. and Calzolari, G. (2003): "Maximum likelihood estimation and inference in multivariate conditionally heteroskedastic dynamic regression models with Student t innovations", *Journal of Business and Economic Statistics* 21, 532-546.

Francq, C. and Zakoïan, J-M. (2010): *GARCH models: structure, statistical inference and financial applications*, Wiley.

Giacomini, R., Politis, D.N. and White, H. (2013): "A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators", *Econometric Theory* 29, 567-589.

Glejser, H. (1969): "A new test for heteroskedasticity", *Journal of the American Statistical Association* 64, 316-323.

Godfrey, L.G. (1988): *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*, Econometric Society Monographs.

González-Rivera, G. and Drost, F.C. (1999): "Efficiency comparisons of maximum-likelihood-based estimators in GARCH models", *Journal of Econometrics* 93, 93-111.

González-Rivera, G. and Ullah, A. (2001): "Rao's score test with nonparametric density estimators", *Journal of Statistical Planning and Inference* 97, 85-100.

Gouriéroux 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.

Hafner, C.M. and Rombouts, J.V.K. (2007): "Semiparametric multivariate volatility models", *Econometric Theory* 23, 251-280.

Harvey, A.C. (2013): *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Econometric Society Monographs, Cambridge University Press.

Harvey, A. C. and Chakravarty, T. (2008): "Beta-t-(e)-GARCH", Cambridge Working Papers in Economics 08340.

Henriksson, R.D. and Merton, R.C. (1981): "On market timing and investment performance II: statistical procedures for evaluating forecasting skills", *Journal of Business* 54, 513-533.

Hodgson, D.J. and Vorkink, K.P. (2003): "Efficient estimation of conditional asset pricing models", *Journal of Business and Economic Statistics* 21, 269-283.

Hodrick, R. J. (1992): "Dividend yields and stock returns: alternative procedures for inference and measurement", *Review of Financial Studies* 5, 357-386.

Hong, Y. (1996): "Consistent testing for serial correlation of unknown form", *Econometrica* 64, 837-864.

Hong Y. and R.S. Shehadeh (1999): "A new test for ARCH effects and its finite sample performance", *Journal of Business and Economic Statistics* 17, 91–108.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economic Letters*, 6, 255-259.

Jegadeesh, N. (1989): "On testing for slow decaying components in stock prices", mimeo, Anderson Graduate School of Management, University of California at Los Angeles.

Jondeau, E. and Rockinger, M. (2001): "Conditional volatility, skewness and kurtosis: Existence, persistence and comovements", *Journal of Economics Dynamics and Control* 27, 1699-1737.

Koenker, R. (1981): "A note on studentizing a test for heteroskedasticity", *Journal of Econometrics* 17, 107-112.

Kotz, S. (1975). "Multivariate distributions at a cross-road", in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical distributions in scientific work*, vol. I, 247-270, Reidel.

León, A., Mencía, J. and Sentana, E. (2009): "Parametric properties of seminonparametric distributions, with applications to option valuation", *Journal of Business and Economic Statistics* 27, 176-192.

Linton, O. (1993): "Adaptive estimation in ARCH models", *Econometric Theory* 9, 539-569.

Linton, O. and Steigerwald, D. (2000): "Adaptive testing in ARCH models", *Econometric Reviews* 19, 145-174.

Machado, J.A.F. and J. M. C. Santos Silva (2000): "Glejser's test revisited", *Journal of Econometrics* 97, 189-202.

Mardia, K. (1970). "Measures of multivariate skewness and kurtosis with applications", *Biometrika*, 57, 519–530.

Maronna, R., Martin, D. and Yohai, V. (2006): *Robust statistics - theory and methods*, Wiley.

Mencía, J. and E. Sentana (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, 105-121.

Mencía, J. and E. Sentana (2012): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", *Review of Economics and Statistics* 94, 133-152.

Mencía, J. and E. Sentana (2013): "Valuation of VIX derivatives", *Journal of Financial Economics* 108, 367-391.

Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H. (2012): "Time series momentum", *Journal of Financial Economics* 104, 228–250.

Numerical Algorithm Group (2001): *NAG Fortran 77 Library Mark 19 Reference Manual.*

Newey, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047-70.

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood

estimators in conditional heteroskedasticity models", *Econometrica* 65, 587-99.

Prakasa Rao, B. L. S. (1983): *Non-parametric functional estimation*, Academic Press.

RiskMetrics Group (1996): *RiskMetrics technical document.*

Robinson, P.M. (2010): "The efficient estimation of the semiparametric spatial autoregressive model", *Journal of Econometrics* 157, 6-17.

Ruud, P.A. (2000): *An introduction to classical econometric theory*, Oxford University Press.

Silverman B.W. (1986): *Density estimation*, Chapman and Hall.

Spiegel, M. (2008): "Forecasting the equity premium: where we stand today", *Review of Financial Studies* 24, 1453-1454.

Stuart, A. and K. Ord (1977): *Kendall's advanced theory of statistics* (6th ed.), Volume 1, Griffin, London.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics* 30, 415-443.

White, H. (1982). "Maximum likelihood estimation of missspecified models", *Econometrica* 50, 1-26.

B Local power calculations

B.1 General results

Let $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the h influence functions used to develop the following moment test of $H_0 : \boldsymbol{\theta}_2 = \mathbf{0}$:

$$M_T = T\bar{\mathbf{m}}_T'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}), \tag{B7}$$

where $\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0})$ is the sample average of $\mathbf{m}_t(\boldsymbol{\theta})$ evaluated under the null, $\boldsymbol{\Psi}$ is the corresponding asymptotic covariance matrix and $\boldsymbol{\theta}_{10}$ the true values of the remaining model parameters. In order to obtain the non-centrality parameter of this test under Pitman sequences of local alternatives of the form $H_0 : \boldsymbol{\theta}_{2T} = \bar{\boldsymbol{\theta}}_2/\sqrt{T}$, it is convenient to linearise $\mathbf{m}_t(\boldsymbol{\theta}_{10}, \mathbf{0})$ with respect to $\boldsymbol{\theta}_2$ around its true value $\boldsymbol{\theta}_{2T}$. This linearisation yields

$$\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}) = \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{2T}) + \frac{1}{T}\sum\nolimits_{t=1}^{T} \frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{2T}^*)}{\partial \boldsymbol{\theta}_2'}\bar{\boldsymbol{\theta}}_2,$$

where $\boldsymbol{\theta}_{2T}^*$ is some "intermediate" value between $\boldsymbol{\theta}_{2T}$ and $\mathbf{0}$. As a result,

$$\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}) \to N[\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2, \boldsymbol{\Psi}],$$

under standard regularity conditions, where

$$\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0}) = E[\partial \mathbf{m}_t(\boldsymbol{\theta}_{10}, \mathbf{0})/\partial \boldsymbol{\theta}_2'],$$

so that the non-centrality parameter of the moment test (B7) will be

$$\bar{\boldsymbol{\theta}}_2'\mathbf{M}'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2 \tag{B8}$$

when $\boldsymbol{\theta}_{10}$ is known. On this basis, we can easily obtain the limiting probability of M_T exceeding some pre-specified quantile of a central χ_h^2 distribution from the cdf of a non-central χ^2 distribution with h degrees of freedom and non-centrality parameter (B8).

Often, though, $\boldsymbol{\theta}_{10}$ will be unknown, and we will have to replace it by some estimator $\bar{\boldsymbol{\theta}}_{1T}$. Let $\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the $\dim(\boldsymbol{\theta}_1)$ influence functions used to estimate $\boldsymbol{\theta}_{10}$ subject to the restriction $\boldsymbol{\theta}_2 = \mathbf{0}$. For convenience, we replace the original influence functions by

$$\mathbf{m}_t^{\perp}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) - E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1'}\right)\left[E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1'}\right)\right]^{-1}\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

which are unaffected by the sampling uncertainty in the estimator of $\boldsymbol{\theta}_1$. In this way, the test statistic will be

$$M_T = T\bar{\mathbf{m}}_T^{\perp\prime}(\bar{\boldsymbol{\theta}}_{1T}, \mathbf{0})\boldsymbol{\Upsilon}^{-1}\bar{\mathbf{m}}_T^{\perp}(\bar{\boldsymbol{\theta}}_{1T}, \mathbf{0}),$$

where $\boldsymbol{\Upsilon}$ is the relevant asymptotic covariance matrix, which takes into account the possible (long-run) correlation between $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. As a result, the non-centrality parameter will be

$$\bar{\boldsymbol{\theta}}_2' \mathbf{M}^{\perp\prime}(\boldsymbol{\theta}_{10}, \mathbf{0}) \boldsymbol{\Upsilon}^{-1} \mathbf{M}^{\perp}(\boldsymbol{\theta}_{10}, \mathbf{0}) \bar{\boldsymbol{\theta}}_2,$$

where

$$\mathbf{M}^{\perp}(\boldsymbol{\theta}_{10}, \mathbf{0}) = E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2'}\right) - E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1'}\right) \left[E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1'}\right)\right]^{-1} E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2'}\right).$$

In the special case in which $\bar{\boldsymbol{\theta}}_{1T}$ is the ML estimator of $\boldsymbol{\theta}_{10}$ under the null, and $\mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0})$ and the scores corresponding to $\boldsymbol{\theta}_1$ are asymptotically uncorrelated when H_0 is true, as in all our tests under correct specification, then no adjustment will be required because $E[\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_1']$ will be 0 by the generalised information matrix equality. In addition, both $\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})$ and $\boldsymbol{\Psi}$ coincide with the (2,2) block of the information matrix when $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are the scores with respect to $\boldsymbol{\theta}_2$.

If on the other hand $\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ coincide with the scores with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ but these are not uncorrelated under the null, as in our tests under incorrect specification, then we should we work with $\mathbf{m}_t^{\perp}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, although we could still exploit the fact that

$$E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1'}\right)' = E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2'}\right)$$

by the symmetry of the Hessian matrix. In either case, though, the non-centrality parameters of LM and Wald tests will be the same under sequences of local alternatives, at least under the assumption that $\boldsymbol{\theta}_2$ is consistently estimated not only under the null but also under those sequences (see White (1982)).

B.2 Gaussian tests

B.2.1 Serial correlation tests

Let us assume without loss of generality that $\pi = 0$. The first-order serial correlation test is effectively based on the influence functions

$$m_{lt}(\boldsymbol{\theta}_s, \rho) = y_t y_{t-1} - G_{yy}(1)$$

evaluated at $\rho = 0$. But since

$$y_t = \left(1 + \sum\nolimits_{l=1}^{h} \rho L^l\right) \varepsilon_t,$$

we will have that

$$G_{yy}(0) = [1 + (h-1)\rho^2]\sigma^2$$

The Yule-Walker equations of the model considered in (11) will be given by

$$\begin{array}{rcl}
\frac{G_{yy}(1)}{G_{yy}(0)} = & \rho\left[1+\frac{G_{yy}(1)}{G_{yy}(0)}+\ldots+\frac{G_{yy}(h-1)}{G_{yy}(0)}\right] \\
\frac{G_{yy}(2)}{G_{yy}(0)} = & \rho\left[\frac{G_{yy}(1)}{G_{yy}(0)}+1+\ldots+\frac{G_{yy}(h-2)}{G_{yy}(0)}\right] \\
\vdots & \vdots \\
\frac{G_{yy}(h-1)}{G_{yy}(0)} = & \rho\left[\frac{G_{yy}(h-2)}{G_{yy}(0)}+\frac{G_{yy}(h-3)}{G_{yy}(0)}+\ldots+\frac{G_{yy}(1)}{G_{yy}(0)}\right]
\end{array}$$

whence

$$G_{yy}(1) = \frac{\rho}{1-(h-1)\rho}[1+(h-1)\rho^2]\sigma^2.$$

Hence, it trivially follows that

$$M_l(\boldsymbol{\theta}_s, \mathbf{0}) = E[\partial m_{lt}(\boldsymbol{\theta}_s, 0)/\partial\rho] = -\sigma^2.$$

As for the asymptotic covariance matrix, the proof of Proposition 2 implies that if $\rho = 0$, then

$$\sqrt{T}m_{lt}(\boldsymbol{\theta}_s, 0) = \frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} y_t y'_{t-1} \to N(0, \sigma^4)$$

irrespective of the distribution of y_t. As a result, the non-centrality parameter will be ρ^2 regardless of h.

In contrast, the test that uses the influence function

$$y_t \sum\nolimits_{l=1}^{h} y_{t-l} - \sum\nolimits_{l=1}^{h} G_{yy}(l)$$

will be asymptotically equivalent to the Wald test based on the Gaussian PML estimator ρ, whose non-centrality parameter is $h\rho^2$, which is clearly bigger than ρ^2 for any $h > 1$.

It is also interesting to study the opposite situation in which we decide to use the influence function that involves $h-$period returns when in fact the true model is an AR(1). Since $G_{yy}(l) = \rho^l\sigma^2$ in that case, $\sum_{l=1}^{h} G_{yy}(l)$ will be equal to $(1-\rho^{h+1})\sigma^2/(1-\rho)$. Therefore, $M_l(\boldsymbol{\theta}_s, \mathbf{0})$ will also be equal to $-\sigma^2$. But since the asymptotic covariance of the sample average of $y_t \sum_{l=1}^{h} y_{t-l}$ is $h\sigma^4$ under the null, the non-centrality parameter will be $h^{-1}\rho^2$, which is clearly below ρ^2 for any $h > 1$.

B.2.2 GARCH tests

To keep the algebra simple, we assume once again that $\pi = 0$, that the conditional variance has been generated according to a GARCH(1,1) process and that the conditional distribution has constant kurtosis coefficient κ. The fixed-$\bar{\beta}$ GARCH test is based on the following influence function:

$$m_{st}(\sigma^2, \bar{\beta}) = (x_t^2 - \sigma^2)\sum\nolimits_{j=0}^{\infty} \bar{\beta}^j (x_{t-j}^2 - \sigma^2)$$

As is well known, Bollerslev (1986) showed that a GARCH$(1,1)$ model implies the following ARMA$(1,1)$ process for x_t^2:

$$(x_t^2 - \sigma^2) = (\alpha + \beta)(x_{t-1}^2 - \sigma^2) + \eta_t - \beta\eta_{t-1},$$

where η_t is the martingale difference sequence $x_t^2 - \sigma_t^2$. As a result,

$$\begin{aligned} V(x_t^2) &= \frac{1 - 2\alpha\beta - \beta^2}{1 - (\alpha + \beta)^2} V(\eta_t), \\ cov(x_t^2, x_{t-1}^2) &= \frac{[1 - (\alpha + \beta)\beta]}{1 - (\alpha + \beta)^2} \alpha V(\eta_t), \end{aligned}$$

and

$$cov(x_t^2, x_{t-j-1}^2) = (\alpha + \beta) cov(x_t^2, x_{t-j}^2) = (\alpha + \beta)^{j-1} cov(x_t^2, x_{t-1}^2)$$

for any $j \geq 1$, so that

$$\begin{aligned} cor(x_t^2, x_{t-1}^2) &= \frac{[1 - (\alpha + \beta)\beta]}{1 - 2\alpha\beta - \beta^2} \alpha, \\ cor(x_t^2, x_{t-j-1}^2) &= (\alpha + \beta)^{j-1} cor(x_t^2, x_{t-1}^2). \end{aligned}$$

But since we know that

$$V(x_t^2) = \frac{1 - 2\alpha\beta - \beta^2}{1 - \kappa\alpha^2 - 2\alpha\beta - \beta^2} (\kappa - 1)\sigma^4$$

when $\kappa\alpha^2 + 2\alpha\beta + \beta^2 < 1$, it immediately follows that

$$V(\eta_t) = \frac{1 - (\alpha + \beta)^2}{1 - \kappa\alpha^2 - 2\alpha\beta - \beta^2} (\kappa - 1)\sigma^4.$$

As a result, the expected value of $m_{st}(\sigma^2, \bar{\beta})$ under the alternative will be given by

$$\sum_{j=0}^{\infty} \bar{\beta}^j (\alpha + \beta)^j E[(x_t^2 - \sigma^2)(x_{t-1}^2 - \sigma^2)] = \frac{\alpha}{1 - \bar{\beta}(\alpha + \beta)} \frac{[1 - (\alpha + \beta)\beta]}{1 - \kappa\alpha^2 - 2\alpha\beta - \beta^2} (\kappa - 1)\sigma^4.$$

If we expand this expression with respect to α at $\alpha = 0$, we finally obtain

$$\frac{\alpha}{1 - \bar{\beta}\beta} (\kappa - 1)\sigma^4.$$

Hence, the non-centrality parameter will be

$$\frac{1 - \bar{\beta}^2}{(1 - \bar{\beta}\beta)^2} \alpha^2.$$

Specifically, for $\bar{\beta} = 0$ the non-centrality parameter will be α^2, while for $\bar{\beta} = 1$ the non-centrality parameter becomes 0 because the regressor has infinite variance while the regressand does not. In fact, $\bar{\beta}$ bigger than $2\beta/(1 + \beta^2)$ will result in local power losses relative to $\bar{\beta} = 0$. Not surprisingly, the maximum of this expression is achieved for $\bar{\beta} = \beta$, in which case its value is

$$\frac{\alpha^2}{1 - \beta^2},$$

which is bigger than α^2, the more so the closer β is to 1.

B.3 Student t tests

Under correct specification, the non-centrality parameters are trivial to find because they effectively depend on the $\rho\rho$ or $\alpha\alpha$ elements of the information matrix under the null of mean and variance unpredictability, which we have already discussed in Lemmas 1 and 2. Under distributional misspecification, the calculations are substantially more elaborate.

B.3.1 Normal mixtures

For any given value of the mixing probability λ, the ratio of variances υ and the relative differences in means δ, the first thing we do is to compute the pseudo true values of the Student t pseudo ML estimators under the null, namely π_∞, ω_∞ and η_∞. We obtain these pseudo true values by solving a nonlinear system of three equations that sets to zero the expected value of the scores with respect to π, θ and η. We compute the integrals with respect to the true normal mixture measure as the weighted average of two integrals with respect to the two underlying Gaussian measures, as in Amengual and Sentana (2010). We obtain each of those integrals by Gauss-Hermite quadrature with infinite support using the NAG D01BAF routine with 64 points, $a = \mu_i$ and $b = .5\sigma_i^{-2}$ $(i = 1, 2)$. We solve the resulting nonlinear system of equations in two steps. First, we define a non-uniform grid of 70 values for η between 0.001 and .4995, which is finer close to the two extremes, and then solve the bivariate system for π and ω keeping η fixed. Next, we feed the "best" triplet as starting values for solving the trivariate system using the NAG C05NCF routine.

Once we have thus obtained π_∞, ω_∞ and η_∞, we compute the expected value of the Hessian ($\mathcal{H}$) and variance of the score ($\mathcal{K}$), including the elements involving ρ or γ using the expressions in the proofs of Propositions 3 and 7. We then compute the usual sandwich formulas $\mathcal{H}^{-1}B\mathcal{H}^{-1}$ and take the appropriate diagonal element to obtain the ratio of noncentrality parameters of the Student $t-$based test to the Gaussian one. Although we can repeat these calculations for any possible triplet $(\lambda, \upsilon, \delta)$, in practice we fix $\lambda = .05$ and define a bivariate grid (on a log-scale) on δ and υ of 300×80 points. We then find out the skewness and kurtosis values that those parameters imply using the bounds described in appendix C.1.2.

There are two further controls in the program. On the one hand, when η_∞ is less or equal then 0.001, then we simply set the ratios of noncentrality parameters equal to one. On the other hand, when η_∞ is greater or equal than .4995, then we drop η from the calculations and compute the expected Hessian and variance of the score matrices for the remaining three parameters.

B.3.2 Gram-Charlier expansions

The procedure for the fourth-order Gram-Charlier density is similar to the one we have just described for discrete normal mixtures. The most relevant differences are (i) that the shape parameters of the true measure are now c_3 and c_4, so that we need to find out first the admissible range of values of these parameters which are compatible with a non-negative density; and (ii) the values of a and b in the Gauss-Hermite numerical quadrature NAG D01BAF routine are no longer optimal.

C Standardised random variables

C.1 Discrete location scale mixtures of normals

C.1.1 Definition and simulation

Let s_t denote an *i.i.d.* Bernoulli variate with $P(s_t = 1) = \lambda$. If $z_t|s_t$ is *i.i.d.* $N(0,1)$, then

$$\varepsilon_t^* = \frac{1}{\sqrt{1+\lambda(1-\lambda)\delta^2}}\left[\delta(s_t-\lambda) + \frac{s_t + (1-s_t)\sqrt{\upsilon}}{\sqrt{\lambda+(1-\lambda)\upsilon}} z_t\right],$$

where $\delta \in \mathbb{R}$ and $\upsilon > 0$, is a two component mixture of normals whose first two unconditional moments are 0 and 1, respectively. The intuition is as follows. First, note that $\delta(s_t - \lambda)$ is a shifted and scaled Bernoulli random variable with 0 mean and variance $\lambda(1-\lambda)\delta^2$. But since

$$\frac{s_t + (1-s_t)\sqrt{\upsilon}}{\sqrt{\lambda+(1-\lambda)\upsilon}} z_t$$

is a discrete scale mixture of normals with 0 unconditional mean and unit unconditional variance that is orthogonal to $\delta(s_t - \lambda)$, the sum of the two random variables will have variance $1 + \lambda(1-\lambda)\delta^2$, which explains the scaling factor.

An equivalent way to define and simulate the same standardised random variable is as follows

$$\varepsilon_t^* = \begin{cases} N[\mu_1^*(\boldsymbol{\eta}), \sigma_1^{*2}(\boldsymbol{\eta})] \text{ with probability } \lambda \\ N[\mu_2^*(\boldsymbol{\eta}), \sigma_2^{*2}(\boldsymbol{\eta})] \text{ with probability } 1-\lambda \end{cases} \tag{C9}$$

where $\boldsymbol{\eta} = (\delta, \upsilon, \lambda)'$ and

$$\begin{aligned}
\mu_1^*(\boldsymbol{\eta}) &= \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \\
\mu_2^*(\boldsymbol{\eta}) &= -\frac{\delta\lambda}{\sqrt{1+\lambda(1-\lambda)\delta^2}} = -\frac{\lambda}{1-\lambda}\mu_1^*(\boldsymbol{\eta}), \\
\sigma_1^{*2}(\boldsymbol{\eta}) &= \frac{1}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\upsilon]}, \\
\sigma_2^{*2}(\boldsymbol{\eta}) &= \frac{\upsilon}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\upsilon]} = \upsilon\sigma_1^{*2}(\boldsymbol{\eta}).
\end{aligned}$$

Therefore, we can immediately interpret υ as the ratio of the two variances. Similarly, since

$$\delta = \frac{\mu_1^*(\boldsymbol{\eta}) - \mu_2^*(\boldsymbol{\eta})}{\sqrt{\lambda\sigma_1^{*2}(\boldsymbol{\eta}) + (1-\lambda)\sigma_1^{*2}(\boldsymbol{\eta})}},$$

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components relative to the mean of the two conditional variances.

We can trivially extended this procedure to define and simulate standardised mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (C9) by a k-component normal mixture with mean and variance given by $\mu_1^*(\boldsymbol{\eta})$ and $\sigma_1^{*2}(\boldsymbol{\eta})$, respectively, then the resulting random variable will be a $(k+1)$-component Gaussian mixture with zero mean and unit variance.

Finally, note that we can also use the above expressions to generate a two component mixture of normals with mean π and variance ω^2 as

$$y_t = \begin{cases} N(\mu_1, \sigma_1^2) \text{ with probability } \lambda \\ N(\mu_2, \sigma_2^2) \text{ with probability } 1-\lambda \end{cases}$$

with

$$\begin{aligned} \mu_1 &= \pi + \omega\mu_1^*(\boldsymbol{\eta}) \\ \mu_2 &= \pi + \omega\mu_2^*(\boldsymbol{\eta}) \\ \sigma_1^2 &= \omega\sigma_1^{*2}(\boldsymbol{\eta}), \\ \sigma_2^2 &= \omega\sigma_2^{*2}(\boldsymbol{\eta}). \end{aligned}$$

Interestingly, the expressions for υ and δ above continue to be valid if we replace $\mu_1^*(\boldsymbol{\eta})$, $\mu_2^*(\boldsymbol{\eta})$, $\sigma_1^{*2}(\boldsymbol{\eta})$ and $\sigma_2^{*2}(\boldsymbol{\eta})$ by μ_1, μ_2, σ_1^2 and σ_2^2.

C.1.2 Skewness-kurtosis bounds

In the case of two-component Gaussian mixtures, the parameters λ, δ and υ determine the higher order moments of ε_t^* through the relationship

$$E(\varepsilon_t^{*j}) = \lambda E(\varepsilon_t^{*j}|s_t = 1) + (1-\lambda)E(\varepsilon_t^{*j}|s_t = 0),$$

where $E(\varepsilon_t^{*j}|s_t = 1)$ can be obtained from the usual normal expressions

$$\begin{aligned} E(\varepsilon_t^*|s_t = 1) &= \mu_1^*(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*2}|s_t = 1) &= \mu_1^{*2}(\boldsymbol{\eta}) + \sigma_1^{*2}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*3}|s_t = 1) &= \mu_1^{*3}(\boldsymbol{\eta}) + 3\mu_1^*(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*4}|s_t = 1) &= \mu_1^{*4}(\boldsymbol{\eta}) + 6\mu_1^{*2}(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) + 3\sigma_1^{*4}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*5}|s_t = 1) &= \mu_1^{*5}(\boldsymbol{\eta}) + 10\mu_1^{*3}(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) + 15\mu_1^*(\boldsymbol{\eta})\sigma_1^{*4}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*6}|s_t = 1) &= \mu_1^{*6}(\boldsymbol{\eta}) + 15\mu_1^{*4}(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) + 45\mu_1^{*2}(\boldsymbol{\eta})\sigma_1^{*4}(\boldsymbol{\eta}) + 15\sigma_1^{*6}(\boldsymbol{\eta}) \end{aligned}$$

etc. But since $E(\varepsilon_t^*) = 0$ and $E(\varepsilon_t^{*2}) = 1$ by construction, straightforward algebra shows that the skewness and kurtosis coefficients will be given by

$$E(\varepsilon_t^{*3}) = \frac{3\delta\lambda(1-\lambda)(1-\upsilon)}{[\lambda+(1-\lambda)\upsilon][1+\lambda(1-\lambda)\delta^2]^{3/2}} + \frac{\delta^3(1-\lambda)\lambda(1-2\lambda)}{[1+\lambda(1-\lambda)\delta^2]^{3/2}} = a(\delta,\upsilon,\lambda) \tag{C10}$$

and

$$\begin{aligned} E(\varepsilon_t^{*4}) &= \frac{3[\lambda+(1-\lambda)\upsilon^2]}{[\lambda+(1-\lambda)\upsilon]^2[1+\lambda(1-\lambda)\delta^2]^2} + \frac{6\delta^2\lambda(1-\lambda)[(1-\lambda)+\upsilon\lambda]}{[\lambda+(1-\lambda)\upsilon][1+\lambda(1-\lambda)\delta^2]^2} \\ &+ \frac{\delta^4\lambda(1-\lambda)[1-3\lambda(1-\lambda)]}{[1+\lambda(1-\lambda)\delta^2]^2} = b(\delta,\upsilon,\lambda). \end{aligned} \tag{C11}$$

Two issues are worth pointing out. First, $a(\delta,\upsilon,\lambda)$ is an odd function of δ, which means that δ and $-\delta$ yield the same skewness in absolute value. In this sense, if we set $\delta = 0$ then we will obtain a discrete scale mixture of normals, which is always symmetric but leptokurtic. Another way of obtaining discrete normal mixture distributions that are symmetric is by making $\lambda = \frac{1}{2}$ and $\upsilon = 1$. Second, $b(\delta,\upsilon,\lambda)$ is an even function of δ, which implies that δ and $-\delta$ give rise to the same kurtosis. For that reason, in what follows we mostly consider the case of $\delta \geq 0$.

A useful property of two component normal mixtures is that they span the entire unconditional skewness-kurtosis frontier given by the parabola $E(\varepsilon_t^{*4}) \geq 1 + E^2(\varepsilon_t^{*3})$ (see Stuart and Ord (1977)). More specifically, for a fixed value of λ, skewness, which is 0 for $\delta = 0$, reaches its frontier value as $\delta \to \infty$, in which case

$$\lim_{\delta\to\infty} a(\delta,\upsilon,\lambda) = \frac{2(\frac{1}{2}-\lambda)}{\sqrt{\lambda(1-\lambda)}}$$

regardless of υ. Clearly, for $\lambda < .5$ this limiting skewness value is positive, while it is negative for $\lambda > .5$. In any case, we can achieve the mirror point on the frontier as $\delta \to -\infty$.

The corresponding kurtosis values are

$$b(0,\upsilon,\lambda) = \frac{3(\lambda+(1-\lambda)\upsilon^2)}{(\lambda+(1-\lambda)\upsilon)^2} = 3\left(\frac{\lambda(1-\lambda)(1-\upsilon)^2}{(\lambda+(1-\lambda)\upsilon)^2}+1\right)$$

and

$$\lim_{\delta\to\pm\infty} b(\delta,\upsilon,\lambda) = -3 + \frac{1}{\lambda(1-\lambda)} = 1 + \left(\frac{2(\frac{1}{2}-\lambda)}{\sqrt{\lambda(1-\lambda)}}\right)^2,$$

which again does not depend on υ. Intuitively, the reason is that a standardised two component normal mixture converges in distribution to a standardised Bernoulli random variable with parameter λ as $\delta \to \infty$ regardless of υ. Interestingly, $\lim_{\delta\to\infty} b(\delta,\upsilon,\lambda) = 3$ for $\lambda = \frac{1}{2} \pm \frac{1}{6}\sqrt{3}$.

Nevertheless, to create Figures 4B and 7B, we need to find out the range of skewness and kurtosis that this distribution can generate when λ is fixed. In this sense, notice that kurtosis is

always larger or equal than 3 for $\delta = 0$, which reflects the fact that a scale mixture of normals is always leptokurtic. The boundary case is of course $\upsilon = 1$, in which case

$$b(0, 1, \lambda) = 3.$$

In fact, maximum kurtosis when $\delta = 0$ is achieved for $\upsilon = 0$ or for $\upsilon \to \infty$, in which case we obtain either

$$b(0, 0, \lambda) = \frac{3}{\lambda} \text{ or } \lim_{\upsilon \to \infty} b(0, \upsilon, \lambda) = \frac{3}{1 - \lambda}.$$

Obviously, this kurtosis can be made arbitrarily large as λ approaches 0 or 1, but it is clearly bounded for fixed λ.

The other interesting cases arise when $\upsilon = 0$ and $\upsilon = 1$. In the first case

$$a(\delta, 0, \lambda) = \delta\,(1 - \lambda)\,\frac{3 + (1 - 2\lambda)\lambda\delta^2}{(1 + \lambda(1 - \lambda)\delta^2)^{3/2}}$$

and

$$b(\delta, 0, \lambda) = \frac{3}{\lambda(1 + \lambda(1 - \lambda)\delta^2)^2} + \frac{6\delta^2(1 - \lambda)^2}{(1 + \lambda(1 - \lambda)\delta^2)^2} + \frac{\delta^4\lambda(1 - \lambda)(1 - 3\lambda(1 - \lambda))}{(1 + \lambda(1 - \lambda)\delta^2))^2},$$

while in the second case

$$a(\delta, 1, \lambda) = \frac{\delta^3(1 - \lambda)\lambda(1 - 2\lambda)}{(1 + \lambda(1 - \lambda)\delta^2)^{3/2}}$$

and

$$b(\delta, 1, \lambda) = \frac{3}{(1 + \lambda(1 - \lambda)\delta^2)^2} + \frac{6\delta^2\lambda(1 - \lambda)}{(1 + \lambda(1 - \lambda)\delta^2)^2} + \frac{\delta^4\lambda(1 - \lambda)(1 - 3\lambda(1 - \lambda))}{(1 + \lambda(1 - \lambda)\delta^2))^2}.$$

It turns out that the range of skewness and kurtosis that a standardised mixture of two normals can generate seems to be bounded by the following two parametric curves:

$$(a(\delta, 1, \lambda), b(\delta, 1, \lambda))$$

and

$$(a(\delta, 0, \lambda), b(\delta, 0, \lambda)),$$

where the range of δ is $[0, \infty)$. In fact, these curves intersect at the unconditional skewness-frontier boundary when $\delta \to \infty$.

Interestingly, it seems that skewness is always non-negative when $\lambda \leq 1/2$. In contrast, for $\lambda > 1/2$ skewness is initially positive for small values of δ, but then becomes negative as δ increases. In turn, kurtosis bounded from below by 3 when $\lambda \leq \frac{1}{2} - \frac{1}{6}\sqrt{3}$, while it is bounded from above by 3 on the negative skewness side if $\frac{1}{2} \leq \lambda \leq \frac{1}{2} + \frac{1}{6}\sqrt{3}$.

As we explained before, the mirror curves

$$(-a(|\delta|, 1, \lambda), b(|\delta|, 1, \lambda))$$

and

$$(-a(|\delta|,0,\lambda), b(|\delta|,0,\lambda)),$$

give us the skewness-kurtosis range when δ if negative.

C.2 Gram-Charlier distributions

C.2.1 Definition and moments

The first five raw Hermite polynomials are:

$$\begin{aligned} H_0(z) &= 1, \\ H_1(z) &= z, \\ H_2(z) &= z^2 - 1, \\ H_3(z) &= z^3 - 3z, \\ H_4(z) &= z^4 - 6z^2 + 3. \end{aligned}$$

When $z \sim N(0,1)$, these have 0 mean and are orthogonal to each other. In turn,

$$\begin{aligned} H_2^*(z) &= \frac{z^2-1}{\sqrt{2}}, \\ H_3^*(z) &= \frac{z^3-3z}{\sqrt{6}}, \\ H_4^*(z) &= \frac{z^4-6z^2+3}{\sqrt{24}}. \end{aligned}$$

are called the standardised Hermite polynomials because their variance will be 1 for a standard normal.

The Gram-Charlier density is defined as:

$$f(z) = \phi(z)P(z), \tag{C12}$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2},$$

$$P(z) = 1 + \frac{\varphi}{\sqrt{6}} H_3^*(z) + \frac{\upsilon}{\sqrt{24}} H_4^*(z) = 1 + \frac{\varphi}{6}\left(z^3 - 3z\right) + \frac{\kappa}{24}\left(z^4 - 6z^2 + 3\right). \tag{C13}$$

This density is such that

$$\begin{aligned} E_f(z) &= 0, \\ E_f(z^2) &= 1, \\ E_f(z^3) &= \varphi, \\ E_f(z^4) &= 3 + \kappa. \end{aligned}$$

C.2.2 Positivity restrictions

The problem is that $P(z)$ in (C13) can be negative, in which case $f(z)$ in (C12) will not be a proper density.

For a given z, the skewness-excess kurtosis frontier that guarantees positivity must satisfy the following two equations:

$$\begin{aligned} 1 + \frac{\varphi}{6}\left(z^3 - 3z\right) + \frac{\kappa}{24}\left(z^4 - 6z^2 + 3\right) &= 0, \\ \frac{\varphi}{2}\left(z^2 - 1\right) + \frac{\kappa}{6}\left(z^3 - 3z\right) &= 0. \end{aligned}$$

The first equation, which is given by $P(z) = 0$, defines a straight line in (φ, κ) space such that in any neighbourhood of the solution we will find positive and negative densities. In contrast, the second equation, which corresponds to $\partial P(z)/\partial z = 0$, guarantees that we remain in the frontier as we move in (φ, κ) space.

The solution to the above system of equations in terms of φ and κ as a function of z is

$$\begin{aligned} \varphi(z) &= -24\frac{z^3 - 3z}{z^6 - 3z^4 + 9z^2 + 9}, \\ \kappa(z) &= 72\frac{z^2 - 1}{z^6 - 3z^4 + 9z^2 + 9}, \end{aligned}$$

where the denominator is

$$d(z) = 4\left(z^3 - 3z\right)^2 - 3\left(z^2 - 1\right)\left(z^4 - 6z^2 + 3\right) = z^6 - 3z^4 + 9z^2 + 9.$$

This solution can be regarded as the parametric representation of the admissible skewness-kurtosis frontier.

The simplest way to find the frontier values is to carry out a grid over z, and for each value of z find out the corresponding values of $\varphi(z)$ and $\kappa(z)$. However, this does not work as expected because we will often end up with two different values of $\varphi(z)$ for the same value of $\kappa(z)$. Following Jondeau and Rockinger (2001), the solution is to restrict the range of z to be $[\sqrt{3}, \infty)$. When $z = \sqrt{3}$, $\varphi(z)$ and $\kappa(z)$ become 0 and 4, respectively. In contrast, when $z \to \infty$ both $\varphi(z)$ and $\kappa(z)$ converge to 0. In practice, the grid should probably be logarithmic between $\sqrt{3}$ and 10^3 or so. The maximum skewness that can be achieved is 1.0493. Obviously, we get the mirror image by changing the sign of z.

C.2.3 Simulation

A very simple way to simulate random variables with a Gram-Charlier distribution is by using the usual inversion method, which exploits the fact that if Z is a random variable with

absolutely continuous distribution function $F_Z(.)$ and quantile function $F_Z^{-1}(.)$, then $U = F_Z(Z)$ is uniformly distributed between 0 and 1, while $F_Z^{-1}(U)$ will follow the distribution of Z.

Given that

$$\int H_k^*(x)\,\phi(x)\,dx = \frac{-1}{\sqrt{k}} H_{k-1}^*(x)\,\phi(x) \qquad k \geq 1 \tag{C14}$$

(see León, Mencía and Sentana (2009)), and that $H_k^*(x)\,\phi(x) \to 0$ when $x \to -\infty$ by virtue of L'Hôpital rule, then

$$\int_{-\infty}^{z} H_k^*(x)\,\phi(x)\,dz = -\frac{1}{\sqrt{k}} H_{k-1}^*(z)\,\phi(z), \qquad k \geq 1. \tag{C15}$$

Consequently,

$$F_Z(z) = \int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \phi(x)P(x)dx = \Phi(z) - \frac{\varphi}{6} H_2^*(z)\,\phi(z) - \frac{\kappa}{24} H_3^*(z)\,\phi(z).$$

In practice, we simulate a uniform variate u, and numerically solve the equation

$$F_Z(z) = u$$

with $\Phi^{-1}(u)$ as starting value.

C.3 Generalised hyperbolic

Let ξ_t denote an *i.i.d.* Generalised Inverse Gaussian (GIG) random variable with parameters $-\nu, \tau$ and 1, or $GIG(-\nu, \tau, 1)$ for short. Mencía and Sentana (2012) show that if $z_t|\xi_t$ is *i.i.d.* $N(0,1)$, then

$$\varepsilon_t^* = c(\beta,\nu,\tau)\beta\left[\frac{\tau\xi_t^{-1}}{R_\nu(\tau)} - 1\right] + \sqrt{\frac{\tau\xi_t^{-1}}{R_\nu(\tau)}}\sqrt{c(\beta,\nu,\tau)}z_t$$

is a standardised Generalised Hyperbolic (GH) distribution with parameters β, ν and τ, where

$$\begin{aligned} c(\beta,\nu,\tau) &= \frac{-1 + \sqrt{1 + 4\beta^2[D_{\nu+1}(\tau) - 1]}}{2\beta^2[D_{\nu+1}(\tau) - 1]} \\ R_\nu(\tau) &= \frac{K_{\nu+1}(\tau)}{K_\nu(\tau)}, \\ D_{\nu+1}(\tau) &= \frac{K_{\nu+2}(\tau)K_\nu(\tau)}{K_{\nu+1}(\tau)}, \end{aligned}$$

and $K_\nu(.)$ is the modified Bessel function of the third kind. In turn, the GH distribution is a special case of the more general location scale mixtures of normals considered in Mencía and Sentana (2009), in which ξ_t is a positive random variable with an arbitrary distribution.

Mencía and Sentana (2012) also provide expressions for the third and fourth moments of the GH distribution, which in the univariate case reduce to

$$E(\varepsilon_t^{*3}) = c^3(\beta,\nu,\tau)\left[\frac{K_{\nu+3}(\tau)\,K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} - 3D_{\nu+1}(\tau) + 2\right]\beta^3 + 3c^2(\beta,\nu,\tau)\left[D_{\nu+1}(\tau) - 1\right]\beta$$

and

$$E(\varepsilon_t^{*4}) = c^4(\beta,\nu,\tau)\left[\frac{K_{\nu+4}(\tau)K_\nu^3(\tau)}{K_{\nu+1}^4(\tau)} - 4\frac{K_{\nu+3}(\tau)K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} + 6D_{\nu+1}(\tau) - 3\right]\beta^4$$
$$+6c^3(\beta,\nu,\tau)\left[\frac{K_{\nu+3}(\tau)K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} - 2D_{\nu+1}(\tau) + 1\right]\beta^2 + 3D_{\nu+1}(\tau)\,c^2(\beta,\nu,\tau).$$

C.3.1 Asymmetric and symmetric versions of the Student t

The asymmetric t distribution is nested within the GH family when $\tau = 0$ and $-\infty < \nu < -2$. If we define $\eta = -1/(2\nu)$, then for $\eta < 1/4$ we will have that

$$\begin{aligned} c(\beta,\nu,\tau) &= \frac{1-4\eta}{2\eta}\frac{\sqrt{1+8\beta^2\eta/(1-4\eta)}-1}{2\beta^2}, \\ \lim_{\tau\to\infty}\frac{R_\nu(\tau)}{\tau} &= \lim_{\tau\to\infty}\frac{K_{\nu+1}(\tau)}{\tau K_\nu(\tau)} = \frac{\eta}{1-2\eta}, \\ D_{\nu+1}(\tau) &= \frac{K_{\nu+2}(\tau)K_\nu(\tau)}{K_{\nu+1}(\tau)} = \frac{1-2\eta}{1-4\eta}. \end{aligned}$$

Therefore, we can easily simulate an asymmetric standardised Student t distribution as:

$$\varepsilon_t^* = c(\beta,\nu,\tau)\beta\left[\frac{(1-2\eta)}{\eta\xi_t} - 1\right] + \sqrt{\frac{(1-2\eta)}{\eta\xi_t}}\sqrt{c(\beta,\nu,\tau)}z_t,$$

where $\xi_t \sim i.i.d.$ *Gamma* with mean η^{-1} and variance $2\eta^{-1}$, and $z_t|\xi_t$ is $i.i.d.$ $N(0,1)$.

If we further assume that $\eta < 1/8$, then

$$\begin{aligned} \frac{K_{\nu+3}(\tau)K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} &= \frac{(1-2\eta)^2}{(1-4\eta)(1-6\eta)} \\ \frac{K_{\nu+4}(\tau)K_\nu^3(\tau)}{K_{\nu+1}^4(\tau)} &= \frac{(1-2\eta)^3}{(1-4\eta)(1-6\eta)(1-8\eta)} \end{aligned}$$

so the skewness and kurtosis coefficients of the asymmetric t distribution will be:

$$E(\varepsilon_t^{*3}) = 16c^3(\beta,\nu,\tau)\frac{\eta^2}{(1-4\eta)(1-6\eta)}\beta^3 + 6c^2(\beta,\nu,\tau)\frac{\eta}{1-4\eta}\beta$$

and

$$E(\varepsilon_t^{*4}) = 12c^4(\beta,\nu,\tau)\frac{\eta^2(10\eta+1)}{(1-4\eta)(1-6\eta)(1-8\eta)}\beta^4$$
$$+12c^3(\beta,\nu,\tau)\frac{\eta(2\eta+1)}{(1-4\eta)(1-6\eta)}\beta^2 + 3\frac{1-2\eta}{1-4\eta}c^2(\beta,\nu,\tau).$$

Not surprisingly, we can obtain maximum asymmetry for a given kurtosis by letting $|\beta| \to \infty$. In contrast, a standardised version of the usual symmetric Student t with $1/\eta$ degrees of freedom is achieved when $\beta = 0$. Since $\lim_{\beta\to 0} c(\beta,\nu,\tau) = 1$, in that case the coefficient of kurtosis becomes

$$E(\varepsilon_t^{*4}) = 3\frac{1-2\eta}{1-4\eta}$$

for any $\eta < 1/4$.

C.3.2 Symmetric Laplace distribution

The asymmetric Laplace distribution is another special case of the GH distribution, which is achieved when $\tau = 0$ and $\nu = 1$. In fact, it is a special case of the asymmetric normal-gamma mixture, which allows ν to be any positive parameter. As is well known, the kurtosis coefficient of a symmetric Laplace distribution is 6. In the univariate case, the Laplace distribution is also a special case of the generalised error distribution (GED) with shape parameter fixed at 1, in contrast to the Gaussian distribution, which is also a special GED case with parameter 2.

The symmetric Laplace distribution is very easy to generate as

$$\varepsilon_t^* = \sqrt{\xi_t} z_t,$$

where ξ_t is an *i.i.d.* exponential (i.e. a *Gamma* with mean 1 and variance 1), and $z_t|\xi_t$ is *i.i.d.* $N(0,1)$. Alternatively, if u_t denotes a $(0,1)$ uniform variate, then we can also simulate a standardised symmetric Laplace random variable ε_t^* as

$$-\frac{1}{\sqrt{2}} sign\left(u_t - \frac{1}{2}\right) \ln\left(1 - 2\left|u_t - \frac{1}{2}\right|\right).$$

In effect, this procedure uses the fact that the absolute value of a Laplace is exponential, with a closed-form quantile function, while its sign is a shifted and scaled Bernoulli random variable that the values ± 1 with probability $1/2$ each.

C.4 Construction of the quarterly portfolios

We follow exactly the same procedure as Ken French uses to create annual returns from monthly ones. The first thing we do is to add the monthly gross return on the 1-month Tbill rate to the excess returns of the 6 value-weighted portfolios formed on size and book-to-market, the 6 value-weighted portfolios formed on size and operating profitability, and the 6 value-weighted portfolios formed on size and investment to transform each of them into monthly gross returns. Then we compound the monthly gross returns into quarterly gross returns by multiplication, and subtract the quarterly gross return on the 3-month Tbill (from the FRED database) to obtain our quarterly excess returns. Form those, we create the five FF factors using the appropriate long or short weights.

More formally, let $X_{t,i}^{(K,J,D)}$ be the net % return over month i, year t of some value-weight portfolio, with $i = 1, \ldots, 12$, where $D = SMALL$, BIG, $K = BM$, OP, INV and $J = LOW$, $NEUTRAL$, $HIGH$, with LOW and $HIGH$ denoting growth and value for BM portfolios, weak and robust for OP portfolios, and conservative and aggressive for INV portfolios. We

then calculate the quarterly portfolios as:

$$X_{t,I}^{(K,J,D)} = 100 \left[\prod_{i=3(I-1)+1}^{3I} \left(\frac{X_{t,i}^{(K,J,D)}}{100} + 1 \right) - 1 \right],$$

for $I = 1, 2, 3, 4$. Next, we apply the FF factor definitions. Specifically, the small minus big factor is

$$SMB = 1/3(SMB_{BM} + SMB_{OP} + SMP_{INV}),$$

where

$$\begin{aligned} SMB_K &= \frac{X^{(K,LOW,SMALL)} + X^{(K,NEUTRAL,SMALL)} + X^{(K,HIGH,SMALL)}}{3} \\ &\quad - \frac{X^{(K,LOW,BIG)} + X^{(K,NEUTRAL,BIG)} + X^{(K,HIGH,BIG)}}{3}. \end{aligned}$$

Similarly, the high minus low factor is obtained as

$$HML = \frac{X^{(BM,HIGH,SMALL)} + X^{(K,HIGH,BIG)}}{2} - \frac{X^{(BM,LOW,SMALL)} + X^{(K,LOW,BIG)}}{2},$$

the robust minus weak as

$$RMW = \frac{X^{(OP,HIGH,SMALL)} + X^{(OP,HIGH,BIG)}}{2} - \frac{X^{(OP,LOW,SMALL)} + X^{(OP,LOW,BIG)}}{2},$$

and the conservative minus aggressive as

$$CMA = \frac{X^{(INV,LOW,SMALL)} + X^{(INV,LOW,BIG)}}{2} - \frac{X^{(INV,HIGH,SMALL)} + X^{(OP,HIGH,BIG)}}{2}.$$

Finally, the quarterly excess return on the market can be obtained aggregating directly the monthly factor

$$Rm_{t,I} = 100 \left[\prod_{i=3(I-1)+1}^{3I} \left(\frac{Rm_{t,i} + Rf_{t,i}}{100} + 1 \right) - 1 \right] - Rf_{t,I}$$

where $Rf_{t,i}$ and $Rf_{t,I}$ are the one-month and three-month riskfree rate, respectively.

C.5 The symmetry component of the Jarque-Bera (1980) test without imposing normality

Consider a moment test based on the influence function

$$n(y; \pi, \omega) = \epsilon_t^3(\boldsymbol{\theta}_s) - 3\epsilon_t(\boldsymbol{\theta}_s)$$

where $\epsilon_t(\boldsymbol{\theta}_s) = \omega^{-1/2}(y_t - \pi)$, evaluated at the sample mean and variance. This influence function coincides with the third Hermite polynomial.

Using standard results (see e.g. Newey and McFadden (1994)), the asymptotic variance of

$$\begin{aligned}
&\frac{\sqrt{T}}{T}\sum_{t=1}^{T} n(y_t;\hat{\pi},\hat{\omega}) \\
= \quad &\frac{\sqrt{T}}{T}\sum_{t=1}^{T} n(y_t;\pi_0,\omega_0) + E\left(\begin{array}{cc} \frac{\partial n(y;\pi_0,\omega_0)}{\partial \pi} & \frac{\partial n(y;\pi_0,\omega_0)}{\partial \omega} \end{array} \right) \sqrt{T}\left(\begin{array}{c} \hat{\pi}-\pi_0 \\ \hat{\omega}-\omega_0 \end{array} \right) + o_p(1)
\end{aligned}$$

But the expected Jacobian matrix evaluated at the true value of the parameters is 0 under symmetry because

$$\begin{aligned}
\frac{\partial n(y;\pi,\omega)}{\partial \pi} &= -\frac{3}{\omega^{\frac{1}{2}}}\left[\epsilon_t^2(\boldsymbol{\theta}_s)-1\right], \\
\frac{\partial n(y;\pi,\omega)}{\partial \omega} &= -\frac{3}{2\omega}\left[\epsilon_t^2(\boldsymbol{\theta}_s)-1\right]\epsilon_t(\boldsymbol{\theta}_s).
\end{aligned}$$

Therefore, the asymptotic covariance matrix of the sample mean of the third Hermite polynomial evaluated at the sample mean and variance will be the same as if we could evaluate it at the true values. Consequently, a moment test of $H_0 : E[n(y;\pi,\omega)] = 0$ can be simply computed as the $t-$ratio of the sample mean of $n(y_t;\hat{\pi},\hat{\omega})$.

Interestingly, this moment test coincides with the outer product of the score version of the asymmetry component of the test of the null hypothesis of normality versus generalised hyperbolic alternatives in Mencía and Sentana (2012), which they argue remains valid under as long the true distribution is symmetric.

D Econometric methods

D.1 Log-likelihood function, score vector, Hessian and information matrices

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $p + r$ parameters of interest, which we assume variation free. Ignoring initial conditions, and assuming that $\sigma_t^2(\boldsymbol{\theta})$ is strictly positive, the log-likelihood function of a sample of size T based on a particular parametric distributional assumption will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, where $d_t(\boldsymbol{\theta}) = -1/2 \ln \sigma_t^2(\boldsymbol{\theta})$, $\varepsilon_t^*(\boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\theta})/\sigma_t(\boldsymbol{\theta})$ and $\varepsilon_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta})$.

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 $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. If $\mu_t(\boldsymbol{\theta})$, $\sigma_t^2(\boldsymbol{\theta})$ and $f(\varepsilon^*, \boldsymbol{\eta})$ are differentiable, then we can use the fact that

$$\partial d_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = -\tfrac{1}{2} \cdot \sigma_t^{-2}(\boldsymbol{\theta}) \cdot \partial \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = -\mathbf{Z}_{st}(\boldsymbol{\theta})$$

and

$$\begin{aligned}
\partial \varepsilon_t^*(\boldsymbol{\theta})/\partial \boldsymbol{\theta} &= -\sigma_t^{-1}(\boldsymbol{\theta}) \cdot \partial \mu_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} - \tfrac{1}{2}\sigma_t^{-2}(\boldsymbol{\theta}) \cdot \partial \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \varepsilon_t^*(\boldsymbol{\theta}) \\
&= -\mathbf{Z}_{lt}(\boldsymbol{\theta}) - \mathbf{Z}_{st}(\boldsymbol{\theta})\varepsilon_t^*(\boldsymbol{\theta}),
\end{aligned}$$

to show that

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \boldsymbol{\theta}} = \left[\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})\right] \left[\begin{array}{c} e_{lt}(\boldsymbol{\phi}) \\ e_{st}(\boldsymbol{\phi}) \end{array}\right] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \\
\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}),
\end{aligned}$$

where

$$\begin{aligned}
e_{lt}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= -\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^*, \\
e_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= -\left\{1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^*\right\},
\end{aligned}$$

depend on the specific distributional assumption.

Let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial \mathbf{s}_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'$. Assuming twice differentiability of the different functions involved, we will have

$$\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\phi}) = \frac{\partial \mathbf{Z}_{lt}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} e_{lt}(\boldsymbol{\phi}) + \frac{\partial \mathbf{Z}_{st}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} e_{st}(\boldsymbol{\phi}) + \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial e_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial e_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} \quad \text{(D16)}$$

$$\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial e_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\eta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial e_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\eta}'} \quad \text{(D17)}$$

$$\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi}) = \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \boldsymbol{\eta} \partial \boldsymbol{\eta}',$$

where

$$\begin{aligned}
\partial \mathbf{Z}_{lt}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}' &= -\tfrac{1}{2} \cdot \sigma_t^{-3}(\boldsymbol{\theta}) \cdot \partial \mu_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \partial \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta}' + \sigma_t^{-1}(\boldsymbol{\theta}) \cdot \partial^2 \mu_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}', \\
\partial \mathbf{Z}_{st}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}' &= -\tfrac{1}{2} \cdot \sigma_t^{-4}(\boldsymbol{\theta}) \cdot \partial \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \partial \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta}' + \tfrac{1}{2} \cdot \sigma_t^{-2}(\boldsymbol{\theta}) \cdot \partial^2 \sigma_t^2(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}', \\
\partial e_{lt}(\boldsymbol{\phi})/\partial \boldsymbol{\theta}' &= \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^* \partial \varepsilon^* \cdot \mathbf{Z}_{lt}'(\boldsymbol{\theta}) + \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^* \partial \varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta}) \cdot \mathbf{Z}_{st}'(\boldsymbol{\theta}) \\
\partial e_{st}(\boldsymbol{\phi})/\partial \boldsymbol{\theta}' &= \{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^* + \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^* \partial \varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta})\} \mathbf{Z}_{lt}'(\boldsymbol{\theta}) \\
&+ \{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta}) + \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right] / \partial \varepsilon^* \partial \varepsilon^* \cdot \varepsilon_t^{2*}(\boldsymbol{\theta})\} \cdot \mathbf{Z}_{st}'(\boldsymbol{\theta})
\end{aligned}$$

and $\partial^2 \ln f(\varepsilon^*, \eta)/\partial \varepsilon^* \partial \varepsilon^*$, $\partial^2 \ln f(\varepsilon^*, \eta)/\partial \varepsilon^* \partial \boldsymbol{\eta}'$ and $\partial \ln f(\varepsilon^*, \eta)/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'$ depend on the specific distribution assumed for estimation purposes (see FSC for the Student t).

Given correct specification, $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}_{dt}'(\boldsymbol{\phi}), \mathbf{e}_{rt}(\boldsymbol{\phi})]'$ evaluated at the true parameter values is an *iid* sequence, and therefore, the score vector $\mathbf{s}_t(\boldsymbol{\phi})$ will be a vector martingale difference sequence. Then, the results in Crowder (1976) imply that, under suitable regularity conditions, the asymptotic distribution of the feasible ML estimator will be $\sqrt{T}(\boldsymbol{\phi}_T - \boldsymbol{\phi}_0) \rightarrow N[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)]$, where $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$, where

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{\phi}) &= -E\left[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\right] = V\left[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta}) \mathcal{M}(\boldsymbol{\eta}) \mathbf{Z}_t'(\boldsymbol{\theta}), \\
\mathbf{Z}_t(\boldsymbol{\theta}) &= \left(\begin{array}{cc} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{array}\right) = \left(\begin{array}{ccc} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{array}\right),
\end{aligned}$$

and

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathcal{M}'_{lr}(\boldsymbol{\eta}) & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix}.$$

In the Student t case, this matrix is simply

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \frac{\nu(\nu+1)}{(\nu-2)(\nu+3)} & 0 & 0 \\ 0 & \frac{(\nu+1)}{(\nu+3)} & -\frac{6\nu^2}{(\nu-2)(\nu+1)(\nu+3)} \\ 0 & -\frac{6\nu^2}{(\nu-2)(\nu+1)(\nu+3)} & \frac{\nu^4}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right)\right] - \frac{\nu^4\left[\nu^2+(\nu-4)-8\right]}{2(\nu-2)^2(\nu+1)(\nu+3)} \end{pmatrix}.$$

where $\psi(.)$ is the di-gamma function (see Abramowitz and Stegun (1964)), which under normality reduces to

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3/2 \end{pmatrix}.$$

D.2 Gaussian pseudo maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote the Gaussian pseudo-ML (PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ in which $\boldsymbol{\varrho}$ is set to zero. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\mu_t(\boldsymbol{\theta})$ and $\sigma_t^2(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments. Proposition 2 in Fiorentini and Sentana (2007) derives the asymptotic distribution of the pseudo-ML estimator of $\boldsymbol{\theta}$ when $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is *i.i.d.*:

Proposition 10 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is i.i.d. $D(0,1,\varrho_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)]$, where*

$$\begin{aligned} \mathcal{C}(\boldsymbol{\phi}) &= \mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi}), \\ \mathcal{A}(\boldsymbol{\phi}) &= -E\left[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\ \mathcal{A}_t(\boldsymbol{\phi}) &= -E[\mathbf{h}_{\boldsymbol{\theta}\boldsymbol{\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\,\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta}), \\ \mathcal{B}(\boldsymbol{\phi}) &= V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right], \\ \mathcal{B}_t(\boldsymbol{\phi}) &= V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\,\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta}), \\ \text{and } \mathcal{K}(\varphi,\kappa) &= V[\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\,\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \begin{bmatrix} 1 & \varphi(\varrho) \\ \varphi(\boldsymbol{\varrho}) & \kappa(\boldsymbol{\varrho}) - 1 \end{bmatrix}, \end{aligned} \tag{D18}$$

which only depends on $\boldsymbol{\varrho}$ through the population coefficients of asymmetry and kurtosis

$$\varphi(\boldsymbol{\varrho}) = E(\varepsilon_t^{*3}|\boldsymbol{\varrho}). \tag{D19}$$

$$\kappa(\boldsymbol{\varrho}) = E(\varepsilon_t^{*4}|\boldsymbol{\varrho}). \tag{D20}$$

Given that $\varphi(\boldsymbol{\varrho}) = 0$ and $\kappa = 2/(\nu-4)$ for the Student t distribution with ν degrees of freedom, it trivially follows that in that case $\mathcal{B}_t(\boldsymbol{\phi})$ reduces to

$$\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\mu_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\nu-1}{2(\nu-4)}\frac{1}{\sigma_t^4(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}.$$

D.3 Semiparametric estimators of $\boldsymbol{\theta}$

González-Rivera and Drost (1999) obtain the semiparametric efficient score and the corresponding efficiency bound for univariate models:

Proposition 11 *If $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0$ is i.i.d. $(1,0)$ with density function $f(\varepsilon_t^*; \boldsymbol{\varrho})$, where $\boldsymbol{\varrho}$ are some shape parameters and $\boldsymbol{\varrho} = \mathbf{0}$ denotes normality, such that both its Fisher information matrix for location and scale*

$$\mathcal{M}_{dd}(\boldsymbol{\varrho}) = V\left[\mathbf{e}_{dt}(\boldsymbol{\theta}, \boldsymbol{\varrho})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}\right]$$
$$= V\left\{ \left[\begin{array}{c} e_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ e_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \end{array} \right] \middle| \boldsymbol{\theta}, \boldsymbol{\varrho} \right\} = V\left\{ \left[\begin{array}{c} -\partial \ln f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \varepsilon^* \\ -vec\left\{\mathbf{I}_N + \partial \ln f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta})\right\} \end{array} \right] \middle| \boldsymbol{\theta}, \boldsymbol{\varrho} \right\}$$

and the matrix of third and fourth order central moments

$$\mathcal{K}(\boldsymbol{\varrho}) = V[\,\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\,\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}] \tag{D21}$$

are bounded, then the semiparametric efficient score will be given by:

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^{-1}(\varphi, \kappa)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})\right], \tag{D22}$$

while the semiparametric efficiency bound is

$$\mathcal{S}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^1(\varphi, \kappa)\mathcal{K}(0)\right]\mathbf{Z}_d'(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0), \tag{D23}$$

where $+$ denotes Moore-Penrose inverses, and $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{M}_{dd}(\boldsymbol{\varrho})\mathbf{Z}_{dt}'(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{\varrho}\right]$.

In practice, $f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]$ has to be replaced by a non-parametric density estimator, which is typically obtained by kernel methods.

Hodgson and Vorkink (2001), Hafner and Rombouts (2007) and other authors have suggested semi-parametric estimators of $\boldsymbol{\theta}$ which limit the admissible distributions of $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ to the class of symmetric ones. Proposition 7 in Fiorentini and Sentana (2007) provides the resulting elliptically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition 12 *When $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0$ is i.i.d. $s(0,1,\boldsymbol{\varrho}_0)$ with $1 < \kappa_0 < \infty$, the elliptically symmetric semiparametric efficient score is given by:*

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0) = \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\left\{-\left[1 + \varepsilon_t(\boldsymbol{\theta}_0)\partial \ln f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial \varepsilon^*\right] - \frac{2}{\kappa_0 - 1}\left[\varepsilon_t^2(\boldsymbol{\theta}_0) - 1\right]\right\}, \tag{D24}$$

where

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\phi}_0)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]\begin{pmatrix} 0 \\ 1 \end{pmatrix} = E\left\{\frac{1}{2\sigma_t^2(\boldsymbol{\theta})}\frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\middle|\boldsymbol{\phi}_0\right\}, \tag{D25}$$

while the elliptically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0) \cdot \left[\mathcal{M}_{ss}(\boldsymbol{\varrho}_0) - \frac{4}{\kappa_0 - 1}\right]. \tag{D26}$$

In practice, $e_{dt}(\boldsymbol{\phi})$ has to be replaced by a semiparametric estimate obtained from the density of ε_t^* that imposes symmetry. The simplest way to do this is by averaging the non-parametric density estimators at ε_t^* and $-\varepsilon_t^*$. Alternatively, one can estimate the common density of $\pm\varepsilon_t^*$ from the density of the Box-Cox transformation $k^{-1}|\varepsilon_t^*|^k - 1$ for some $k \geq 0$.

D.4 Student t-based (pseudo) maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta},\eta} L_T(\boldsymbol{\theta},\eta)$ denote the t-based pseudo-ML (t-PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ obtained by assuming that the conditional distribution is $t(0,1,\eta)$. Proposition 5 in Fiorentini and Sentana (2018) shows that this estimator is asymptotically equivalent to the Gaussian PML estimator when the conditional distribution is platykurtic. They also show that if the conditional mean and variance can be parametrised as in Linton (1993) and Newey and Steigerwald (1997), then some of the reparametrised mean and variance parameters will be consistently estimated even if the true conditional distribution is not a Student t. In our context, the robustness of the Student t serial correlation tests under conditional symmetry follows from the fact that the only parameter that is inconsistently estimated is ω in those circumstances. More generally, its robustness under possibly asymmetric distributions derives from the fact that we can reparametrise the mean of (1) as $\delta\sqrt{\omega} + \rho y_{t-1}$. Therefore, the t-based ML estimator of ρ continues to be consistent even if the estimators of ω and π are inconsistent. The argument for the α is slightly different, because a Student log-likelihood function can only estimate $\gamma = \alpha/\omega$ consistently in those circumstances. Nevertheless, given that α is 0 under the null, the t-based ML estimator of α continues to be consistent even if the estimators of ω and π are inconsistent.

D.5 Kotz-based (pseudo) maximum likelihood estimators

The original Kotz distribution (see Kotz (1975)) is a member of the spherical family, and thereby symmetric in the univariate case. Its main distinctive characteristic is that ε^{*2} follows a gamma distribution with mean 1 and variance $(3\varkappa_0 + 2)$, where

$$\varkappa = E(\varepsilon^{*4}|\boldsymbol{\eta})/3 - 1$$

is the coefficient of multivariate excess kurtosis of ε^* (see Mardia (1970)), which is trivially 0 under normality. In fact, the Kotz distribution nests the normal distribution when $\varkappa = 0$, in which ε^{*2} follows with a chi square distribution with one degree of freedom, but it can also be either platykurtic ($\varkappa < 0$) or leptokurtic ($\varkappa > 0$), although in the second case the Jensen inequality restriction $E(\varepsilon^{*4}) \geq E(\varepsilon^{*2}) = 1$ implies that $\varkappa \geq -2/3$. Such a nesting provides an analytically convenient generalisation of the normal. Specifically, the kernel of the distribution of ε^{*2} is

$$g(\varepsilon^{*2};\varkappa) = -\frac{3\varkappa}{2(3\varkappa+2)}\ln\varepsilon^{*2} - \frac{1}{3\varkappa+2}\varepsilon^{*2},$$

while the constant of integration becomes

$$c(\varkappa) = -\ln\Gamma\left(\frac{1}{3\varkappa+2}\right) - \frac{1}{3\varkappa+2}\ln(3\varkappa+2)$$

(see Amengual and Sentana (2011)). Therefore, the density of a leptokurtic Kotz distribution has a pole at 0, and an antimode in the platykurtic case, which is a potential drawback from an empirical point of view.

The contribution of the t^{th} observation to the log-likelihood function is

$$l_t(\boldsymbol{\theta}, \varkappa) = -\frac{1}{2}\ln \sigma_t^2(\boldsymbol{\theta}) + c(\varkappa) + g(\varepsilon_t^{*2}; \varkappa).$$

As a result, the damping factor becomes

$$\delta(\varepsilon^{*2}; \varkappa) = \frac{1}{3\varkappa + 2}\left(\frac{3\varkappa}{\varepsilon^{*2}} + 2\right).$$

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\varkappa} L_T(\boldsymbol{\theta},\varkappa)$ denote the t-based pseudo-ML (t-PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ obtained by assuming that the conditional distribution is a standardised version of the univariate $Kotz(0, 1, \varkappa)$.

Straightforward algebra shows that the ML estimator of the mean sets to 0 the following moment condition

$$\frac{1}{3\varkappa + 2}[3\varkappa\breve{\varepsilon}_T^{*-1}(\boldsymbol{\theta}) + 2\bar{\varepsilon}_T^{*}(\boldsymbol{\theta})] = 0,$$

where $\breve{\varepsilon}_T^{*-1}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\varepsilon_t^{*-1}(\boldsymbol{\theta})$ is the reciprocal of the harmonic mean of the standardised residuals and $\bar{\varepsilon}_T^{*}(\boldsymbol{\theta})$ their arithmetic one. Therefore, the estimator is such that it makes a combination of the arithmetic and harmonic mean of the standardised residuals equal to 0. In contrast, the ML estimator of the variance can be concentrated out of the log-likelihood function as:

$$\omega(\pi) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \pi)^2$$

Finally, the score with respect to the excess kurtosis parameter $\varkappa$ is

$$s_{\varkappa t}(\boldsymbol{\theta}, \varkappa) = \varepsilon_t^{*2} - \ln \varepsilon_t^{*2} + \left[\psi\left(\frac{1}{3\varkappa + 2}\right) + \ln(3\varkappa + 2) - 1\right],$$

where $\psi(.)$ is the digamma (or Gauss psi) function (see Abramowitz and Stegun (1964)).

We can combine the moments of the gamma and reciprocal gamma random variables to show that

$$\mathrm{M}_{ll}(\varkappa) = \frac{9\varkappa + 2}{(3\varkappa + 1)(3\varkappa + 2)}, \tag{D27}$$

as long as $\varkappa > -1/3$,

$$\mathrm{M}_{ss}(\varkappa) = \frac{\varkappa + 2}{3\varkappa + 2},$$

and $\mathrm{M}_{sr}(\varkappa) = 0\ \forall\varkappa$, as in the Gaussian case, so that the information matrix is block diagonal between the mean, variance and shape parameters.

To sample the Kotz innovations, we exploit the fact that $\varepsilon_t^* = \sqrt{\xi_t} u_t$, where u_t is a shifted and scaled Bernoulli random variable that the values ± 1 with probability $1/2$ each, and ξ_t is a univariate Gamma with mean 1 and variance $(3\varkappa + 2)$.

Like in the Student t case, all mean parameters will be consistently estimated if the true conditional distribution is symmetric, while only ρ will remain consistent under asymmetry. And while ω will be inconsistently estimated unless the true distribution is Kotz, $\gamma = \alpha/\omega$ will be consistently estimated regardless.

D.6 Laplace-based (pseudo) maximum likelihood estimators

The Laplace (or double exponential) distribution, which is also a member of the generalised hyperbolic distribution, contains no shape parameters. As is well known, the ML estimator of the location parameter is given by the sample median, $med(y_1, \ldots, y_T)$. In turn, the estimator of the variance parameter ω is given by the twice the square of the mean absolute deviation around the median. Specifically,

$$\hat{\omega}_T = 2 \left[\frac{1}{T} \sum_{t=1}^{T} |y_t - med(y_1, \ldots, y_T)| \right]^2 .$$

Although the lack of shape parameters implies that the Laplace distribution is not very flexible, the fact that it is symmetric implies that the robustness properties of the pseudo ML estimators of ρ and γ are exactly the same as in the Student and Kotz-based log-likelihood functions.

D.7 Discrete mixtures of normals-based (pseudo) maximum likelihood estimators

The EM algorithm discussed by Dempster, Laird and Rubin (1977) allows us to obtain initial values as close to the optimum as desired. The recursions are as follows:

$$\begin{aligned}
\hat{\lambda}^{(n} &= \frac{1}{T} \sum\nolimits_{t=1}^{T} w(y_t; \boldsymbol{\phi}^{(n-1)}) \\
\hat{\mu}_1^{(n} &= \frac{1}{\hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t w(y_t; \boldsymbol{\phi}^{(n-1)}), \\
\hat{\mu}_2^{(n} &= \frac{1}{1 - \hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t [1 - w(y_t; \boldsymbol{\phi}^{(n-1)})], \\
\hat{\sigma}_1^{2(n} &= \frac{1}{\hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t^2 w(y_t; \boldsymbol{\phi}^{(n-1)}) - \left(\hat{\mu}_1^{(n} \right)^2, \\
\hat{\sigma}_2^{2(n} &= \frac{1}{1 - \hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t^2 [1 - w(y_t; \boldsymbol{\phi}^{(n-1)})] - \left(\hat{\mu}_2^{(n} \right)^2
\end{aligned}$$

where

$$\begin{aligned} w(y_t;\boldsymbol{\phi}) &= \frac{\frac{\lambda}{\sigma_1}\phi\left(\frac{y_t-\mu_1}{\sigma_1}\right)}{\frac{\lambda}{\sigma_1}\phi\left(\frac{y_t-\mu_1}{\sigma_1}\right)+\frac{1-\lambda}{\sigma_2}\phi\left(\frac{y_t-\mu_2}{\sigma_2}\right)} \\ &= \frac{\frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon_t^*(\boldsymbol{\theta}_s)-\mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right]}{\frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon_t^*(\boldsymbol{\theta}_s)-\mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right]+\frac{1-\lambda}{\sigma_2^*(\boldsymbol{\eta})}\phi\left[\frac{\varepsilon_t^*(\boldsymbol{\theta}_s)-\mu_2^*(\boldsymbol{\eta})}{\sigma_2^*(\boldsymbol{\eta})}\right]} = w[\epsilon_t^*(\boldsymbol{\theta}_s);\boldsymbol{\eta}] \end{aligned}$$

and $\phi(.)$ denotes the standard normal density.

From those recursions it is easy to check that

$$\begin{aligned} \hat{\pi}^{(n} &= \hat{\mu}_1^{(n}\hat{\lambda}^{(n} + \hat{\mu}_2^{(n}(1-\hat{\lambda}^{(n}) = \frac{1}{T}\sum\nolimits_{t=1}^{T} y_t, \\ \hat{\sigma}^{2(n} &= [(\hat{\mu}_1^{(n})^2 + \hat{\sigma}_1^{2(n}]\hat{\lambda}^{(n} + [(\hat{\mu}_2^{(n})^2 + \hat{\sigma}_2^{2(n}](1-\hat{\lambda}^{(n}) - (\hat{\pi}^{(n})^2 = \frac{1}{T}\sum\nolimits_{t=1}^{T} y_t^2 - \left(\frac{1}{T}\sum\nolimits_{t=1}^{T} y_t\right)^2, \end{aligned}$$

for all n regardless of the values of $\boldsymbol{\phi}^{(n-1}$. This means that $\hat{\lambda}^{(n}$, $\hat{\upsilon}^{(n} = \hat{\sigma}_2^{2(n}/\hat{\sigma}_1^{2(n}$ and

$$\hat{\delta}^{(n} = \frac{\hat{\mu}_1^{(n} - \hat{\mu}_2^{(n}}{\sqrt{\hat{\lambda}^{(n}\hat{\sigma}_1^{2(n} + (1-\hat{\lambda}^{(n})\hat{\sigma}_2^{2(n}}}$$

will yield the EM recursions for a mixture model parametrised in terms of π, ω^2 and λ, δ and υ, which are the parameters of the standardised version in appendix C.1.

Since the ML estimators constitute the fixed point of the EM recursions, (i.e. $\boldsymbol{\phi} = \boldsymbol{\phi}^{(\infty}$), another implication of the above result is that $\hat{\pi}$ and $\hat{\omega}$ coincide with the Gaussian PML estimators. As a result, we can maximise the log-likelihood function with respect to λ, δ and υ keeping $\hat{\pi}$ and $\hat{\sigma}^2$ fixed at their Gaussian pseudo ML values. Interestingly, this somewhat surprising result will continue to be true even in a complete log-likelihood situation in which we would observe not only y_t but also s_t. In addition, it is straightforward to prove that the same result holds for finite mixtures of normals with more than two components.

As a result, the ML estimators of π and ω continue to be consistent under distributional misspecification. Similarly, the estimators of ρ and $\alpha = \omega\gamma$ will also remain consistent in that case too, as explained in Fiorentini and Sentana (2018).

Nevertheless, the log-likelihood function of a mixture distribution has a pole for each observation. Specifically, it will go to infinity if we set $\hat{\mu}_1 = y_t$ and let $\hat{\sigma}_1^2$ go to 0. In practice, we deal with this issue by starting the EM algorithm from many different starting values. In addition, there is a trivial identification issue that arises by exchanging the labels of the components. We solve this problem by restricting υ to the range $(0, 1)$ so that the first component is the one with the largest variance.

TABLE 1A: Monte Carlo size of predictability tests. Asymptotic critical values.

Test against	AR(1)			AR(12)			ARCH(1)			GARCH(1,1)		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP Gaussian												
Gaussian	9.61	4.73	0.98	13.62	7.43	1.69	7.98	3.71	0.84	11.61	5.77	1.13
Student t	9.72	4.78	0.92	13.56	7.47	1.69	8.28	3.84	0.83	11.84	5.97	1.14
DLSMN	9.71	4.79	0.85	12.54	6.83	1.40	7.90	3.86	1.13	10.47	4.93	0.77
GED(1)	9.99	5.00	0.98	12.45	6.68	1.34	9.22	4.37	0.73	12.06	6.24	1.22
Semiparametric	9.73	4.98	0.88	13.37	7.36	1.69	8.76	4.00	0.92	11.78	5.79	1.04
Sym. Semipar.	9.96	4.94	0.89	13.58	7.40	1.70	8.49	3.88	0.89	11.88	5.92	1.05
DGP Student $t(6)$												
Gaussian	9.69	4.69	0.85	13.62	7.29	1.63	5.70	3.01	0.97	8.13	3.46	0.40
Student t	9.80	4.77	0.91	13.58	7.24	1.63	7.96	3.64	0.85	10.64	4.88	0.63
DLSMN	9.61	4.59	0.98	12.83	6.86	1.41	7.07	3.86	1.37	9.45	4.17	0.60
Laplace	10.04	4.91	0.86	12.35	6.59	1.35	7.90	3.69	0.73	10.40	4.71	0.67
Semiparametric	9.69	4.76	0.88	13.07	7.11	1.47	8.28	4.15	1.41	10.44	4.75	0.70
Sym. Semipar.	9.88	4.78	0.96	13.29	7.27	1.57	8.08	4.21	1.38	10.46	4.92	0.75
DGP DLSMN(-0.85,0.16,0.05)												
Gaussian	9.59	4.73	0.83	13.44	7.40	1.61	5.22	2.84	1.01	7.09	2.92	0.37
Student t	9.79	4.75	0.90	13.61	7.36	1.60	7.26	3.46	0.83	10.21	4.56	0.52
DLSMN	9.64	4.76	0.86	13.15	6.93	1.57	7.28	4.23	1.56	9.30	4.17	0.55
Laplace	9.98	4.91	0.97	12.30	6.65	1.31	7.13	3.40	0.81	9.50	4.39	0.63
Semiparametric	9.74	4.83	0.97	13.49	7.34	1.70	8.16	4.47	1.59	10.15	4.95	0.78
Sym. Semipar.	9.80	4.88	0.95	13.51	7.38	1.70	7.90	4.30	1.56	10.20	4.89	0.75
DGP Gram-Charlier(0,3.0)												
Gaussian	9.01	4.60	0.90	12.93	7.08	1.51	5.80	3.62	1.37	9.20	3.88	0.40
Student t	10.14	4.86	0.89	12.47	6.49	1.42	8.71	3.89	0.80	10.80	5.38	0.80
DLSMN	10.06	4.91	0.97	12.06	6.15	1.39	8.56	4.09	0.99	10.08	4.75	0.73
Laplace	9.98	4.96	0.92	11.88	6.34	1.29	7.14	3.48	0.99	10.59	4.88	0.61
Semiparametric	9.96	4.84	0.96	11.97	6.18	1.39	8.39	4.52	1.35	9.77	4.52	0.66
Sym. Semipar.	9.88	4.82	1.00	12.03	6.20	1.37	8.03	4.32	1.34	9.56	4.28	0.60
DGP Gram-Charlier(-0.8,3.0)												
Gaussian	9.18	4.50	0.89	12.96	7.01	1.59	5.63	3.44	1.40	8.92	3.55	0.40
Student t	9.95	4.91	0.94	12.30	6.33	1.39	8.69	3.90	0.78	10.84	4.99	0.80
DLSMN	9.84	4.81	0.93	11.72	6.09	1.28	8.56	4.32	1.31	9.55	4.42	0.65
Laplace	9.92	5.00	0.95	11.85	6.17	1.19	6.92	3.28	1.00	10.33	4.50	0.62
Semiparametric	9.92	4.87	1.00	11.73	6.29	1.29	8.12	4.28	1.31	9.19	4.21	0.61
Sym. Semipar.	10.10	4.96	0.98	11.75	6.28	1.32	7.75	3.93	1.27	9.10	3.94	0.56
DGP Gram-Charlier(-0.8,3.0) with outliers												
Gaussian	18.20	9.75	1.90	10.79	5.15	0.81	22.93	13.24	3.40	1.00	0.23	0.01
Student t	8.73	4.04	0.66	10.96	5.55	1.03	12.89	7.00	1.57	6.35	2.37	0.22
DLSMN	8.80	4.07	0.68	11.14	5.70	1.03	10.76	5.82	1.46	7.32	3.31	0.40
Laplace	9.32	4.34	0.68	10.79	5.46	1.04	20.70	11.73	3.16	2.23	0.55	0.04
Semiparametric	9.38	4.45	0.78	10.65	5.26	1.00	18.81	11.62	3.86	8.33	3.94	0.71
Sym. Semipar.	9.73	4.76	0.89	10.48	5.17	0.98	33.06	22.18	8.46	8.29	4.00	0.79

Monte Carlo empirical rejection rates of mean variance predictability tests. Sample length=100. Replications=20,000.

TABLE 1B: Monte Carlo size of predictability tests. WARP bootstrap critical values.

Test against	AR(1)			AR(12)			ARCH(1)			GARCH(1,1)		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP Gaussian												
Gaussian	9.29	4.69	1.00	9.99	5.11	0.94	10.35	5.16	1.11	10.11	5.17	1.19
Student t	9.30	4.71	0.92	10.05	5.12	0.91	10.32	5.20	1.09	10.11	5.15	1.17
DLSMN	9.67	4.84	0.88	9.98	5.04	0.97	9.92	4.59	0.92	10.32	5.34	1.12
Laplace	9.56	4.84	1.01	10.35	5.04	0.88	10.21	5.14	1.15	9.96	5.06	1.07
Semiparametric	9.42	4.90	0.90	10.30	5.41	1.03	10.57	5.10	1.08	9.93	5.15	1.19
Sym. Semipar.	9.59	4.79	0.91	10.48	5.38	1.03	10.53	5.04	1.05	10.23	5.19	1.11
DGP Student $t(6)$												
Gaussian	10.01	4.96	1.03	10.53	5.23	0.99	9.35	4.52	0.91	8.97	4.38	0.84
Student t	9.71	4.75	0.99	10.57	4.94	1.03	10.10	4.95	1.04	9.59	4.69	0.81
DLSMN	9.69	4.58	1.10	10.57	5.29	0.97	9.64	4.80	1.03	9.84	4.95	0.98
Laplace	9.45	4.66	0.89	10.03	4.92	0.92	9.98	5.04	0.90	9.54	4.58	0.82
Semiparametric	9.44	4.67	0.91	10.42	5.14	0.92	10.38	5.22	1.31	9.63	4.59	0.91
Sym. Semipar.	9.75	4.81	1.06	10.57	5.01	0.95	10.27	5.19	1.26	9.74	4.71	0.94
DGP DLSMN(-0.85,0.16,0.05)												
Gaussian	9.79	4.93	0.95	10.25	5.06	0.92	8.96	4.61	0.97	8.71	4.21	0.70
Student t	9.40	4.63	0.92	10.30	5.10	0.92	9.53	4.92	1.01	9.75	4.72	0.85
DLSMN	9.63	4.79	0.92	10.64	5.41	1.13	9.58	4.98	0.97	10.08	5.10	0.95
Laplace	9.62	4.57	1.08	10.05	5.15	1.01	9.78	4.69	0.92	9.28	4.71	0.84
Semiparametric	9.46	4.56	0.97	10.64	5.58	1.03	10.46	5.40	1.20	9.98	5.24	1.02
Sym. Semipar.	9.25	4.81	0.96	10.54	5.42	1.03	9.80	5.27	1.17	9.93	5.13	0.96
DGP Gram-Charlier(0,3.0)												
Gaussian	9.41	4.87	1.07	10.09	5.15	1.04	9.90	5.25	0.98	10.05	5.06	0.97
Student t	9.85	4.72	0.92	9.87	4.87	1.01	10.37	5.49	1.03	9.90	5.06	1.11
DLSMN	9.88	4.89	0.94	10.23	5.07	1.10	10.41	5.08	0.83	10.40	5.38	1.28
Laplace	9.74	4.88	1.06	9.99	5.08	0.92	10.06	5.13	0.94	9.96	5.03	1.05
Semiparametric	9.66	4.92	0.97	10.13	4.85	1.02	9.86	5.10	1.05	9.93	5.10	1.03
Sym. Semipar.	9.74	4.92	0.99	10.07	4.93	0.98	9.95	5.29	1.17	9.97	4.84	0.96
DGP Gram-Charlier(-0.8,3.0)												
Gaussian	9.49	4.81	1.04	10.11	5.20	1.14	9.80	4.67	0.97	10.13	4.94	0.93
Student t	9.77	4.88	1.07	10.08	4.70	0.88	10.39	5.27	1.09	9.96	4.97	0.91
DLSMN	9.83	4.85	1.05	10.00	5.13	1.10	10.62	5.04	1.06	10.44	5.27	1.05
Laplace	9.73	5.00	1.06	10.12	5.01	0.89	9.95	4.84	1.02	9.89	4.88	0.82
Semiparametric	10.02	4.93	1.16	9.88	5.39	0.95	9.83	4.88	1.01	9.79	5.01	0.91
Sym. Semipar.	10.14	5.10	1.13	10.06	5.19	1.03	9.58	4.52	0.92	9.88	4.71	0.93
DGP Gram-Charlier(-0.8,3.0) with outliers												
Gaussian	18.87	10.26	1.82	7.92	3.37	0.49	34.14	17.98	1.58	1.13	0.33	0.03
Student t	8.24	3.87	0.71	9.21	4.58	0.78	13.53	8.06	1.98	5.40	2.05	0.24
DLSMN	8.45	4.01	0.68	9.74	4.71	0.82	11.78	6.71	1.60	7.12	3.31	0.55
Laplace	8.96	4.36	0.81	8.85	4.32	0.78	25.25	15.83	2.57	2.04	0.59	0.07
Semiparametric	9.01	4.23	0.80	9.18	4.51	0.81	20.52	11.67	2.54	8.73	4.47	0.95
Sym. Semipar.	9.74	4.53	0.87	8.92	4.61	0.74	36.18	24.10	6.26	8.54	4.54	1.12

Monte Carlo empirical rejection rates of mean variance predictability tests. Sample length=100. Replications=20,000.

TABLE 1C: Monte Carlo power of mean predictability tests. WARP bootstrap critical values.

True process	AR(1): $\rho = 0.2$						ARMA(1,1): $\rho = 0.98$ $\varphi = -0.92$					
Test against	AR(1)			AR(12)			AR(1)			AR(12)		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP Gaussian												
Gaussian	58.49	44.96	22.82	9.88	4.93	1.15	18.59	11.88	4.43	29.09	21.29	10.94
Student t	57.95	44.60	22.85	9.93	4.96	1.17	18.54	11.83	4.44	29.06	21.19	11.00
DLSMN	47.66	35.38	17.29	9.96	5.04	1.15	16.98	10.46	3.67	25.29	17.82	8.69
Laplace	42.77	31.36	13.74	10.32	5.20	1.04	16.04	9.56	3.33	24.71	17.81	7.81
Semiparametric	54.00	40.74	19.22	10.06	5.16	1.23	17.85	11.07	4.04	28.19	20.17	10.58
Sym. Semipar.	56.17	43.21	20.86	10.18	5.05	1.23	18.22	11.58	4.33	28.57	20.58	10.56
DGP Student $t(6)$												
Gaussian	59.28	46.67	23.81	9.70	5.04	0.98	19.39	12.07	4.78	29.92	21.53	11.31
Student t	62.71	50.63	28.13	10.47	5.39	1.19	20.47	13.30	5.39	33.77	25.25	14.01
DLSMN	54.03	42.34	21.80	10.51	5.21	1.23	18.87	11.62	4.40	30.70	22.28	11.96
Laplace	52.10	39.50	18.61	10.62	5.41	1.25	18.61	11.51	4.01	31.71	23.10	11.56
Semiparametric	59.89	47.21	24.00	10.61	5.55	1.05	19.92	12.34	4.78	33.18	24.68	13.72
Sym. Semipar.	61.39	49.19	26.03	10.70	5.50	0.97	20.26	12.93	5.01	33.01	25.21	13.81
DGP DLSMN(-0.85,0.16,0.05)												
Gaussian	58.87	46.15	24.04	9.74	4.91	0.96	18.94	11.82	4.54	29.95	22.04	10.93
Student t	62.36	49.29	26.45	10.31	5.51	1.15	20.02	12.64	5.33	33.23	25.09	13.06
DLSMN	55.27	43.53	22.31	10.73	5.41	1.17	18.30	11.49	4.46	30.70	23.08	12.23
Laplace	48.55	35.52	16.54	11.01	5.62	1.25	17.34	10.70	3.77	29.13	21.17	10.40
Semiparametric	60.26	46.97	24.47	10.94	5.79	1.29	19.76	12.31	4.98	33.50	25.01	13.55
Sym. Semipar.	61.48	49.36	25.66	10.87	5.80	1.23	19.95	12.56	5.21	33.45	25.11	13.71
DGP Gram-Charlier(0,3.0)												
Gaussian	60.62	47.51	23.44	9.46	4.75	1.12	19.21	12.57	4.37	29.60	21.75	11.05
Student t	76.32	65.15	41.03	12.18	6.73	1.67	25.07	16.89	7.29	44.53	36.20	21.57
DLSMN	73.63	62.62	39.37	12.74	7.29	1.95	24.61	16.23	6.44	45.27	36.31	22.38
Laplace	64.50	51.50	27.01	11.76	6.25	1.49	21.95	14.12	5.44	39.38	30.74	17.34
Semiparametric	78.36	67.94	43.97	12.91	7.35	1.78	25.41	16.91	6.86	47.38	38.86	24.06
Sym. Semipar.	79.35	69.36	45.52	12.91	7.44	1.69	25.56	17.39	7.26	47.40	38.81	24.09
DGP Gram-Charlier(-0.8,3.0)												
Gaussian	59.48	46.67	23.13	9.74	4.96	1.03	18.98	12.55	4.42	29.62	21.97	11.47
Student t	78.50	67.83	44.50	12.42	6.75	1.83	25.83	17.69	7.72	46.52	37.68	23.92
DLSMN	78.46	68.47	46.17	13.73	7.46	2.08	26.82	18.21	7.61	50.08	41.31	25.92
Laplace	65.92	53.75	29.78	11.71	6.44	1.50	22.07	14.70	5.60	40.44	32.20	18.06
Semiparametric	82.56	73.47	51.09	13.77	7.96	2.00	28.32	19.39	8.06	52.81	43.65	28.11
Sym. Semipar.	82.97	74.00	52.51	13.78	7.78	2.02	28.13	19.20	8.31	51.70	42.52	27.24

Monte Carlo empirical rejection rates of mean predictability tests. Sample length=100. Replications=20,000.

TABLE 1D: Monte Carlo power of variance predictability tests. WARP bootstrap critical values.

True process	ARCH(1): $\alpha = 0.25$						GARCH(1,1): $\alpha = 0.1$ $\beta = 0.88$					
Test against	ARCH(1)			GARCH(1,1)			ARCH(1)			GARCH(1,1)		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP Gaussian												
Gaussian	55.52	46.78	29.48	15.84	10.08	3.86	26.62	19.72	9.36	39.63	31.50	18.95
Student t	53.34	44.95	28.31	14.64	8.70	3.29	26.10	19.66	9.86	39.07	31.48	18.93
DLSMN	44.49	35.45	18.83	14.67	8.65	2.94	23.29	16.06	6.02	31.49	24.07	13.77
Laplace	49.69	41.11	25.83	13.72	8.19	2.77	24.76	17.77	8.80	37.72	29.76	17.93
Semiparametric	44.05	34.89	18.82	13.46	8.02	2.93	23.61	16.98	7.48	36.23	28.38	16.41
Sym. Semipar.	48.43	39.72	22.25	13.58	8.46	3.09	24.84	18.12	8.06	37.76	29.96	17.85
DGP Student $t(6)$												
Gaussian	47.24	39.14	20.45	14.43	9.05	3.77	24.88	17.61	6.42	33.99	26.77	15.98
Student t	48.08	39.80	23.81	14.66	8.82	3.61	26.84	20.16	9.62	39.72	32.50	21.14
DLSMN	39.93	30.39	13.08	14.70	8.89	3.15	22.68	15.40	4.81	33.14	25.58	15.26
Laplace	48.34	39.37	23.49	14.73	8.67	3.58	26.22	19.16	8.94	39.67	32.06	19.93
Semiparametric	37.35	28.93	12.35	14.32	9.05	3.47	23.06	16.41	5.52	36.86	29.38	17.79
Sym. Semipar.	41.27	32.26	14.94	14.76	9.04	3.67	24.09	17.41	7.04	38.45	30.84	18.93
DGP DLSMN(-0.85,0.16,0.05)												
Gaussian	48.23	39.90	21.50	15.42	9.96	4.11	24.98	17.93	7.07	35.99	28.36	16.71
Student t	51.24	43.92	26.53	17.00	11.09	4.70	28.59	22.07	11.70	44.27	37.02	24.62
DLSMN	42.29	32.75	13.82	16.89	10.44	4.36	24.52	16.45	5.36	38.06	30.34	18.93
Laplace	49.83	41.52	25.06	16.48	10.31	4.00	27.47	20.15	8.93	43.06	35.15	22.05
Semiparametric	41.48	32.33	14.90	16.96	11.19	4.70	26.65	19.27	7.83	44.11	36.27	23.31
Sym. Semipar.	44.80	35.77	17.50	17.88	11.70	5.15	27.43	20.24	8.77	45.31	37.23	24.27
DGP Gram-Charlier(0,3.0)												
Gaussian	41.94	33.40	12.72	12.55	7.54	2.61	21.14	14.47	5.21	28.90	21.73	11.39
Student t	45.40	37.21	21.00	14.46	8.60	3.02	24.88	18.06	7.70	42.71	35.52	22.84
DLSMN	42.22	33.06	14.86	15.40	9.41	3.59	24.55	16.73	5.50	43.53	35.34	21.54
Laplace	44.69	35.87	18.45	13.44	8.15	3.09	22.84	16.55	6.54	38.35	30.84	19.47
Semiparametric	41.03	31.44	12.54	17.24	10.98	4.25	24.09	17.01	4.75	43.20	34.73	21.70
Sym. Semipar.	44.86	35.56	14.12	17.16	11.00	4.41	25.83	18.14	5.82	45.33	36.23	22.84
DGP Gram-Charlier(-0.8,3.0)												
Gaussian	42.17	32.97	14.07	12.53	7.74	2.60	21.92	14.89	4.68	28.93	21.49	11.34
Student t	45.77	37.07	19.66	14.41	8.64	2.90	25.04	17.60	7.63	43.34	35.55	22.48
DLSMN	49.73	40.36	19.10	19.57	12.96	5.33	27.42	18.93	7.12	53.43	44.08	28.90
Laplace	44.34	35.37	16.91	13.51	8.01	2.99	23.22	16.01	6.07	37.23	29.68	17.75
Semiparametric	48.94	39.60	19.44	21.70	14.66	6.05	28.44	20.15	6.89	53.36	44.18	28.83
Sym. Semipar.	45.86	36.40	15.63	19.22	12.39	5.08	25.82	18.19	5.63	48.12	38.75	23.79

Monte Carlo empirical rejection rates of mean predictability tests. Sample length=100. Replications=20,000.

TABLE 1E: Monte Carlo power of predictability tests under DLSMN(-0.85,0.161,0.05)

True process	AR(1): $\rho = 0.06$						ARMA(1,1): $\rho = 0.98$ $\varphi = -0.96$					
Test against	AR(1)			AR(12)			AR(1)			AR(12)		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Gaussian	59.62	46.22	23.36	13.14	7.32	1.98	19.94	12.26	4.00	60.28	50.58	31.16
Student t	63.64	51.08	27.14	14.16	8.02	2.08	21.28	13.54	4.54	65.80	55.60	35.96
DLSMN	64.70	51.92	28.00	14.02	7.74	2.12	21.54	13.52	4.66	66.34	56.34	35.90
Laplace	50.28	37.42	16.96	12.96	7.14	1.62	18.12	10.68	3.14	55.10	44.40	25.10
Semiparametric	63.12	50.06	26.62	14.00	8.20	2.38	21.04	13.20	4.50	64.60	54.94	35.42
Sym. Semipar.	63.70	50.90	26.88	14.12	7.98	2.08	21.36	13.26	4.56	65.32	55.50	35.40

True process	ARCH(1): $\alpha = 0.08$						GARCH(1,1): $\alpha = 0.04$ $\beta = 0.88$					
Test against	ARCH(1)			GARCH(1,1)			ARCH(1)			GARCH(1,1)		
Nominal size	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Gaussian	56.74	47.90	33.42	27.74	19.82	10.02	32.26	24.88	14.92	73.68	64.64	46.48
Student t	72.30	64.22	47.22	36.66	26.94	12.46	47.26	37.40	22.56	88.46	83.74	69.66
DLSMN	71.98	63.68	46.22	37.02	27.14	12.56	47.42	38.30	23.30	89.70	85.26	72.18
Laplace	69.68	60.96	43.72	34.96	24.82	11.76	43.82	35.06	20.82	86.60	81.08	65.42
Semiparametric	63.12	55.02	38.16	32.78	23.80	10.60	42.34	33.54	20.02	86.10	79.82	65.28
Sym. Semipar.	65.30	56.38	39.80	34.16	24.70	11.02	43.44	34.00	20.30	87.06	80.88	66.26

Monte Carlo empirical rejection rates of mean variance predictability tests. Sample length=1,000. Replications=5,000.

TABLE 2: Descriptive statistics.

Region	Factor	Mean	Std.dev.	0.74× IQR	Skewness	Kurtosis
North America	Market	2.12	7.98	5.57	**-0.64**	**3.83**
	SMB	0.53	4.36	3.77	0.13	3.01
	HML	0.46	6.59	4.42	0.99	**7.13**
	RMW	0.98	4.33	3.42	1.14	**8.05**
	CMA	0.68	5.47	4.12	1.37	**7.68**
Europe	Market	1.57	9.29	6.68	-0.32	**3.94**
	SMB	0.27	3.83	3.75	-0.03	3.44
	HML	0.95	5.21	3.77	0.43	**6.31**
	RMW	1.14	2.92	2.85	0.04	3.34
	CMA	0.52	3.88	2.10	0.37	**5.82**
Japan	Market	0.36	10.63	10.40	0.01	2.83
	SMB	0.43	6.02	5.17	-0.49	**5.01**
	HML	0.86	5.79	4.09	-0.59	**8.61**
	RMW	0.39	3.97	3.58	0.33	**4.83**
	CMA	0.14	4.88	3.26	-1.59	**12.39**
Asia Pacific ex Japan	Market	2.19	11.19	8.47	0.02	**4.13**
	SMB	-0.15	6.31	4.40	1.46	**8.40**
	HML	1.90	5.66	3.54	1.85	**10.38**
	RMW	0.58	4.91	4.03	0.01	**3.85**
	CMA	1.01	4.38	3.04	-0.44	**6.68**

Sample: 1990Q3-2018Q3. Boldface figure means statistically different from its value under symmetry and mesokurtosis at the 5% level. IQR denotes the Interquartile Range, which under normality equals the standard deviation divided by .74

TABLE 3A: North America Fama and French 5 factors.

Test against	AR(1)	AR(4)	AR(12)	ARCH(1)	GARCH(1,1)
Market Portfolio: *Rm-Rf*					
Gaussian	0.62	0.81	0.82	**0.01**	**0.00**
Student t	0.63	0.67	0.30	**0.00**	**0.00**
DLSMN	0.30	0.57	0.30	**0.01**	**0.00**
Laplace	0.46	0.56	0.30	**0.00**	**0.00**
Semiparametric	0.17	0.48	0.18	**0.01**	**0.00**
Sym. Semipar.	0.56	0.93	0.48	**0.01**	**0.00**
SMB: *small minus big*					
Gaussian	0.20	0.27	0.38	**0.01**	**0.00**
Student t	0.19	0.27	0.38	**0.01**	**0.00**
DLSMN	0.22	0.10	0.27	**0.01**	**0.00**
Laplace	0.16	0.69	0.88	**0.01**	**0.00**
Semiparametric	0.09	0.19	0.42	**0.00**	**0.00**
Sym. Semipar.	0.14	0.26	0.53	**0.01**	**0.00**
HML *high minus low*					
Gaussian	**0.00**	0.74	0.39	**0.00**	**0.00**
Student t	0.69	0.96	0.80	**0.00**	**0.00**
DLSMN	0.84	0.92	0.67	**0.01**	**0.00**
Laplace	0.85	0.94	0.88	**0.00**	**0.00**
Semiparametric	0.67	0.98	0.76	**0.00**	**0.00**
Sym. Semipar.	0.89	0.91	0.74	**0.03**	**0.00**
RMW *robust minus weak*					
Gaussian	0.21	0.68	0.09	**0.05**	**0.00**
Student t	0.97	0.68	0.09	**0.00**	**0.00**
DLSMN	0.72	0.25	**0.02**	**0.03**	**0.00**
Laplace	0.92	0.86	0.16	**0.01**	**0.00**
Semiparametric	0.96	0.78	**0.05**	**0.00**	**0.00**
Sym. Semipar.	0.96	0.95	0.11	**0.00**	**0.00**
CMA *conservative minus aggresive*					
Gaussian	0.34	0.78	0.78	**0.00**	**0.00**
Student t	0.90	0.46	0.80	**0.00**	**0.00**
DLSMN	0.50	0.10	0.95	**0.01**	**0.00**
Laplace	0.59	0.86	0.52	**0.00**	**0.00**
Semiparametric	0.21	0.15	0.95	**0.01**	**0.00**
Sym. Semipar.	0.46	0.46	0.87	**0.02**	**0.00**

P-values of mean variance predictability tests based on 20,000 bootstrap samples.

TABLE 3B: EUROPE Fama and French 5 factors.

Test against	AR(1)	AR(4)	AR(12)	ARCH(1)	GARCH(1,1)
Market Portfolio: *Rm-Rf*					
Gaussian	0.39	0.78	0.35	**0.02**	**0.00**
Student t	0.58	0.79	0.42	**0.04**	**0.00**
DLSMN	0.91	0.16	0.69	0.09	**0.02**
Laplace	0.42	0.95	0.26	**0.02**	**0.00**
Semiparametric	0.87	0.41	0.29	**0.04**	**0.00**
Sym. Semipar.	0.67	0.73	0.55	0.06	**0.01**
SMB: *small minus big*					
Gaussian	0.43	0.37	0.60	0.34	0.12
Student t	0.41	0.39	0.59	0.31	0.06
DLSMN	0.40	0.40	0.59	0.28	**0.05**
Laplace	0.60	0.74	0.82	0.25	**0.03**
Semiparametric	0.48	0.61	0.71	0.15	**0.01**
Sym. Semipar.	0.54	0.54	0.61	0.18	**0.01**
HML *high minus low*					
Gaussian	**0.00**	**0.04**	0.41	**0.01**	**0.00**
Student t	**0.01**	**0.01**	**0.03**	**0.00**	**0.00**
DLSMN	**0.01**	**0.01**	**0.03**	**0.01**	**0.00**
Laplace	**0.01**	**0.01**	**0.01**	**0.00**	**0.00**
Semiparametric	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**
Sym. Semipar.	**0.01**	**0.01**	0.06	**0.00**	**0.00**
RMW *robust minus weak*					
Gaussian	0.60	0.54	0.22	0.13	0.16
Student t	0.83	0.36	0.26	0.14	0.20
DLSMN	0.84	0.15	0.73	0.46	0.37
Laplace	0.70	0.12	0.41	0.13	0.26
Semiparametric	0.80	0.41	0.47	0.33	0.21
Sym. Semipar.	0.92	0.25	0.50	0.38	0.24
CMA *conservative minus aggresive*					
Gaussian	**0.02**	0.14	0.94	**0.00**	**0.00**
Student t	0.16	0.34	0.69	**0.00**	**0.00**
DLSMN	0.06	0.36	0.66	**0.01**	**0.00**
Laplace	0.08	0.28	0.73	**0.00**	**0.00**
Semiparametric	0.21	0.56	0.64	**0.00**	**0.00**
Sym. Semipar.	0.23	0.34	0.67	**0.00**	**0.00**

P-values of mean variance predictability tests based on 20,000 bootstrap samples.

TABLE 3C: JAPAN Fama and French 5 factors.

Test against	AR(1)	AR(4)	AR(12)	ARCH(1)	GARCH(1,1)
Market Portfolio: *Rm-Rf*					
Gaussian	0.17	0.40	0.12	**0.01**	**0.00**
Student t	0.17	0.40	0.12	**0.01**	**0.00**
DLSMN	0.17	0.39	0.11	**0.01**	**0.00**
Laplace	0.33	0.84	0.30	**0.01**	**0.00**
Semiparametric	0.08	0.43	0.24	**0.01**	**0.00**
Sym. Semipar.	0.08	0.48	0.14	**0.01**	**0.00**
SMB: *small minus big*					
Gaussian	0.56	0.27	0.40	0.19	**0.01**
Student t	0.83	0.16	0.59	0.08	**0.01**
DLSMN	0.90	0.53	0.94	0.13	**0.01**
Laplace	0.85	0.06	0.61	0.08	**0.00**
Semiparametric	0.81	0.18	0.84	0.07	**0.02**
Sym. Semipar.	0.77	0.10	0.60	0.10	**0.04**
HML *high minus low*					
Gaussian	**0.05**	0.72	0.75	0.31	**0.00**
Student t	0.27	0.95	0.44	0.35	**0.00**
DLSMN	0.23	0.61	0.31	0.39	**0.01**
Laplace	0.35	1.00	0.79	0.33	**0.00**
Semiparametric	0.32	0.55	0.33	0.33	**0.02**
Sym. Semipar.	0.20	0.37	0.36	0.32	**0.01**
RMW *robust minus weak*					
Gaussian	0.61	0.84	0.41	0.37	**0.00**
Student t	0.86	0.54	0.77	0.41	**0.00**
DLSMN	0.81	0.67	0.77	0.39	**0.00**
Laplace	0.70	0.28	0.70	0.36	**0.00**
Semiparametric	0.70	0.66	0.54	0.36	**0.00**
Sym. Semipar.	0.80	0.72	0.70	0.36	**0.01**
CMA *conservative minus aggresive*					
Gaussian	**0.02**	0.46	0.62	0.62	0.10
Student t	0.57	0.63	0.29	0.29	0.08
DLSMN	0.68	0.75	0.52	0.52	0.10
Laplace	0.33	0.93	0.23	0.23	0.07
Semiparametric	0.35	0.94	0.61	0.61	0.12
Sym. Semipar.	0.45	0.67	0.41	0.42	0.13

P-values of mean variance predictability tests based on 20,000 bootstrap samples.

TABLE 3D: Asia Pacific ex Japan Fama and French 5 factors.

Test against	AR(1)	AR(4)	AR(12)	ARCH(1)	GARCH(1,1)
Market Portfolio: *Rm-Rf*					
Gaussian	0.44	0.41	0.45	0.09	**0.02**
Student t	0.67	0.45	0.79	**0.05**	**0.01**
DLSMN	0.50	0.62	0.88	0.09	**0.04**
Laplace	0.61	0.77	0.93	**0.05**	**0.01**
Semiparametric	0.52	0.44	0.67	**0.01**	**0.00**
Sym. Semipar.	0.67	0.40	0.84	0.07	**0.01**
SMB: *small minus big*					
Gaussian	0.38	0.07	**0.01**	0.33	0.25
Student t	0.92	0.44	0.08	0.42	0.07
DLSMN	0.72	0.19	0.06	0.35	0.08
Laplace	0.69	0.42	**0.05**	0.40	0.13
Semiparametric	0.71	0.21	0.07	0.30	0.06
Sym. Semipar.	0.88	0.36	0.08	0.41	0.09
HML *high minus low*					
Gaussian	**0.02**	**0.02**	0.76	0.08	**0.00**
Student t	0.07	**0.02**	0.78	**0.01**	**0.00**
DLSMN	**0.03**	**0.01**	0.78	**0.01**	**0.01**
Laplace	0.14	**0.04**	0.67	**0.05**	**0.00**
Semiparametric	0.13	**0.03**	0.87	0.08	0.07
Sym. Semipar.	0.09	**0.02**	0.73	0.07	**0.02**
RMW *robust minus weak*					
Gaussian	0.07	0.93	0.65	0.25	**0.00**
Student t	**0.03**	0.92	0.92	0.14	**0.00**
DLSMN	**0.03**	0.76	0.92	0.14	**0.00**
Laplace	**0.04**	0.54	0.78	0.12	**0.00**
Semiparametric	0.06	0.73	0.63	0.34	**0.00**
Sym. Semipar.	**0.05**	0.82	0.78	0.26	**0.00**
CMA *conservative minus aggresive*					
Gaussian	0.84	0.22	0.29	**0.01**	**0.00**
Student t	0.65	0.44	0.48	**0.00**	**0.00**
DLSMN	0.59	0.50	0.53	**0.04**	**0.04**
Laplace	0.66	0.56	0.60	**0.00**	**0.00**
Semiparametric	0.39	0.46	0.36	**0.02**	0.13
Sym. Semipar.	0.79	0.46	0.47	**0.01**	**0.03**

P-values of mean variance predictability tests based on 20,000 bootstrap samples.

FIGURE 1: Tests of predictability in mean

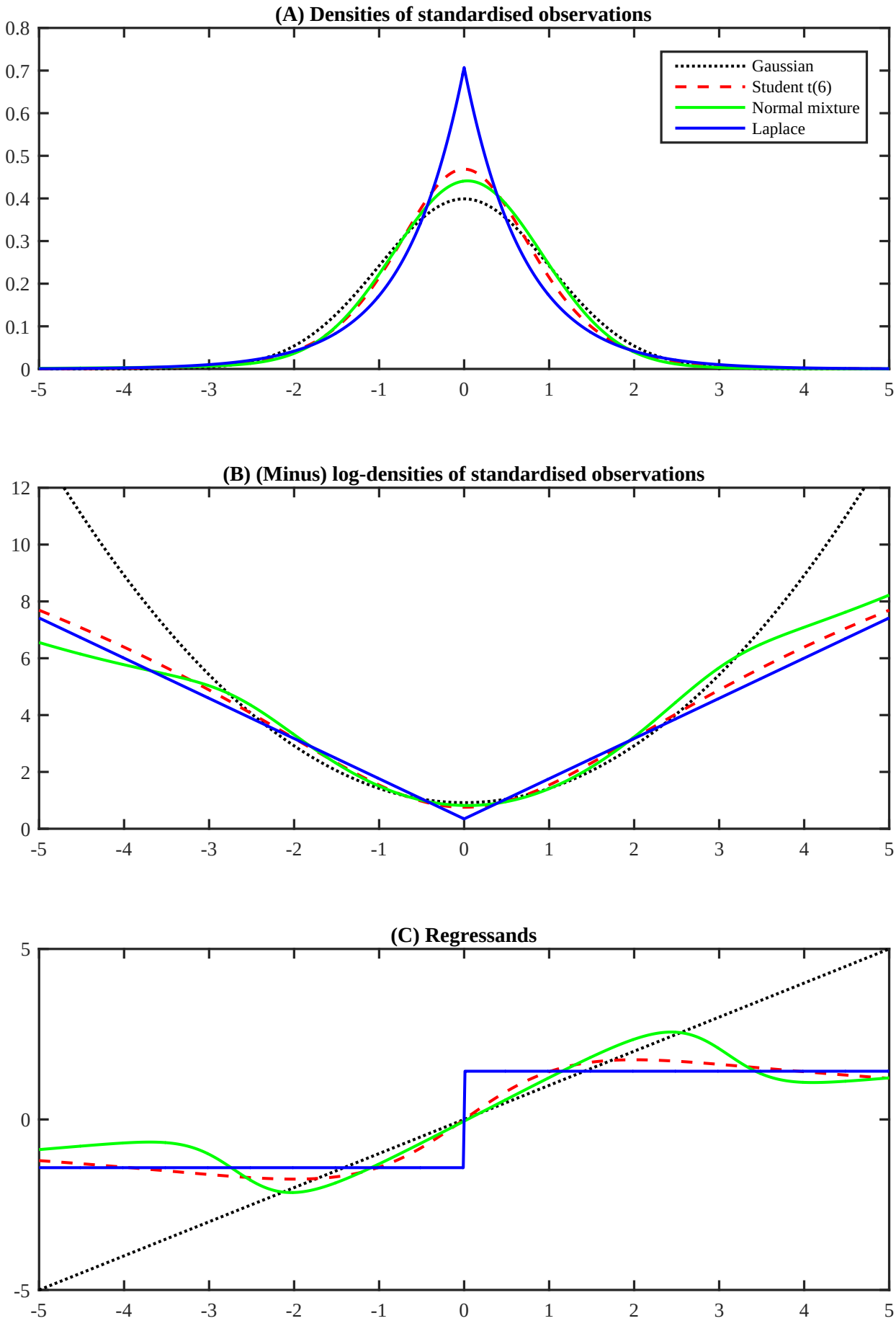


FIGURE 2: ACF of expected and observed returns ($h = 24, \rho = .015$)

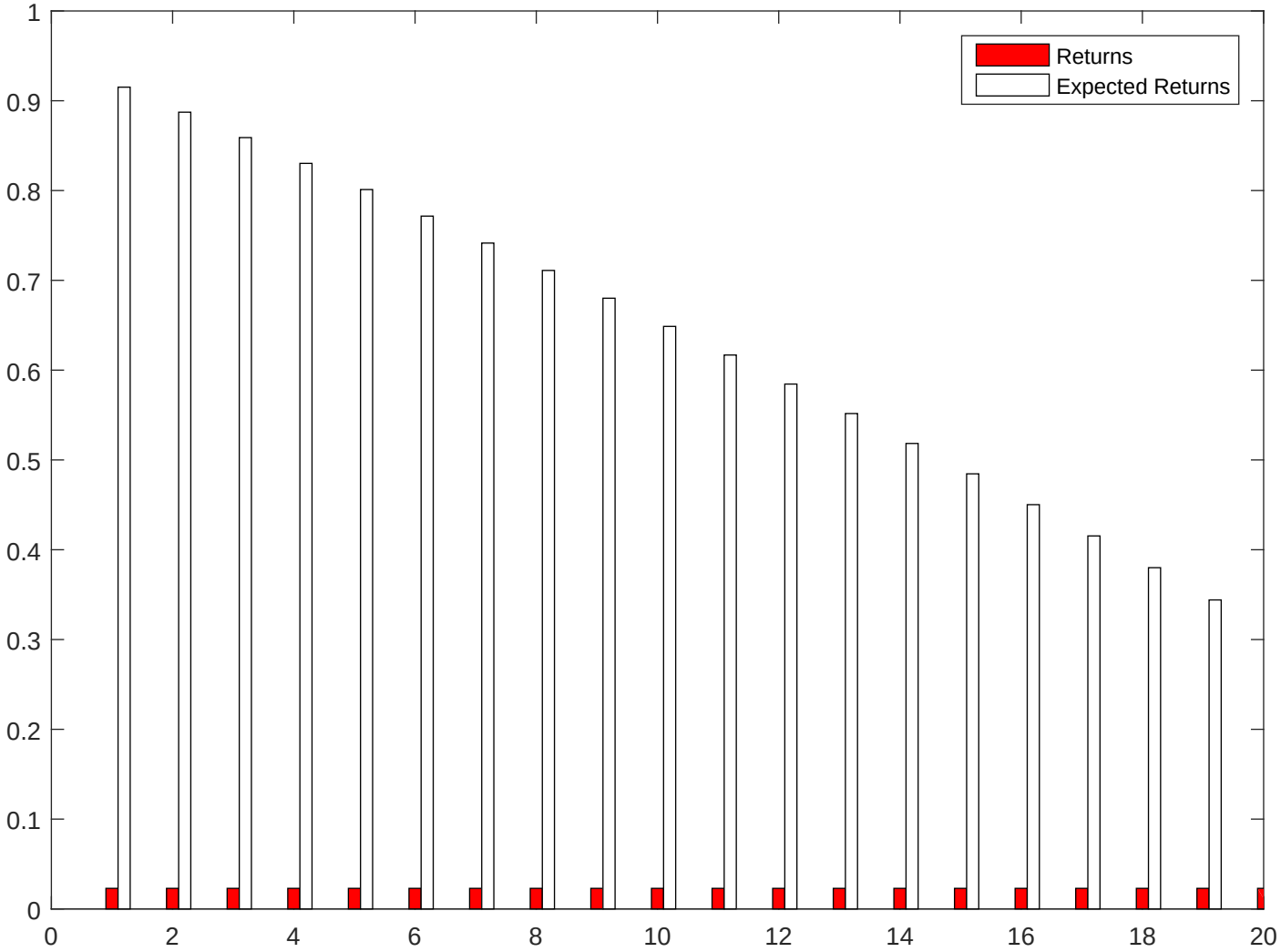
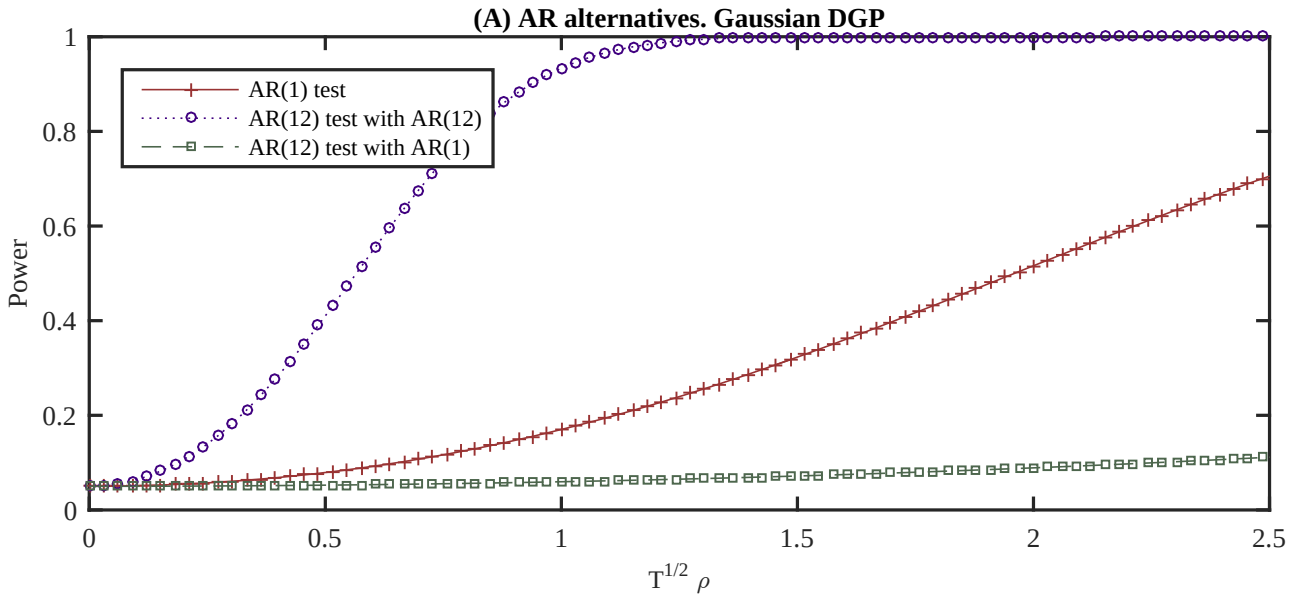
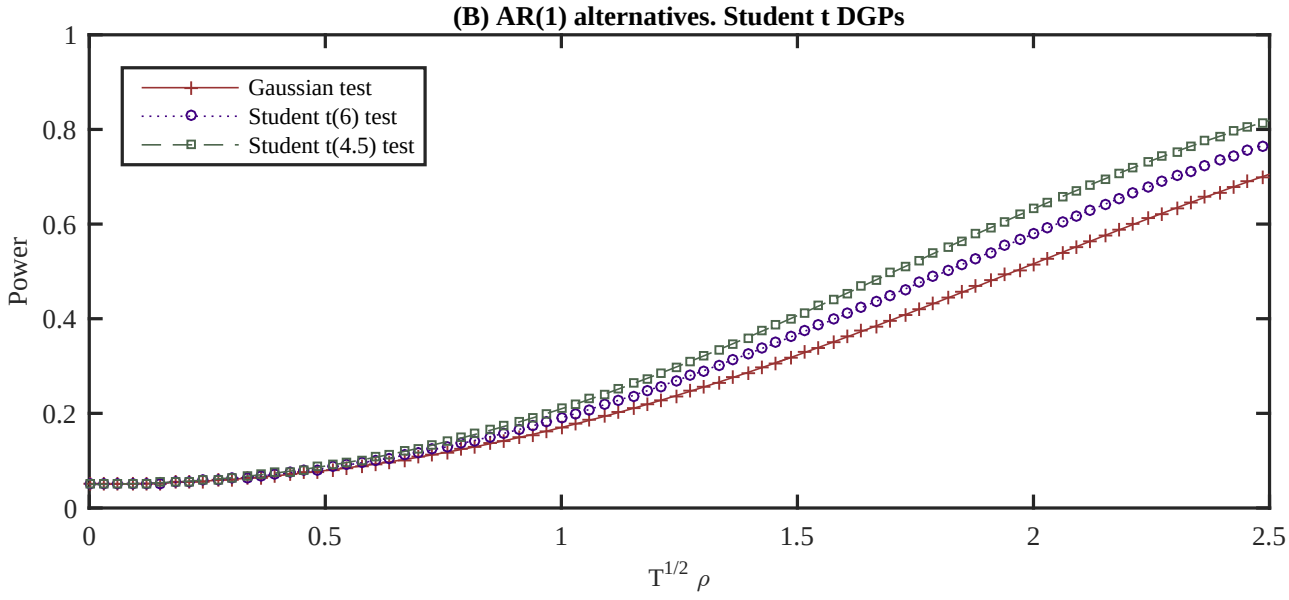


FIGURE 3: Local power of unpredictability in mean tests at 5% level





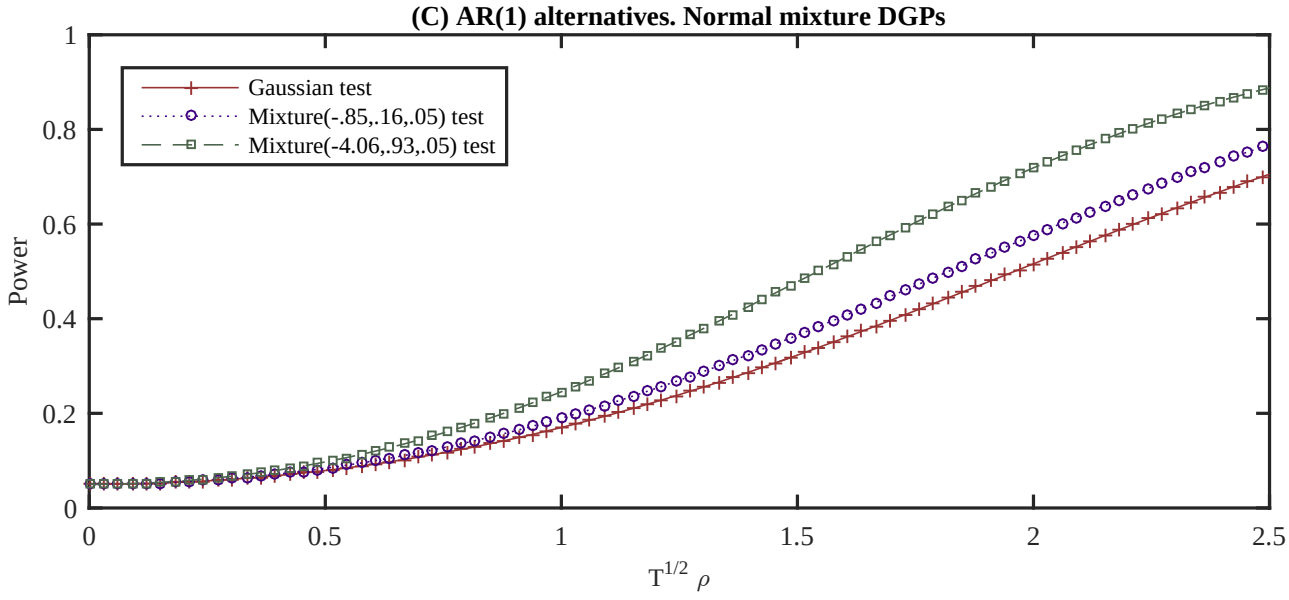


FIGURE 4A: Pitman's asymptotic relative efficiency of mean tests
Gaussian / Student t when DGP is GC(c_3, c_4)

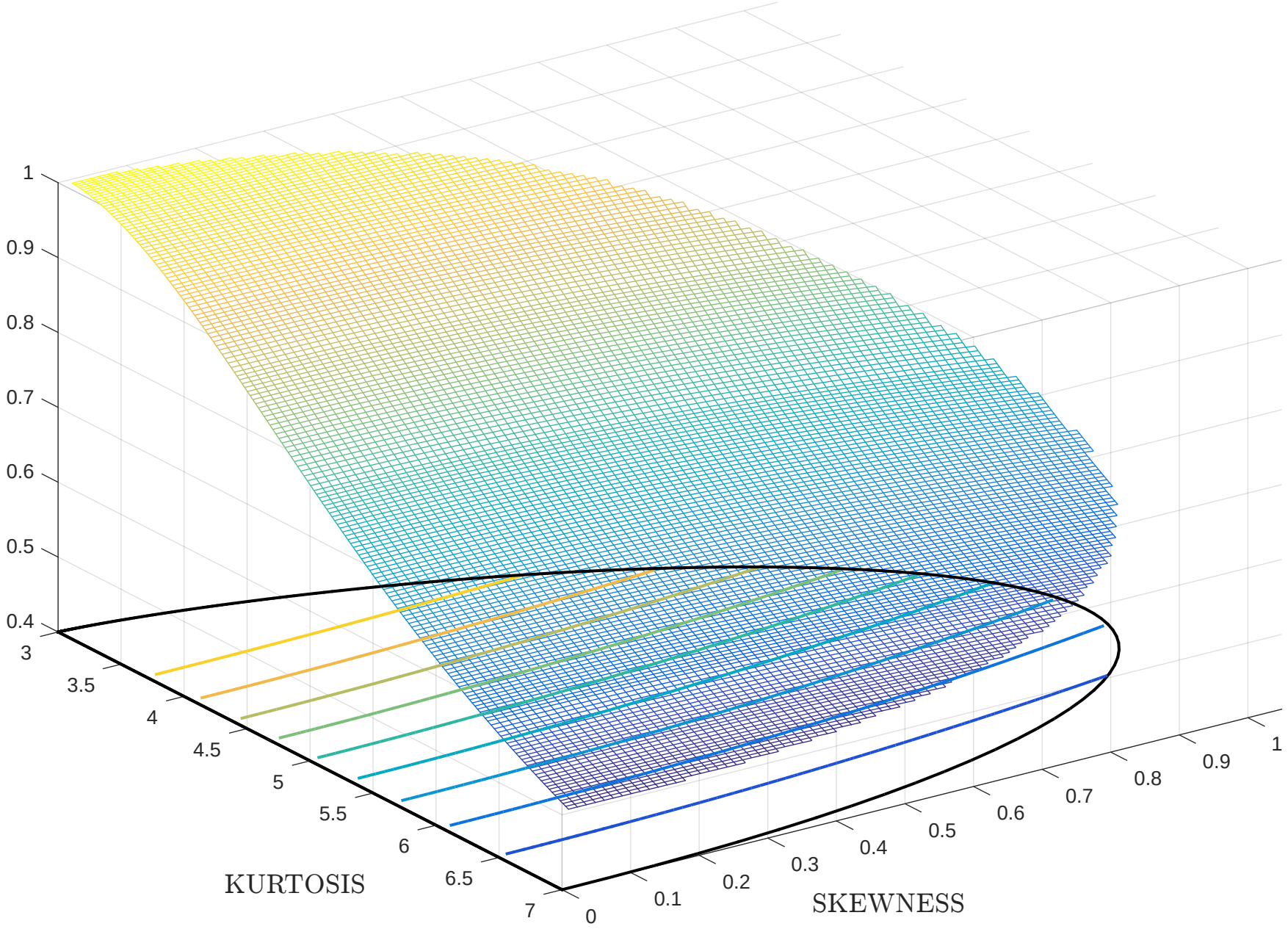


FIGURE 4B: Pitman's asymptotic relative efficiency of mean tests
Gaussian / Student t when DGP is DLSMN$(\delta, \upsilon, \lambda)$

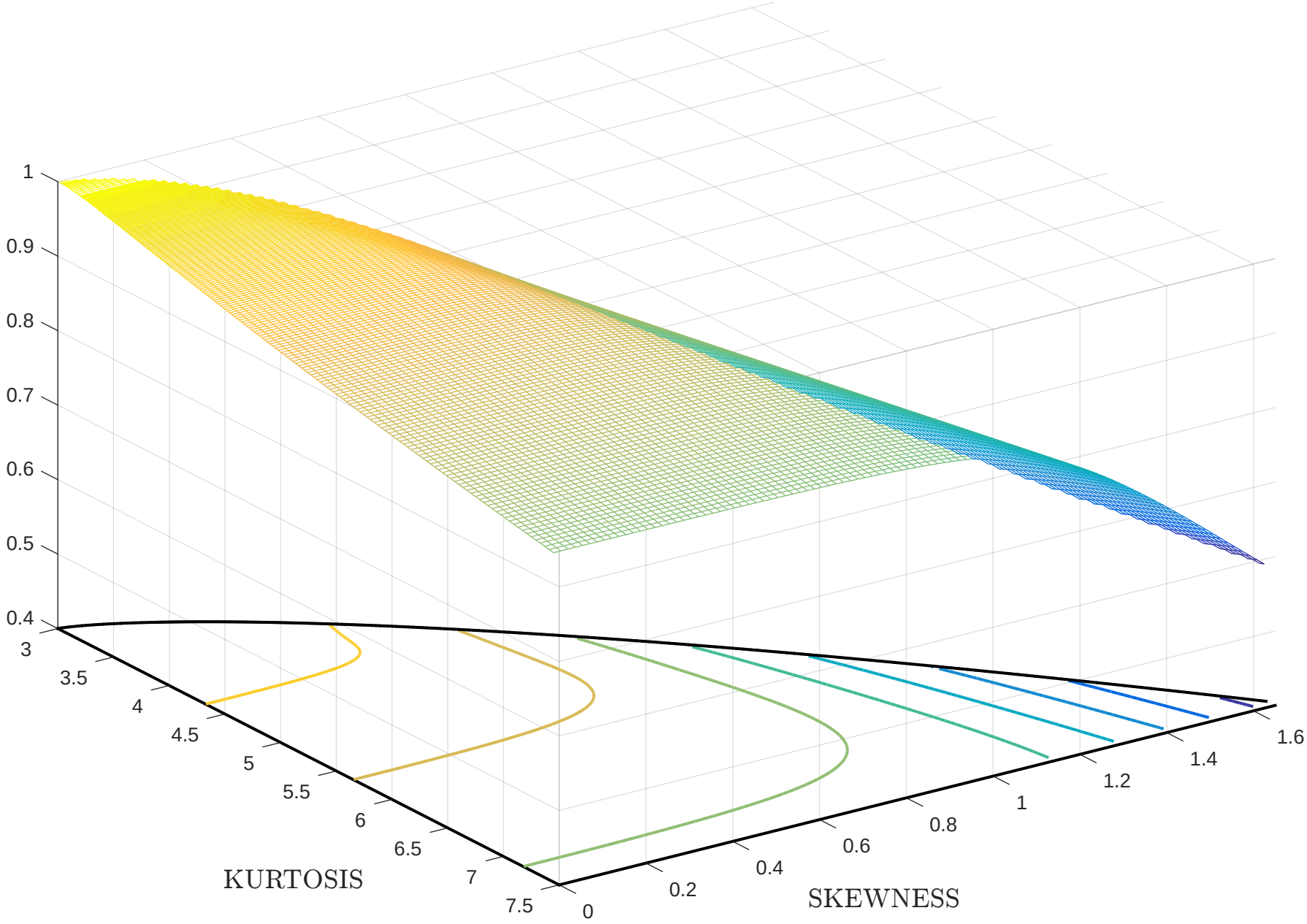


FIGURE 5: Tests of predictability in variance

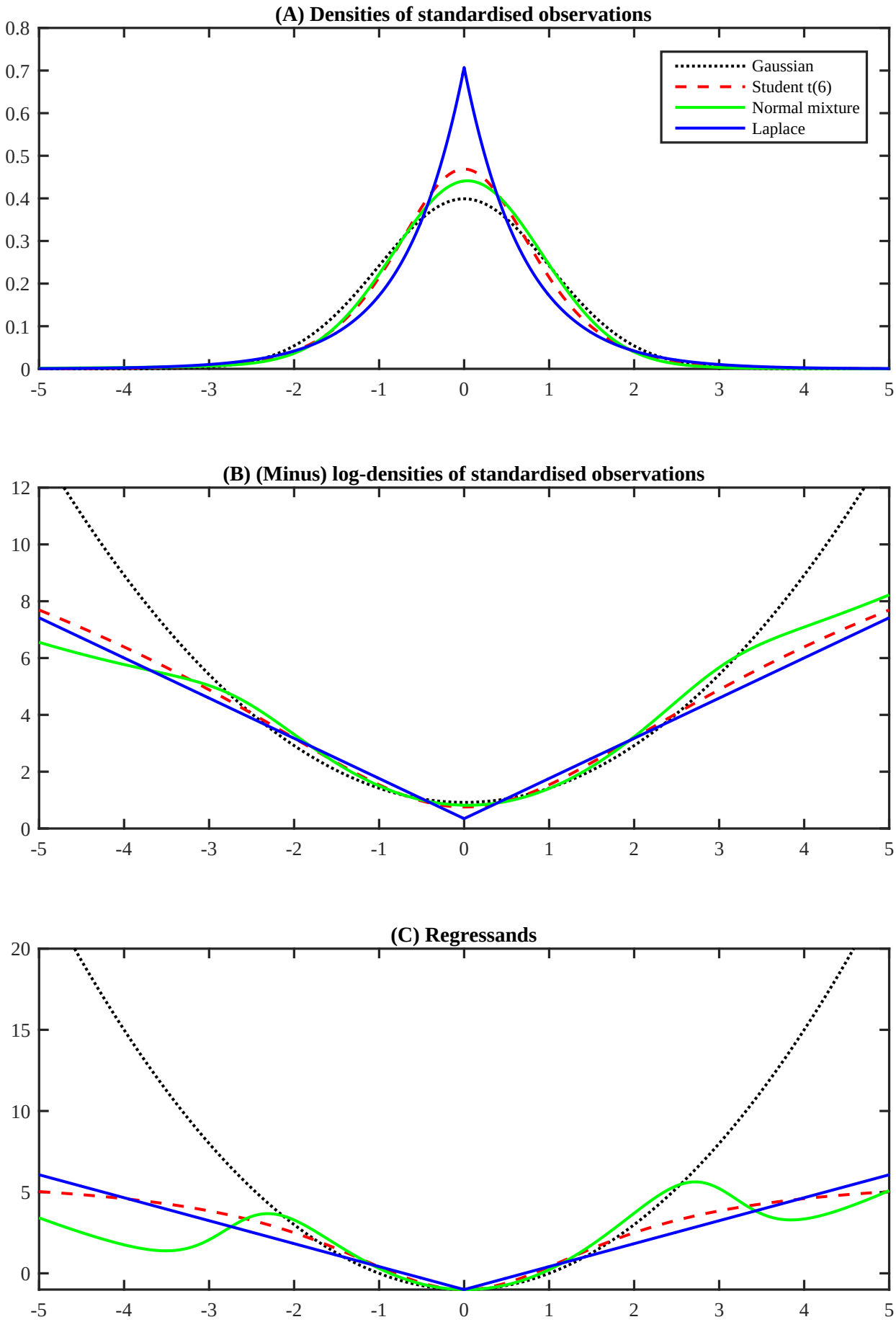
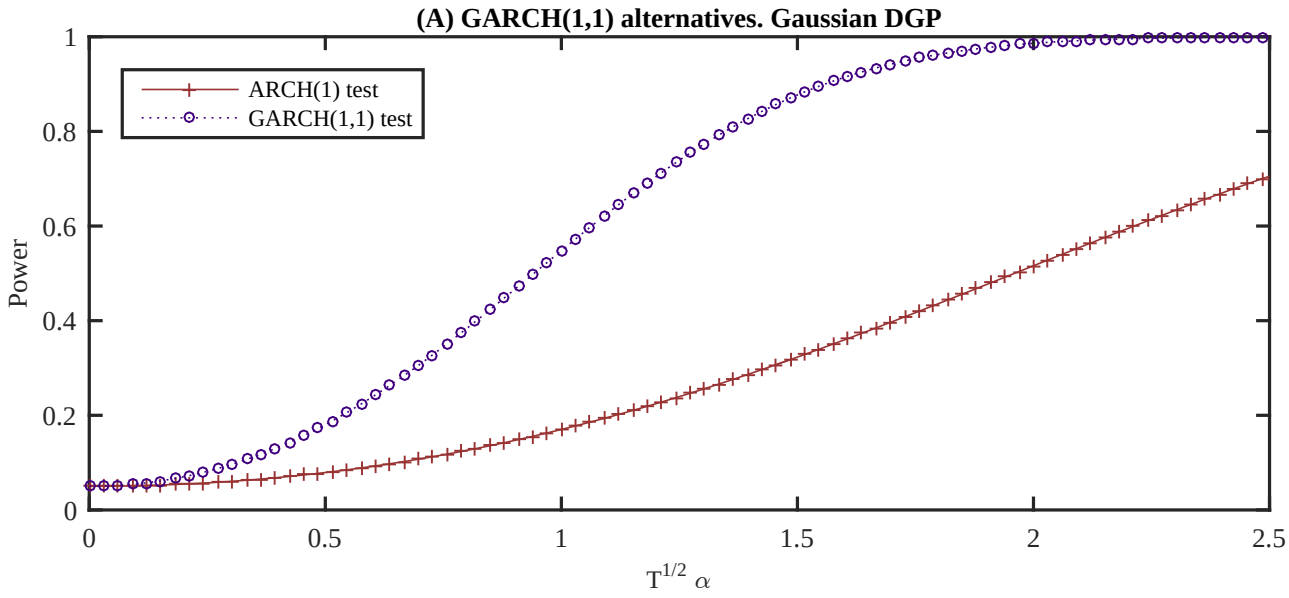
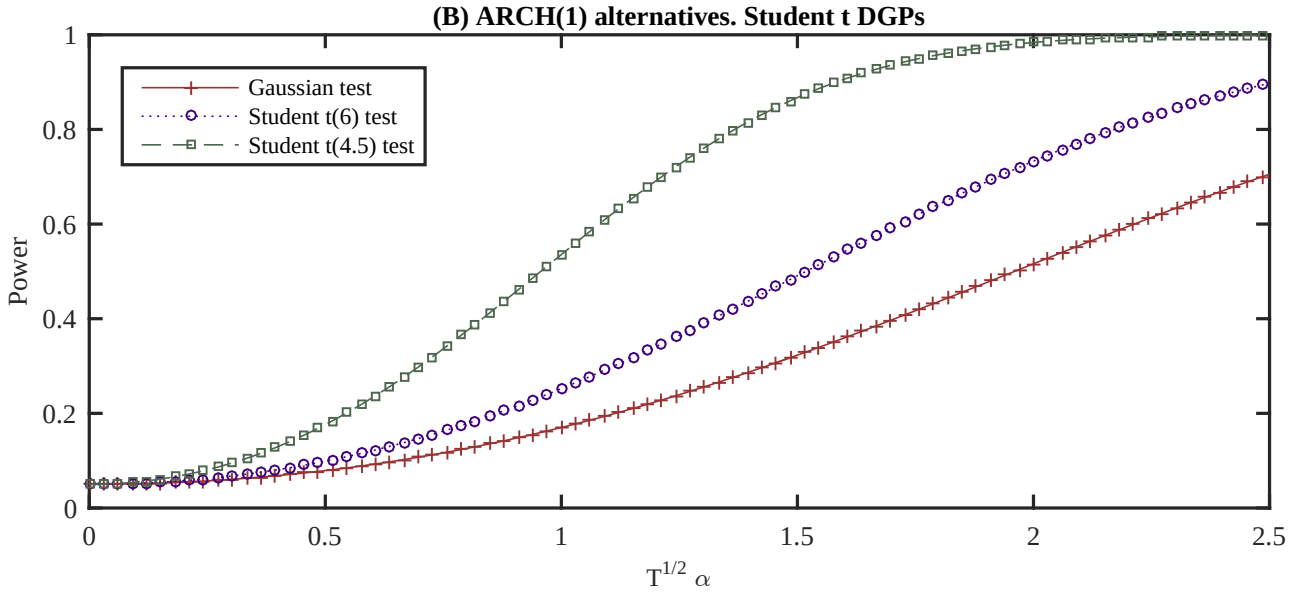


FIGURE 6: Local power of unpredictability in variance tests at 5% level





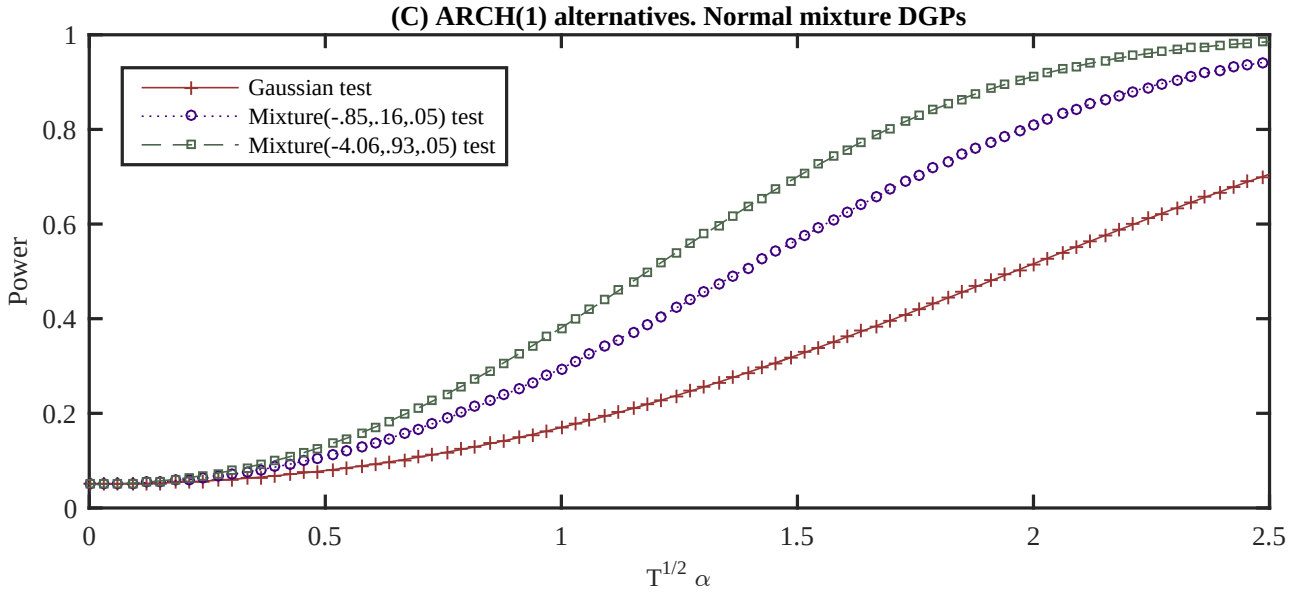


FIGURE 7A: Pitman's asymptotic relative efficiency of variance tests
Gaussian / Student t when DGP is GC(c_3, c_4)

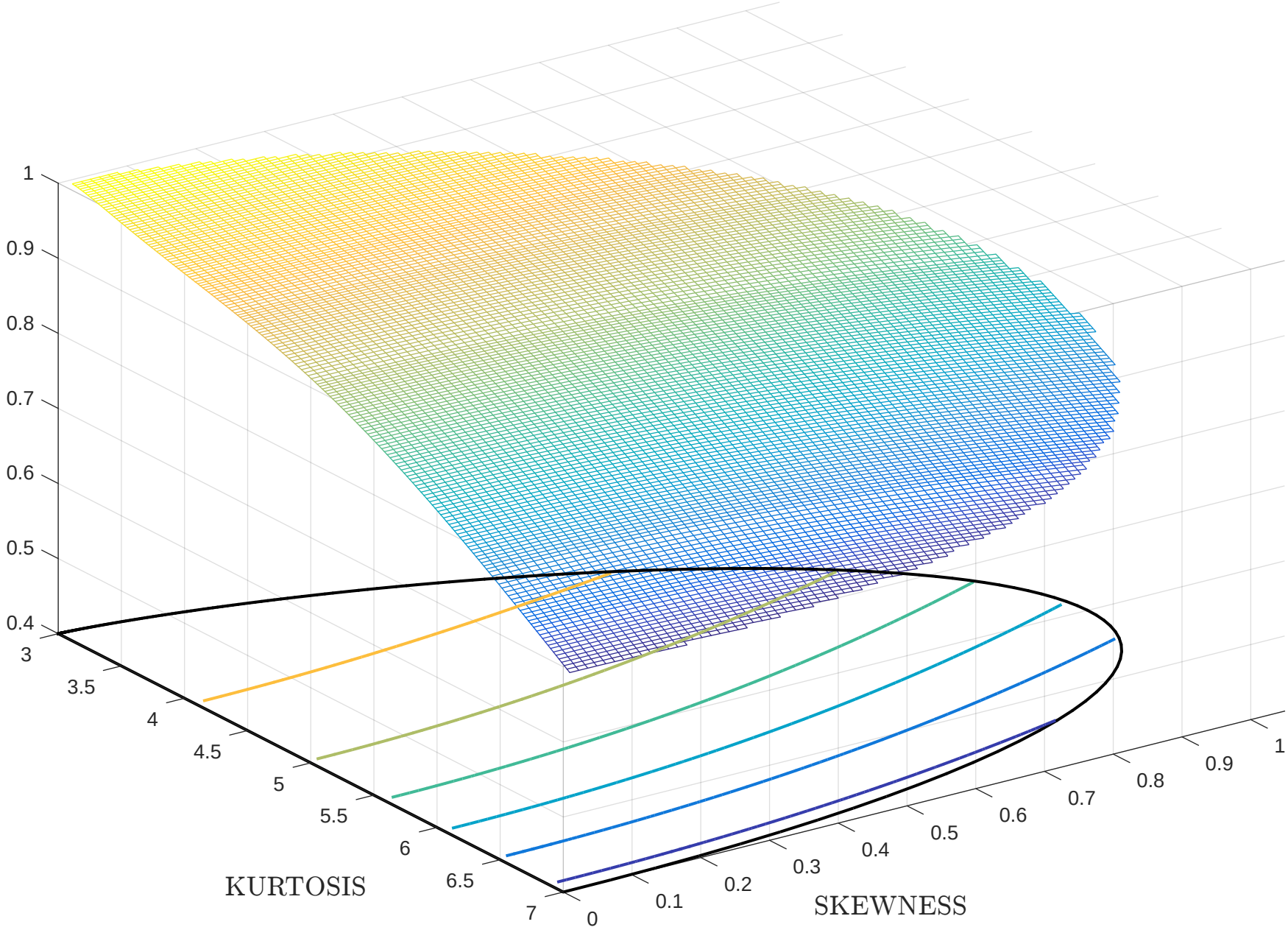


FIGURE 7B: Pitman's asymptotic relative efficiency of variance tests
Gaussian / Student t when DGP is DLSMN$(\delta, \upsilon, \lambda)$

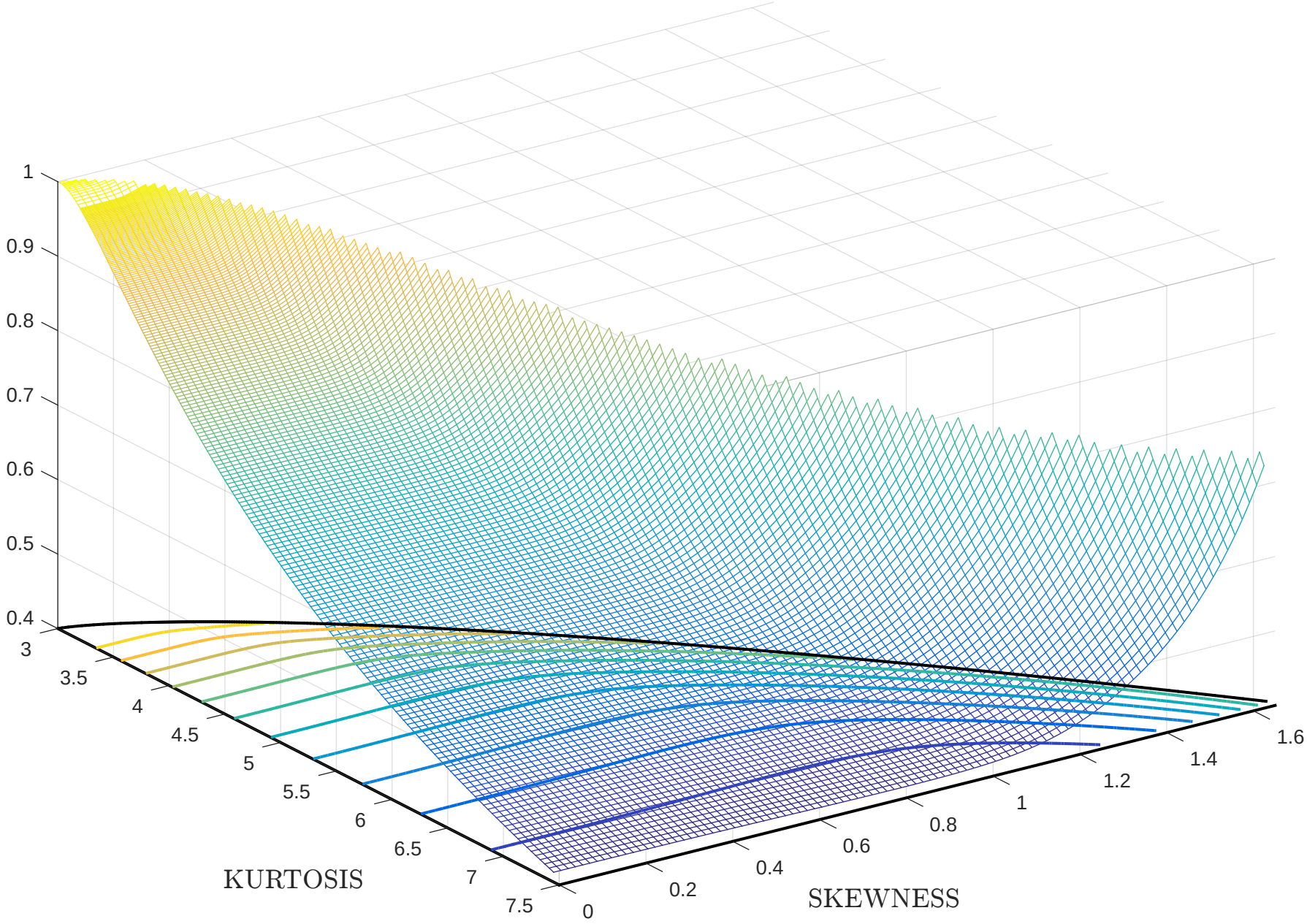


FIGURE 8: Fama and French 5 factor portfolios and robust confidence bands