New testing approaches for mean-variance predictability*

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

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

January 2010
Preliminary and incomplete

Abstract

We propose tests for smooth but persistent serial correlation in risk premia and volatilities that exploit the non-normality of financial asset returns. Our parametric tests are robust to distributional misspecification, while our nonparametric tests are as powerful as if we knew the true distribution of excess returns. Local power analyses confirm their gains over existing methods, while Monte Carlo exercises document their finite sample reliability. We apply our methods to the Fama-French factors for US stocks. We find mean predictability for the size and value factors but not the market, and variance predictability for all of them.

Keywords: ARCH, Financial returns, LM tests, Normal Mixtures, Robustness.

JEL: C13, C12, C14, C16, G17

*We would like to thank Dante Amengual, Christian Bontemps, Nour Meddahi and Javier Mencía, as well as participants at the Toulouse Financial Econometrics Conference (May 2009) 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 (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 three decades, the evidence remains controversial (see Spiegel (2008) and the references therein). There is 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 frequencies. Finally, many empirical studies indicate that the distribution of asset returns is rather leptokurtic, and possibly somewhat asymmetric. Nevertheless, most existing tests for predictability of the mean and volatility of asset returns ignore this fact by implicitly relying on normality. Similarly, 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.

In this context, we propose new testing approaches for mean-variance predictability that explicitly account for 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 non-parametric tests. For a given predictor variable, our tests differ from standard tests in that we effectively change the regressand. But we also transform the predictor variable 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.

Importantly, we 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 Bollerslev and Wooldridge (1992) among many others. We also show that our non-parametric 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. Finally, we also illustrate our methods with an application to the three Fama and French (1993) factors for US stocks.

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. A Monte Carlo evaluation of our proposed procedures can be found in section 4, followed by our empirical application in section 5. Finally, we present our conclusions in section 6. 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

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

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

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.*

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

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

which reduces to ε^* under normality. In contrast, for a standardised Laplace distribution, which does not depend on any additional parameters, we have that

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

which means that (2) effectively becomes a directional prediction test in that case.

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, $\boldsymbol{\eta} = (\delta, \varkappa, \lambda)'$ 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

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

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 $-\partial^2 \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]/\partial\varepsilon^*\partial\varepsilon^*$. As shown by Davidson and MacKinnon (1983) and many others, this choice will affect the finite sample properties of the tests.

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

$$E\left\{\frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}); \boldsymbol{\eta}_0]}{\partial \varepsilon^*}\epsilon_{t-1}(\boldsymbol{\theta}_{s0})\middle| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0\right\} = 0, \qquad (4)$$

which is related to the moment condition used by Bontemps and Meddahi (2007) in their distributional tests.[1] In fact, given that the score with respect to π under the null is proportional to

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

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 $\epsilon_{t-1}(\boldsymbol{\theta}_s)$.

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

Our regressand can be regarded as $\epsilon_t(\boldsymbol{\theta}_s)$ times a damping factor that accounts for skewness and kurtosis, as in the robust estimation literature. Figures 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 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.[2] In this sense, the Student t factor $(\eta+1)/[1-2\eta+\eta\varepsilon_t^{*2}]$ clearly downweights big observations because it is a decreasing function of ε_t^{*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. A notable exception is a discrete mixture of normals, since we prove in Appendix D.5 that the ML estimators of π and ω coincide with the Gaussian ones.

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. $D(0,1,\boldsymbol{\varrho}_0)$, 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 $H_0:\rho=0$.*

But it is important to emphasise that the orthogonality condition (4) underlying our proposed Ar 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 would 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

[2] 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.

uncertainty in those pseudo-MLE 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.

Importantly, a test based on (4) will have non-trivial power even when it will no longer be an LM test. In fact, the evidence presented in Amengual and Sentana (2010) suggests that our proposed tests could be more powerful than the usual regression-based tests in Proposition 2 even though the parametric distribution is misspecified.

In any case, though, the test proposed in Proposition 1 requires to specify a parametric distribution. Given that some researchers might be reluctant to do so, we next consider semi-parametric tests that do not make any specific assumptions about the conditional distribution of the standardised innovations ε_t^*, as in Gonzalez-Rivera and Ullah (2001). There are two possibilities: unrestricted nonparametric density estimates (SP) and nonparametric density estimates that impose symmetry (SSP). It turns out that 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:

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,*

$$\bar{\pi} = \frac{1}{T}\sum\nolimits_{t=1}^{T} y_t \tag{5}$$

and

$$\bar{\omega} = \frac{1}{T}\sum\nolimits_{t=1}^{T} (y_t - \bar{\pi})^2, \tag{6}$$

which coincide with the sample mean and variance of y_t.

2. *The same result is true 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,*

$$\dot{\pi} = \bar{\pi} + \frac{1}{\sqrt{\bar{\omega}}}\left\{\sum\nolimits_{t=1}^{T}\left[\frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial\varepsilon^*}\right]^2\right\}^{-1}\left[\sum\nolimits_{t=1}^{T}\frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial\varepsilon^*}\right] \tag{7}$$

and

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

In addition, the semiparametric predictability test is adaptive, in the sense that it is as powerful as if we knew the distribution of ε_t^*, including the true values of the shape parameters $\boldsymbol{\eta}$. Similarly, the symmetric semiparametric test will be adaptive too if the true distribution is symmetric. In finite samples, though, the power of these semiparametric procedures may not be well approximated by the first-order asymptotic theory that justifies their adaptivity.

2.2 Exploiting the persistence of expected returns

Let us now consider a situation in which

$$y_t = \pi(1 - \sum_{l=1}^{h} \rho_l) + \sum_{l=1}^{h} \rho_l y_{t-l} + \sqrt{\omega}\varepsilon_t^*,$$

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

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

Given that under the null hypothesis y_t is independent and identically distributed, 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

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

for $l = 1, \ldots, h$, whose asymptotic distribution would be a χ_h^2 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.

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 ARMA$(h, h-1)$ process:

$$\mu_{t+1} = \pi(1 - h\rho) + \sum_{j=1}^{h} \rho\mu_{t+1-j} + \rho\left[\varepsilon_t + \sum_{j=1}^{h-1} \varepsilon_{t-j}\right]. \tag{9}$$

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

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

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 $\rho = .015$.

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 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

$$E\left\{ \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \sum_{l=1}^{h} \epsilon_{t-l}(\boldsymbol{\theta}_s) \right| \left. \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0 \right\} = 0, \tag{10}$$

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. The asymptotic variance is analogous to the so-called Hodrick (1992) standard errors used in LM tests for long run return predictability in univariate OLS regressions with overlapping regressands.

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 (9), the non-centrality parameter of the Gaussian pseudo-score test based on the first order serial correlation coefficient 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, the 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 (10) (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 $\boldsymbol{\eta}_0$. If in addition the conditional distribution is symmetric and $\kappa_0 < \infty$, then the elliptically symmetric semiparametric estimator of ρ is also fully efficient. The same is true in general of the semiparametric estimator of ρ. In contrast, the inefficiency ratio of the Gaussian PML estimator of ρ is $\mathcal{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$, where $\mathcal{M}_{ll}(\boldsymbol{\eta}_0)$ is defined in (3).*

This means that the 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.

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

$$\left.\begin{array}{c} y_t = \pi_0 + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha(y_{t-1}-\pi)^2, \\ \varepsilon_t^*|I_{t-1};\pi,\rho,\omega,\boldsymbol{\eta} \sim i.i.d.\ D(0,1,\boldsymbol{\eta}), \\ \text{with density function } f(.,\boldsymbol{\eta}) \end{array}\right\}, \tag{11}$$

where the parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, $\boldsymbol{\theta}' = (\boldsymbol{\theta}_s', \alpha)'$ and $\boldsymbol{\theta}_s = (\pi, \omega)'$. 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*

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

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

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

$$LM_{ARCH(1)} = T \cdot \frac{\bar{G}_s^2(1)}{\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0)} \tag{12}$$

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

$$\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0) = V[\tfrac{1}{2}\epsilon_t^2(\boldsymbol{\theta}_s)|\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}_0] \cdot M_{ss}(\boldsymbol{\eta}_0)$$

and

$$M_{ss}(\boldsymbol{\eta}) = V\left[\left.\frac{\partial \ln f(\varepsilon_t^*, \boldsymbol{\eta})}{\partial \varepsilon^*}\varepsilon_t^*\right| I_{t-1}; \boldsymbol{\eta}\right]. \tag{13}$$

2. *This asymptotic null distribution is unaffected if we replace π_0, ω_0 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 (13), or exploit the information matrix equality and calculate it as twice the sample average of $1 - \partial^2 \ln f\,[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}]\,/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t^2(\boldsymbol{\theta}_s)$. Once again, this choice will affect the finite sample properties of the tests (see Davidson and MacKinnon (1983)).

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

$$E\left[\left.\left\{1 + \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s), \boldsymbol{\eta}_0]}{\partial \varepsilon^*}\epsilon_t(\boldsymbol{\theta}_s)\right\} \epsilon_{t-1}^2(\boldsymbol{\theta}_s)\right| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0\right] = 0, \tag{14}$$

which is also related to the moment conditions used by Bontemps and Meddahi (2007) in their distributional tests. 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 $\epsilon^2_{t-1}(\boldsymbol{\theta}_s)$.

In this light, 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 $\epsilon^2_t(\boldsymbol{\theta}_s)$ times a damping factor that accounts for skewness and kurtosis. Figure 4 illustrates the transformation of the regressands for the same four 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.

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 $\epsilon^2_t(\boldsymbol{\theta}_s)$ 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 (11) we assume that the conditional distribution of ε^*_t is i.i.d. $N(0,1)$, when in fact it is i.i.d. $D(0,1,\boldsymbol{\varrho}_0)$, then (12) 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 $V\left[\epsilon^2_t(\boldsymbol{\theta}_s)\middle|\, \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0\right]$.*

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 $V\left[\epsilon^2_t(\boldsymbol{\theta}_s)\middle|\, \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0\right] = 2$, will be incorrectly sized.

But again, it is important to emphasise that the orthogonality condition (14) underlying our proposed ARCH test also remains valid under the null regardless of whether or not the assumed parametric distribution is correct. Therefore, 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 $\epsilon^2_{t-1}(\boldsymbol{\theta}_s)$ would continue to be asymptotically distributed as a χ^2_1 under the null. In practice, though, researchers would 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 pseudo-MLE 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 (11) 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 $\epsilon^2_{t-1}(\hat{\boldsymbol{\theta}}_s)$ will continue to be distributed as a χ^2 with 1 degree of freedom as T goes to infinity under the null hypothesis $H_0 : \alpha = 0$.*

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 in the Laplace case.

Importantly, a test based on (14) will continue to have non-trivial power even though it will no longer be an LM test. In fact, it might well be the case that our proposed tests are more powerful than the usual regression-based tests in Proposition 6 even though the parametric distribution is misspecified.

The test proposed in Proposition 5, though, 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^*. Once again, there are two possibilities: unrestricted nonparametric density estimates (SP) and nonparametric density estimates that impose symmetry (SSP). It turns out that 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:

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 (5) and (6), which coincide with the sample mean and variance of y_t.*

2. The same result is true 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 (7) and (8).

In addition, the semiparametric predictability test is adaptive, in the sense that it is as powerful as if we knew the distribution of ε_t^*, including the true values of the shape parameters $\boldsymbol{\eta}$ (see Linton and Steigerwald (2000)). Similarly, the symmetric semiparametric test will be adaptive too if the true distribution is symmetric. 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.

3.2 Exploiting the persistence of volatilities

Let us now consider a situation in which

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

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

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

Given that under then null hypothesis y_t is independent and identically distributed, 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

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

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 Gourieroux, 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, though, that since $\sigma_t^2(\boldsymbol{\theta}) = \omega(1-\beta)^{-1} + \alpha\sum_{j=0}^{t-2}\beta^j\varepsilon_{t-j-1}^2(\boldsymbol{\theta})$, α has to be positive under the alternative to guarantee nonnegative variances 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 $T\tilde{R}^2$ 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 $\sum_{j=0}^{t-2}\bar{\beta}^j\epsilon_{t-j-1}^2(\boldsymbol{\theta}_s)$ (see Demos and Sentana (1998)). The one-sided versions of such tests are asymptotically distributed as a $50 : 50$ mixture of χ_0^2 and χ_1^2 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 $\sum_{j=0}^{t-2} \bar{\beta}^j \epsilon_{t-j-1}^2(\boldsymbol{\theta}_s)$, 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 $(1-\bar{\beta}^2)^{-1}\mathcal{I}_{\alpha\alpha}(\boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0)$ 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 begin by comparing 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 $\alpha_{0T} = \bar{\alpha}/\sqrt{T}$. As we show in Appendix B, when the true model is (9), the non-centrality parameter of the Gaussian pseudo-score test based on the first order serial correlation coefficient of $\epsilon_t^2(\boldsymbol{\theta}_s)$ 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 5a 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 (11) with $\alpha_0 = 0$, then the feasible ML estimator of α is as efficient as the infeasible ML estimator, which require knowledge of $\boldsymbol{\eta}_0$. If in addition the conditional distribution is symmetric and $\kappa_0 < \infty$, then the elliptically symmetric semiparametric estimator of α is also fully efficient. The same is true in general of the semiparametric estimator of α. 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 (13).*

Proposition 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 5b 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 5c repeats the same exercise for two discrete location scale mixture of normals whose kurtosis coefficients are both 6, and whose skewness coefficients are -.5 and -1.219, respectively. In this case, our test yield 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.

4 Monte Carlo analysis

4.1 Design

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 empirical application in section 5. Specifically, we consider the following univariate, covariance stationary AR(12)-GARCH(1,1) model:

$$\begin{aligned} y_t &= \mu_t(\pi_0, \rho_0) + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \mu_t(\pi, \rho) &= \pi(1-12\rho) + \rho \textstyle\sum_{j=1}^{12} y_{t-j}, \\ \sigma_t^2(\boldsymbol{\theta}) &= \omega(1-\alpha-\beta)\upsilon(\rho) + \alpha_j[y_{t-1} - \mu_{t-1}(\pi, \rho)]^2 + \beta\sigma_{t-1}^2(\boldsymbol{\theta}), \\ \varepsilon_t^*|I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 &\sim i.i.d.\ D(0, 1, \boldsymbol{\eta}_0). \end{aligned}$$

We choose $\upsilon(\rho)$ in such a way that by construction $E(y_t) = \pi$ and $V(y_t) = \omega$. We set $\pi = .5$ and $\omega = 18$. Although these values are inconsequential for our econometric results, in annualised terms they imply a realistic risk premia of 6%, a standard deviation 14.7%, and a Sharpe ratio .41. For the sake of brevity we only report the results for $T = 720$ observations (plus another 100 for initialisation), which correspond to 60 years of monthly data, roughly the same as in our empirical analysis. We systematically rely on 20,000 replications, which means, for instance, that the 95% confidence interval for a nominal size of 5% would be (4.7%,5.3%). As for $\boldsymbol{\eta}_0$, we consider four different standardised distributions: Gaussian, Student t_6, a standardised mixture of two normals with finite higher order moments, the same kurtosis (=6) but negative skewness (=-.5), and an asymmetric t distribution with the maximum negative skewness compatible with the kurtosis of a univariate t_6 (=-1.219; see Mencia and Sentana (2009a,b) for details).

These distributions allow us to assess the reliability of our asymptotically robust tests, and to shed some light on the advantages of those tests that exploit the leptokurtosis and potential asymmetries of financial returns.

Importantly, we use the same underlying pseudo-random numbers in all designs to minimise experimental error. In particular, we make sure that the standard Gaussian random variables are the same for all four distributions. Given that the usual routines for simulating gamma random variables involve some degree of rejection, which unfortunately can change for different values of η, 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 in closed form using (5) and (6). 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 could 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 constraint $0 \leq \eta \leq .499$. As for the discrete mixture of normals, we use the EM algorithm described in Appendix D.5 to obtain good initial values, and then we numerically maximise the log-likelihood function of y_t in terms of the shape parameters $\boldsymbol{\eta} = (\delta, \varkappa, \lambda)'$ keeping $\boldsymbol{\theta}_s$ fixed at their Gaussian ML estimates.

Computational details for the symmetric and general semiparametric procedures can be found in Appendix B of Fiorentini and Sentana (2007). 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).

4.2 Finite sample size

The size properties under the null of the different LM tests for first-order serial correlation are summarised in Figures 6a-6d using Davidson and MacKinnon's (1998) p-value discrepancy plots, which show the difference between actual and nominal test sizes for every possible nominal size. When the distribution is Gaussian, all tests are very accurate. The same conclusion is obtained when the distribution is a Student t or a discrete mixture of normals, although there are some very minor distortions. Similarly, all tests are remarkably reliable when the conditional distribution is an asymmetric Student t, which confirms the theoretical results in Proposition

3. By and large we reach the same conclusions when we consider the restricted 12^{th}-order serial correlation tests discussed in section 2 (see Figures 7a-7d).

In turn, Figures 8a-8d show the size distortions of the two-sided versions of the different Arch(1) tests. When the distribution is Gaussian, all tests are quite accurate, but when the distribution is a Student t, a normal mixture or an asymmetric t, they tend to over-reject very slightly for low significance values, and underreject for larger ones. This is particularly true of the robust Gaussian tests, and to some extent of the test based on a two component normal mixture, especially in the asymmetric t case.

Finally, Figures 9a-9d describe the finite size properties of the two-sided versions of the different Garch(1) tests calculated with the discount factor $\bar{\beta} = .94$ suggested in Riskmetrics (1996). In this case, the tendency to under-reject is attenuated and sometimes even reversed.

4.3 Finite sample power

In order to gauge the power of the serial correlation tests we look at a design in which $\rho = 2/\sqrt{720}$ but $\alpha = 0$. The evidence at the 5% significance level is presented in Table 1, which includes raw rejection rates, as well as size adjusted ones based on the empirical distribution obtained under the null, which in this case provides the closest match because the Gaussian PML estimators of $\boldsymbol{\theta}_s$ that ignore the dynamics in y_t remain consistent in the presence of serial correlation or conditional heteroskedasticity.

As expected from the theoretical analysis in section 2.3, our proposed tests show clear power gains over standard (i.e. Gaussian) tests in the presence of non-normal distributions, with the parametric tests performing somewhat better than the semiparametric ones even in those situations in which the assumed distribution is misspecified.

We also look at a design with $\rho = 0$ but $\alpha = 2/\sqrt{720}$ and $\beta = 0$ to assess the power of the Arch(1) tests. Once again, we find that the usual Gaussian tests are dominated by all our proposals, but the semiparametric tests fail to achieve maximum power.

5 Empirical application

In this section we apply the procedures considered previously to the three Fama and French factors for US stocks, which we have obtained from Ken French's Data Library (see <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html> for further details). We use monthly data from January 1952 to December 2008, so that our sample begins soon after the March 1951 Treasury - Federal Reserve Accord whereby the Fed stopped its wartime pegging of interest rates. Thus, we consider 672 observations as we reserve 1952 for pre-sample values.

The exact factor definition is as follows:

1. MK: Value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates).
2. SMB: Returns on small cap firms in excess of the returns on large cap firms.
3. HML: Returns on value firms in excess of the returns on growth firms.

Descriptive statistics can be found in Table 2. As can be seen, the most distinctive feature of the distribution of monthly returns to these portfolios is their leptokurtosis. Once we take this feature in to account, we also find some statistically significant evidence for skewness in the market portfolio. The SMB and HML factors, on the other hand, appear to be symmetric.[3]

Table 3 reports the results of the different mean predictability tests discussed in section 2. As can be seen, there is not consistent evidence of serial correlation in the market portfolio. In contrast, there is substantial evidence that the SMB and HML factors are serially correlated. This is true not only when we test against AR(1) alternatives, but also when we consider restricted AR(12) alternatives too. This interesting divergence could be due to the fact that large stocks, which dominate a value weighted portfolio by construction, are more closely followed by investors than small, value stocks. Another result worth mentioning is the resemblance between the Student t and SSP tests on the one hand, and the normal mixture and SP tests on the other.

Finally, Table 4 presents our tests for conditional heteroskedasticity. Not surprisingly, we find very strong evidence of first order ARCH effects. This evidence is typically stronger when we use our fixed β GARCH tests instead. Therefore, the lack of conditional homoskedasticity that is observed at daily frequencies seems to get preserved in monthly data.

6 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. We also explain how to transform the regressor to exploit the persistence of expected returns and volatilities.

Importantly, we 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

[3]More formally, in both cases we find that the log likelihood function corresponding to a symmetric Student t is hardly increased when we estimate an asymmetric t, or an asymmetric Generalised Hyperbolic (see Mencia and Sentana (2009b) for details).

likelihood (PML) testing procedures advocated by Bollerslev and Wooldridge (1992) among many others. We also show that our non-parametric tests should be as efficient as if we knew the true distribution of the data.

We present local power analyses that confirm that there are clear power gains from exploiting the non-normality of financial returns, especially for variance predictability tests, as well as the persistent behaviour of risk premia and volatility. We complement our theoretical results with detailed Monte Carlo exercises that document the reliability of our mean predictability tests in finite samples. In contrast, we find some mild size distortions in the conditionally homoskedasticity tests. Given that y_t is assumed *i.i.d.* under the null, it would be useful to explore bootstrap procedures. In addition, we observe 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.

When we apply our methods to monthly stock returns on the three Fama & French factors for US stocks, we find clear evidence in favour of first order serial correlation in the size and value factors only, slightly weaker evidence for persistent components in those factors, and no evidence that such a component appears in the market portfolio. We also find strong evidence for persistent serial correlation in the volatility of all three series.

It would be interesting to analyse the power gains of tests based on the wrong parametric distribution along the lines of Amengual and Sentana (2010), who focus on mean variance efficiency tests. Relatedly, 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.

Another interesting extension would be to consider nonparametric 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, either in univariate contexts, or in multivariate ones, as in Fiorentini and Sentana (2009). We are currently exploring 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

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

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

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. $\square$

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 $\rho = 0$

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

Given that the pseudo-true values of π, ω and η are implicitly defined in such a way that

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

the law of iterated expectations implies that

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

and

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

where

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

Consequently,

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

where

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

On this basis, we can show that

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

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

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

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

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

Finally we obtain

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

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

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

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

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

$$\begin{aligned}
\mathcal{S}(\boldsymbol{\phi}_0) \;=\; & \begin{bmatrix} \omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) \\ 0 & \mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 \\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & 0 & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} - \\
& \begin{pmatrix} \omega_0^{-1/2} & 0 \\ \mathbf{0} & \mathbf{0} \\ 0 & \frac{1}{2}\omega_0^{-1} \end{pmatrix} \left\{ \begin{bmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) \end{bmatrix} \right. \\
& \left. - \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & \varphi \\ \varphi & \kappa - 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right\} \begin{pmatrix} \omega_0^{-1/2} & 0 & 0 \\ 0 & 0 & \frac{1}{2}\omega_0^{-1} \end{pmatrix} \\
\;=\; & \begin{bmatrix} \omega^{-1} & 0 & \frac{1}{2}\omega^{-3/2}\varphi \\ 0 & \mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 \\ \frac{1}{2}\omega^{-3/2}\varphi & 0 & \frac{1}{4}\omega^{-2}(\kappa - 1) \end{bmatrix}.
\end{aligned}$$

. □

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 (11), 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

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_s, 0) = \begin{bmatrix} \omega^{-1/2} & 0 \\ 0 & \frac{1}{2}\omega^{-1} \\ 0 & \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1] \end{bmatrix}, \tag{A3}$$

so that

$$\mathbf{Z}_d(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta}_0) = \begin{bmatrix} \omega_0^{-1/2} & 0 \\ 0 & \frac{1}{2}\omega_0^{-1} \\ 0 & 0 \end{bmatrix}. \tag{A4}$$

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

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

$$\times \begin{bmatrix} \omega^{-1/2} & 0 & 0 & \mathbf{0} \\ 0 & \frac{1}{2}\omega^{-1} & \frac{1}{2}[\epsilon_{t-1}^2(\boldsymbol{\theta}_s) - 1] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix}$$

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

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 (A3) 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 $\mathcal{B}_{\rho\rho}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0)=\frac{1}{2}[\kappa(\boldsymbol{\varrho})-1]\mathcal{A}_{\rho\rho}(\boldsymbol{\theta}_s,0,\mathbf{0};\boldsymbol{\theta}_0,0,\boldsymbol{\varrho}_0)$. As a result, the Gaussian-based LM test for $H_0:\alpha=0$ 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

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

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

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

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

and $\mathcal{B}_{\alpha\alpha}$, $\mathcal{A}_{\alpha\boldsymbol{\phi}_s}$, 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 $\alpha = 0$:

$$\begin{aligned}
h_{\alpha\pi t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-1/2}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^*\} \cdot [\epsilon^2_{t-1}(\boldsymbol{\theta}_s) - 1] \\
&\quad + \omega^{-1/2}\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + 1\} \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s) \\
h_{\alpha\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{4}\omega^{-1}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon^2_t(\boldsymbol{\theta}_s) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s)\} \cdot [\epsilon^2_{t-1}(\boldsymbol{\theta}_s) - 1] \\
&\quad + \tfrac{1}{2}\omega^{-1}\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) + 1\} \cdot \epsilon^2_{t-1}(\boldsymbol{\theta}_s) \\
\mathbf{h}_{\alpha\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= -\tfrac{1}{2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta} \cdot \epsilon_t(\boldsymbol{\theta}_s) \cdot [\epsilon^2_{t-1}(\boldsymbol{\theta}_s) - 1]
\end{aligned}$$

Given that the pseudo-true values of π, ω and η are implicitly defined in such a way that

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

the law of iterated expectations implies that

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

where $\mathcal{H}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi})$, $\mathcal{H}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi})$ and $\mathcal{H}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi})$ are definied in the proof of Proposition 3, while

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

Similarly, it is clear that when $\alpha = 0$ the expressions for the Hessian elements corresponding to π, ω and $\boldsymbol{\eta}$ will coincide with the ones obtained in the proof of Proposition 3 evaluated at $\rho = 0$, and the same is true of their expected values.

On this basis, we can easily show that

$$\mathcal{A}_{\alpha\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} = \begin{pmatrix} 0 & E[\epsilon^2_t(\boldsymbol{\theta}_{s\infty}) - 1|\boldsymbol{\varphi}_0]\omega_\infty & 0 \end{pmatrix}$$

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

$$\begin{aligned}
\mathcal{B}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= E[\mathcal{B}_t(\boldsymbol{\phi};\boldsymbol{\varphi})|\boldsymbol{\varphi}], \\
\mathcal{B}_t(\boldsymbol{\phi};\boldsymbol{\varphi}) &= V[\mathbf{s}_{\boldsymbol{\phi}t}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta})|\, I_{t-1};\boldsymbol{\varphi}] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{K}(\boldsymbol{\phi};\boldsymbol{\varphi})\mathbf{Z}'_t(\boldsymbol{\theta}),
\end{aligned}$$

where $\mathcal{K}(\boldsymbol{\phi};\boldsymbol{\varphi})$ is also defined in the proof of Proposition 3, we will have that under the null of $H_0 : \alpha = 0$,

$$\begin{bmatrix} \mathcal{B}_{\omega\omega}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{B}_{\omega\alpha}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{B}_{\omega\alpha}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{B}_{\alpha\alpha}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{bmatrix} = \frac{1}{4}\mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) \begin{bmatrix} \omega_\infty^{-2} & \omega_\infty^{-1}E[\epsilon^2_t(\boldsymbol{\theta}_{s\infty}) - 1|\boldsymbol{\varphi}_0] \\ \omega_\infty^{-1}E[\epsilon^2_t(\boldsymbol{\theta}_{s\infty}) - 1|\boldsymbol{\varphi}_0] & E\{[\epsilon^2_t(\boldsymbol{\theta}_{s\infty}) - 1]^2|\boldsymbol{\varphi}_0\} \end{bmatrix}.$$

Finally we obtain

$$\mathcal{F}_{\alpha\alpha}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_{\infty};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0)=\frac{1}{4}\mathcal{K}_{ss}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)V[\epsilon_t^2(\boldsymbol{\theta}_{s\infty})]|\boldsymbol{\varphi}_0],$$

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

Proposition 8

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

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)=\begin{bmatrix}\omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & 0\\ 0 & \frac{1}{(\kappa-1)\omega^2} & 0\\ 0 & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta})\end{bmatrix}.$$

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

$$\begin{aligned}\mathcal{S}(\boldsymbol{\phi}_0) &= \begin{bmatrix}\omega^{-1}\mathcal{M}_{ll}(\boldsymbol{\eta}) & 0 & \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta})\\ 0 & \frac{1}{4}\omega^{-2}\mathcal{M}_{ss}(\boldsymbol{\eta}) & 0\\ \frac{1}{2}\omega^{-3/2}\mathcal{M}_{ls}(\boldsymbol{\eta}) & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta})\end{bmatrix}-\\
&\quad\begin{pmatrix}\omega_0^{-1/2} & 0\\ 0 & \frac{1}{2}\omega_0^{-1}\\ 0 & 0\end{pmatrix}\left\{\begin{bmatrix}\mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta})\\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta})\end{bmatrix}\right.\\
&\quad\left.-\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}1 & \varphi\\ \varphi & \kappa-1\end{pmatrix}^{-1}\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}\right\}\begin{pmatrix}\omega_0^{-1/2} & 0 & 0\\ 0 & \frac{1}{2}\omega_0^{-1} & 0\end{pmatrix}\\
&= \begin{bmatrix}\omega^{-1} & \frac{1}{2}\omega^{-3/2}\varphi & 0\\ \frac{1}{2}\omega^{-3/2}\varphi & \frac{1}{4}\omega^{-2}(\kappa-1) & 0\\ 0 & 0 & \frac{\kappa-1}{4}\mathcal{M}_{ss}(\boldsymbol{\eta})\end{bmatrix}.\end{aligned}$$

. □

Lemma 2

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

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.

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.

Bickel, P.J. (1982): "On adaptive estimation", *Annals of Statistics* 10, 647-671.

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. (2007): "Testing distributional assumptions: a GMM approach", mimeo, Toulouse School of Economics.

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.

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

Crowder, M.J. (1976): "Maximum likelihood estimation for dependent observations", *Journal 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. (1998): "Graphical methods for investigating the size and power of tests statistics", *The Manchester School* 66, 1-26.

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.

Drost, F.C. and Klaassen, C.A.J. (1997): "Efficient estimation in semiparametric GARCH models", *Journal of Econometrics* 80, 193-221.

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

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

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. (2009): "Dynamic specification tests for static factor models", CEMFI Working Paper 0912.

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.

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.

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

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

Gourieroux C., Holly A. and Monfort A. (1980): "Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters", Harvard Institute of Economic Research Discussion Paper 770.

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

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.

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

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

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.

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley, Chichester.

Maronna, R., Martin, D. and Yohai, V. (2006): *Robust Statistics - Theory and Methods*, Wiley.

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

Mencía, J. and E. Sentana (2009b): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", mimeo, CEMFI.

NAG (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.

RiskMetrics Group (1996): *RiskMetrics Technical Document.*

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.

B Local power calculations

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

$$M_T = T\bar{\mathbf{m}}_T'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}), \tag{B5}$$

where $\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0})$ is the sample average of $\mathbf{m}_t(\boldsymbol{\theta})$ evaluated under the null, and $\boldsymbol{\Psi}$ is the corresponding asymptotic covariance matrix. 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 (B5) will be

$$\bar{\boldsymbol{\theta}}_2'\mathbf{M}'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2. \tag{B6}$$

On this basis, we can easily obtain the limiting probability of M_T exceeding some prespecified 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 (B6).

Finally, note that (B6) remains valid when we replace $\boldsymbol{\theta}_{10}$ by its ML estimator under the null if $\mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0})$ and the scores corresponding to $\boldsymbol{\theta}_1$ are asymptotically uncorrelated when H_0 is true, as in all our tests under correct specification. 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 $\boldsymbol{\theta}_2$. This result confirms that the non-centrality parameters of LM and Wald tests will be the same under sequences of local alternatives, which simplifies their computation.

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 (9) will be given by

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

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 $h > 1$.

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

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

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. 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.

C Simulation of standardised random variables

C.1 Discrete location scale mixtures of normals

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

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

where $\delta\in\mathbb{R}$ and $\varkappa>0$, 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

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

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 simulate the same standardised random variable is as follows

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

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

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

Therefore, we can immediately interpret $\varkappa$ as the ratio of the two variances. Similarly, since

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

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components. In particular, if we set $\delta = 0$ then we will obtain a discrete scale mixture of normals, which is always symmetric but leptokurtic.

The parameters λ, δ and $\varkappa$ 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

$$E(\varepsilon_t^{*3}) = \frac{3\delta\lambda(1-\lambda)(1-\varkappa)}{[\lambda + (1-\lambda)\varkappa][1+\lambda(1-\lambda)\delta^2]^{3/2}} + \frac{\delta^3(1-\lambda)\lambda(1-2\lambda)}{[1+\lambda(1-\lambda)\delta^2]^{3/2}}$$

and

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

A useful property of finite normal mixtures is that they span the entire skewness-kurtosis frontier $E(\varepsilon_t^{*4}) \geq 1 + E^2(\varepsilon_t^{*3})$ (see Stuart and Ord (1977)). In this sense, note that another way of obtaining discrete normal mixture distributions that are symmetric is by making $\lambda = \frac{1}{2}$ and $\varkappa = 1$.

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 $\varkappa$ 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.2 Generalised hyperbolic

Let ξ_t denote an *i.i.d.* Generalised Inverse Gaussian (GIG) random variable with parameters $-\nu, \gamma$ and 1, or $GIG(-\nu, \gamma, 1)$ for short. Mencia and Sentana (2009b) show that if $z_t|\xi_t$ is *i.i.d.* $N(0,1)$, then

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

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

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

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 Mencia and Sentana (2009a), in which ξ_t is a positive random variable with an arbitrary distribution.

Mencia and Sentana (2009b) also provide expressions for the third and fourth moments of the GH distribution, which in the univariate case reduce to

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

and

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

C.2.1 Asymmetric and symmetric versions of the Student t

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

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

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

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

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

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

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

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

and

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

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 $\lim_{\beta\to 0} c(\beta,\nu,\gamma) = 1$, 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.2.2 Symmetric Laplace distribution

The asymmetric Laplace distribution is another special case of the GH distribution, which is achieved when $\gamma = 0$ and $\nu = 1$. In fact, it is a special case of the asymmetric normal-gamma mixture, which allows ν to be any positive parameter. 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 $i.i.d.$ $N(0,1)$. As is well known, the kurtosis coefficient of a symmetric Laplace distribution is 6.

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, 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 $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]$, 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 $\sigma_t^2(\boldsymbol{\theta})$ is strictly positive and $\mu_t(\boldsymbol{\theta})$, $\sigma_t^2(\boldsymbol{\theta})$ 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

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

where

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

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

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\partial \mathbf{Z}_{lt}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} e_{lt}(\boldsymbol{\phi}) + \frac{\partial \mathbf{Z}_{st}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} e_{st}(\boldsymbol{\phi}) + \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial e_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial e_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\theta}'} & \text{(D7)} \\ \mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta}) \frac{\partial e_{lt}(\boldsymbol{\phi})}{\partial \boldsymbol{\eta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta}) \frac{\partial e_{st}(\boldsymbol{\phi})}{\partial \boldsymbol{\eta}'} & \text{(D8)} \\ \mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= \partial^2 \ln f[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}]/\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}', \end{aligned}$$

where

$$\partial \mathbf{Z}_{lt}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = -\tfrac{1}{2}\cdot\sigma_t^{-3}(\boldsymbol{\theta})\cdot\partial\mu_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' - \sigma_t^{-1}(\boldsymbol{\theta})\cdot\partial^2\mu_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}',$$

$$\partial \mathbf{Z}_{st}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = -\tfrac{1}{2}\cdot\sigma_t^{-4}(\boldsymbol{\theta})\cdot\partial\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' - \tfrac{1}{2}\cdot\sigma_t^{-2}(\boldsymbol{\theta})\cdot\partial^2\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}',$$

$$\partial e_{lt}(\boldsymbol{\phi})/\partial\boldsymbol{\theta}' = \partial^2\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\mathbf{Z}_{lt}'(\boldsymbol{\theta}) + \partial^2\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\varepsilon_t^*(\boldsymbol{\theta})\cdot\mathbf{Z}_{st}'(\boldsymbol{\theta})$$

$$\partial e_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\theta}' = \{\partial\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varepsilon^* + \partial^2\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\varepsilon_t^*(\boldsymbol{\theta})\}\mathbf{Z}_{lt}'(\boldsymbol{\theta})$$
$$+\{\partial\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varepsilon^*\cdot\varepsilon_t^*(\boldsymbol{\theta}) + \partial^2\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\varepsilon_t^{2*}(\boldsymbol{\theta})\}\cdot\mathbf{Z}_{st}'(\boldsymbol{\theta})$$

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 $\sqrt{T}(\boldsymbol{\phi}_T - \boldsymbol{\phi}_0) \to N[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)]$, where $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$, where

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

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

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

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 $\boldsymbol{\eta}$ 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 9 *If $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}_0$ is i.i.d. $D(0,1,\boldsymbol{\eta}_0)$ with $\kappa_0 < \infty$, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N\left[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)\right]$, where*

$$\mathcal{C}(\boldsymbol{\phi}) = \mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi}),$$
$$\mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],$$
$$\mathcal{A}_t(\boldsymbol{\phi}) = -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\mathbf{0})\mathbf{Z}_{dt}'(\boldsymbol{\theta}),$$
$$\mathcal{B}(\boldsymbol{\phi}) = V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta},\mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],$$
$$\mathcal{B}_t(\boldsymbol{\phi}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta};\mathbf{0})|\,\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}_{dt}'(\boldsymbol{\theta}),$$
$$\text{and } \mathcal{K}\left(\varphi,\kappa\right) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t, I_{t-1};\boldsymbol{\phi}] = \begin{bmatrix} 1 & \varphi(\boldsymbol{\eta}) \\ \varphi(\boldsymbol{\eta}) & \kappa(\boldsymbol{\eta})-1 \end{bmatrix}, \tag{D9}$$

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

$$\varphi(\boldsymbol{\eta}) = E(\varepsilon_t^{*3}|\boldsymbol{\eta}). \tag{D10}$$
$$\kappa(\boldsymbol{\eta}) = E(\varepsilon_t^{*4}|\boldsymbol{\eta}). \tag{D11}$$

Given that $\varphi(\boldsymbol{\eta}) = 0$ 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 $\boldsymbol{\theta}$

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

Proposition 10 *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*

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

and the matrix of third and fourth order central moments

$$\mathcal{K}\left(\boldsymbol{\varrho}\right) = V[\mathbf{e}_{dt}(\boldsymbol{\theta},\mathbf{0})|\,\mathbf{z}_t, I_{t-1};\boldsymbol{\theta},\boldsymbol{\varrho}] \tag{D12}$$

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

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)\left[\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0) - \mathcal{K}\left(0\right)\mathcal{K}^{-1}(\varphi,\kappa)\mathbf{e}_{dt}(\boldsymbol{\theta}_0,\mathbf{0})\right], \tag{D13}$$

while the semiparametric efficiency bound is

$$\mathcal{S}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0)\left[\mathcal{M}_{dd}\left(\boldsymbol{\varrho}_0\right) - \mathcal{K}\left(0\right)\mathcal{K}^{1}(\varphi,\kappa)\mathcal{K}\left(0\right)\right]\mathbf{Z}_d'(\boldsymbol{\theta}_0,\boldsymbol{\varrho}_0), \tag{D14}$$

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

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 11 *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:*

$$\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0)=\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\left\{-[1+\varepsilon_t(\boldsymbol{\theta}_0)\partial\ln f[\varepsilon_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\varepsilon^*]-\frac{2}{\kappa_0-1}\left[\varepsilon_t^2(\boldsymbol{\theta}_0)-1\right]\right\}, \tag{D15}$$

where

$$\mathbf{W}_s(\boldsymbol{\phi}_0)=\mathbf{Z}_d(\boldsymbol{\phi}_0)\begin{pmatrix}0\\1\end{pmatrix}=E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]\begin{pmatrix}0\\1\end{pmatrix}=E\left\{\frac{1}{2\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|\boldsymbol{\phi}_0\Bigg\}, \tag{D16}$$

while the elliptically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0)=\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0)-\mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0)\cdot\left[\mathcal{M}_{ss}(\boldsymbol{\varrho}_0)-\frac{4}{\kappa_0-1}\right]. \tag{D17}$$

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 $\tilde{\boldsymbol{\theta}}_T=\arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta},\eta)$ 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 13 in Fiorentini and Sentana (2007) 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 α/ω 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 Discrete mixtures of normals based pseudo maximum likelihood estimators

The EM algorithm discussed by Dempster, Laird and Rubin, D. (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 $\hat{\lambda}^{(n}$, $\hat{\varkappa}^{(n}=\hat{\sigma}_2^{2(n}/\hat{\sigma}_1^{2(n}$ 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 $\varkappa$, 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 $\hat{\omega}^2$ coincide with the Gaussian PML estimators. As a result, we can maximise the log-likelihood function with respect to λ, δ and $\varkappa$ 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.

Table 1

Test power

(a) AR(1) tests. DGP: Student t_6 (ρ=$2/\sqrt{720}$, α=β=0)

	Normal	Student	DMN	SSP	SP
Rejection rate	0.495	0.563	0.558	0.549	0.539
Size adjusted	0.489	0.552	0.547	0.536	0.529

(b) AR(1) tests. DGP: DMN(φ=-.5,κ=6,λ=.05) (ρ=$2/\sqrt{720}$, α=β=0)

	Normal	Student	DMN	SSP	SP
Rejection rate	0.497	0.551	0.558	0.545	0.539
Size adjusted	0.512	0.561	0.570	0.559	0.548

(c) ARCH(1) tests. DGP: Student t_6 (ρ=0, α=$2/\sqrt{720}$,β=0)

	Normal	Student	DMN	SSP	SP
Rejection rate	0.374	0.442	0.407	0.376	0.351
Size adjusted	0.423	0.458	0.432	0.394	0.362

(d) ARCH(1) tests. DGP: DMN(φ=-.5,κ=6,λ=.05) (ρ=0, α=$2/\sqrt{720}$,β=0)

	Normal	Student	DMN	SSP	SP
Rejection rate	0.357	0.499	0.496	0.445	0.427
Size adjusted	0.411	0.516	0.510	0.450	0.434

Table 2

Descriptive statistics

Portfolio	Means	Std.dev.	Skewness (φ)	Kurtosis (κ)
Market	.485	4.307	-.599*	5.245*
SMB	.189	2.969	.573	9.441*
HML	.397	2.718	.078	5.797*

Notes: Sample: January 1953-December 2008. Definitions: Market: Value-weighted portfolio of all NYSE, AMEX and NASDAQ stocks; SMB: Size factor; HML: Value factor. The symbol * means statistically different from its value under normality at the 5% level.

Table 3

Serial correlation tests (p-values, %)

	AR(1)				
	Normal	Student	SSP	DMN	SP
Market	1.63%	7.89%	7.78%	31.10%	26.59%
SMB	11.60%	0.13%	0.02%	0.03%	0.01%
HML	0.03%	0.01%	0.01%	0.03%	0.03%

	AR(12)				
	Normal	Student	SSP	DMN	SP
Market	22.14%	29.18%	30.42%	96.45%	87.73%
SMB	4.61%	0.11%	0.12%	0.02%	0.27%
HML	3.17%	1.25%	1.07%	0.49%	4.37%

Notes: Sample: January 1953-December 2008. Definitions: Market: Value-weighted portfolio of all NYSE, AMEX and NASDAQ stocks; SMB: Size factor; HML: Value factor. PML refers to the Gaussian-based ML estimators, Student to the t-based ones, DMN to the ones based on a two component mixture of normals, SSP to the symmetric semiparametric estimators and SP to the general semiparametric estimators.

Table 4

Conditional heteroskedasticity tests (p-values, %)

	Arch(1)				
	PML	Student	SSP	DMN	SP
Market	0.06%	0.00%	0.00%	0.09%	0.03%
SMB	0.00%	0.00%	0.12%	0.00%	0.43%
HML	0.00%	0.00%	0.00%	0.00%	0.00%

	Garch(1,1)				
	PML	Student	SSP	DMN	SP
Market	0.14%	0.00%	0.00%	0.00%	0.00%
SMB	0.00%	0.00%	0.00%	0.00%	0.00%
HML	0.00%	0.00%	0.00%	0.00%	0.00%

Notes: Sample: January 1953-December 2008. Definitions: Market: Value-weighted portfolio of all NYSE, AMEX and NASDAQ stocks; SMB: Size factor; HML: Value factor. PML refers to the Gaussian-based ML estimators, Student to the t-based ones, DMN to the ones based on a two component mixture of normals, SSP to the symmetric semiparametric estimators and SP to the general semiparametric estimators.

Figure 1: Tests of predictability in mean

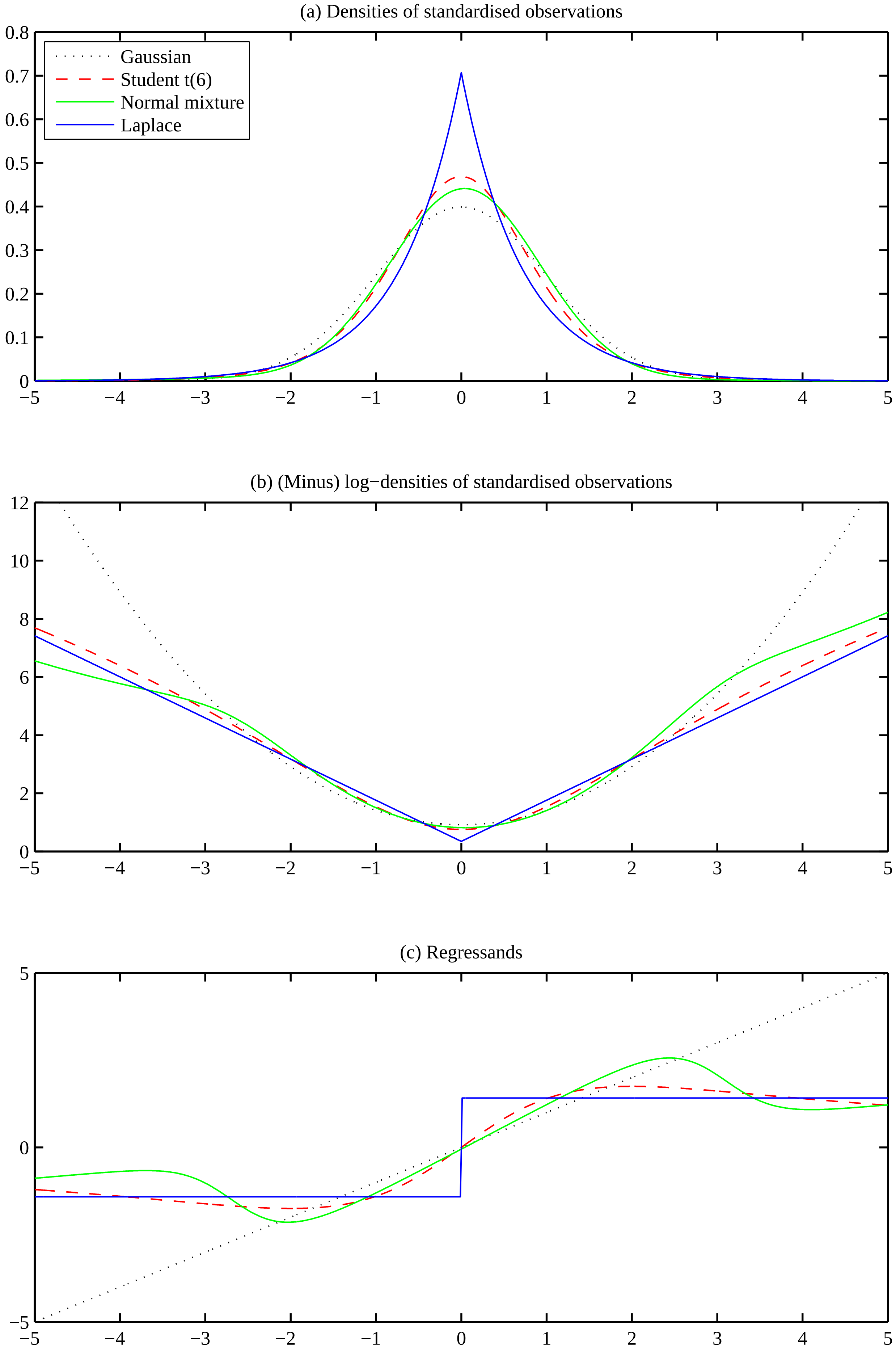

Figure 2: ACF of expected and observed returns (h=24, ρ=.015)

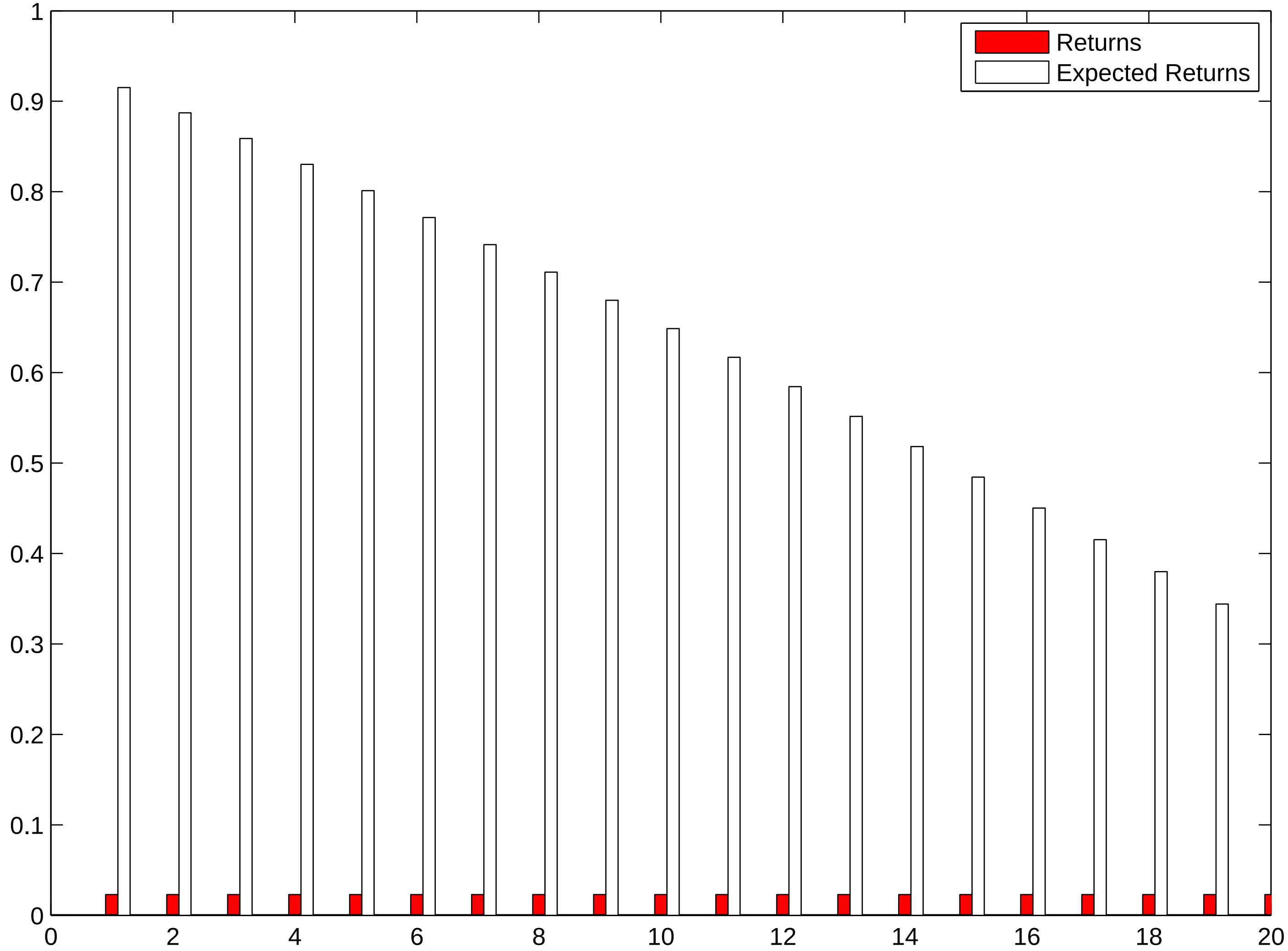

Figure 3: Power of mean dependence tests at 5% level against local alternatives

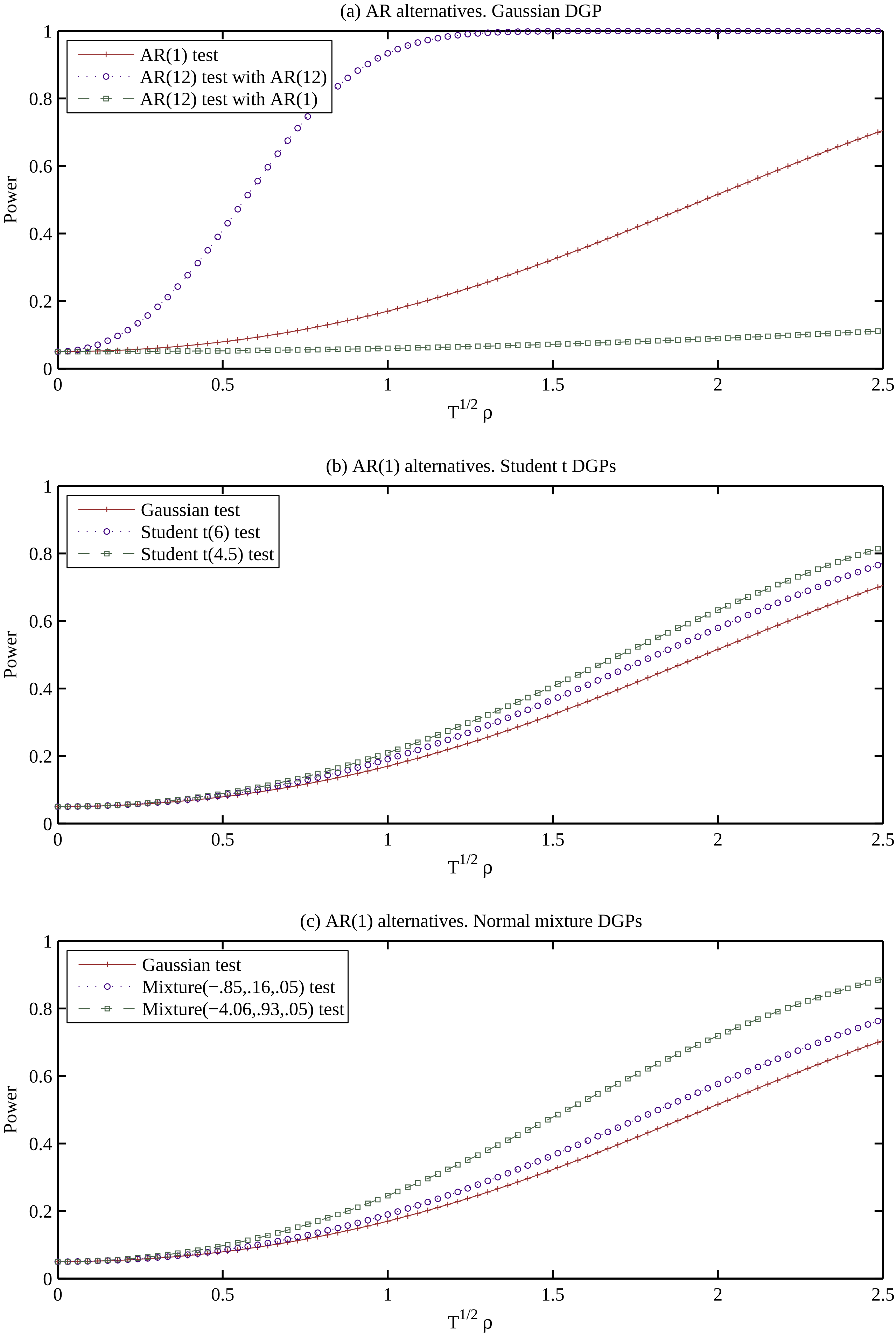

Figure 4: Tests of predictability in variance

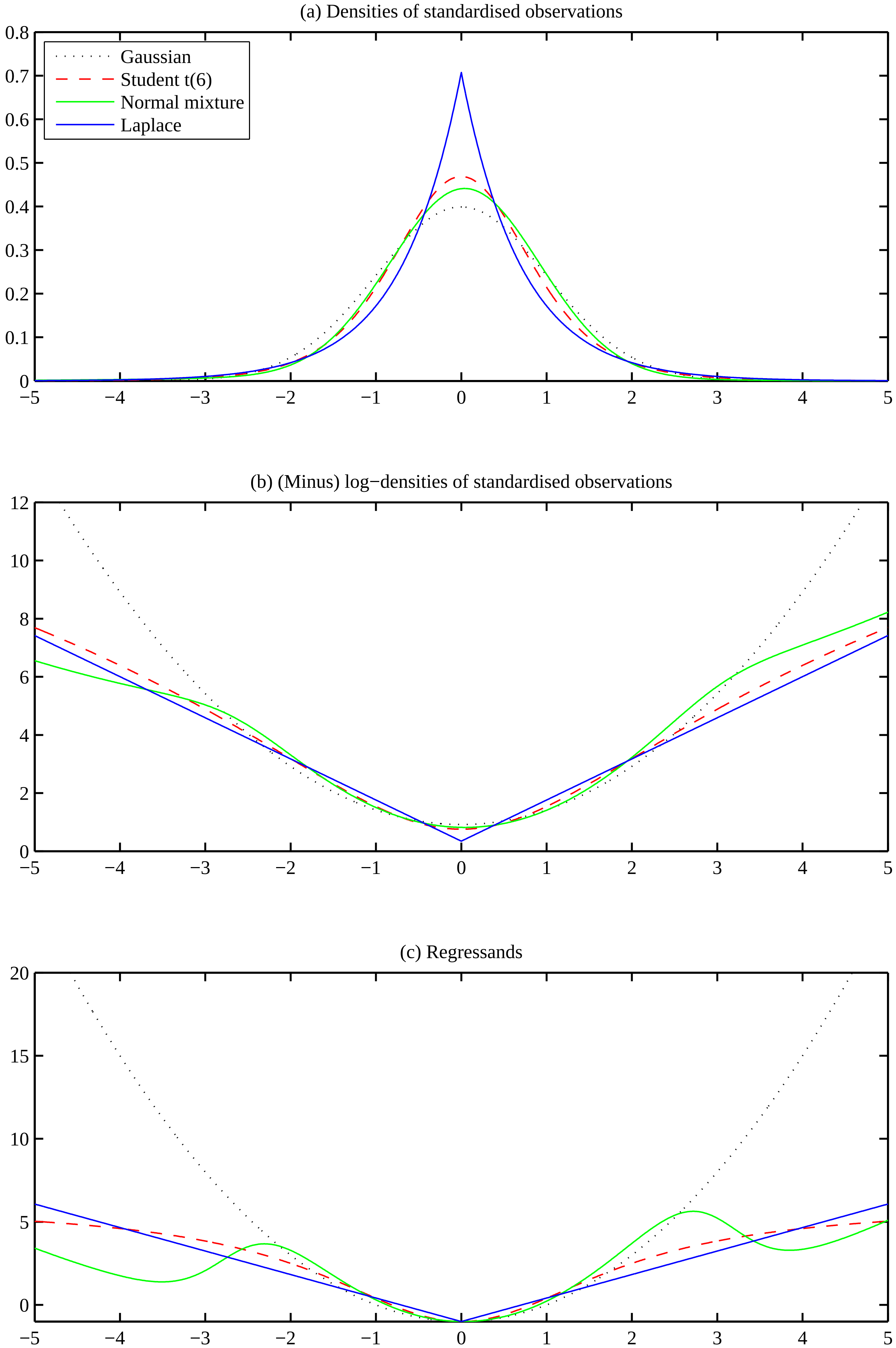

Figure 5: Power of variance dependence tests at 5% level against local alternatives

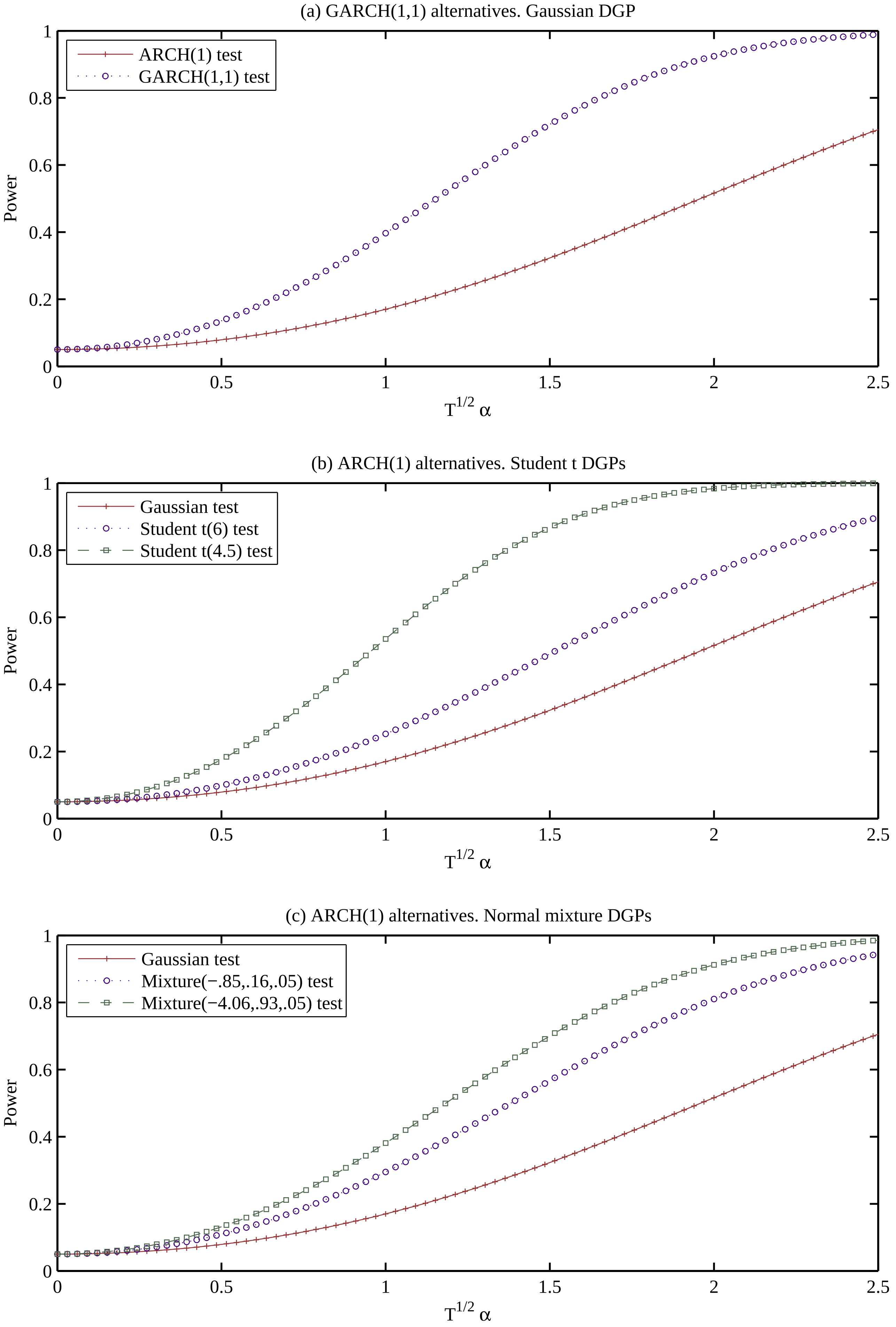

Figure 6: P−value discrepancy plots. Tests against AR(1) alternatives.

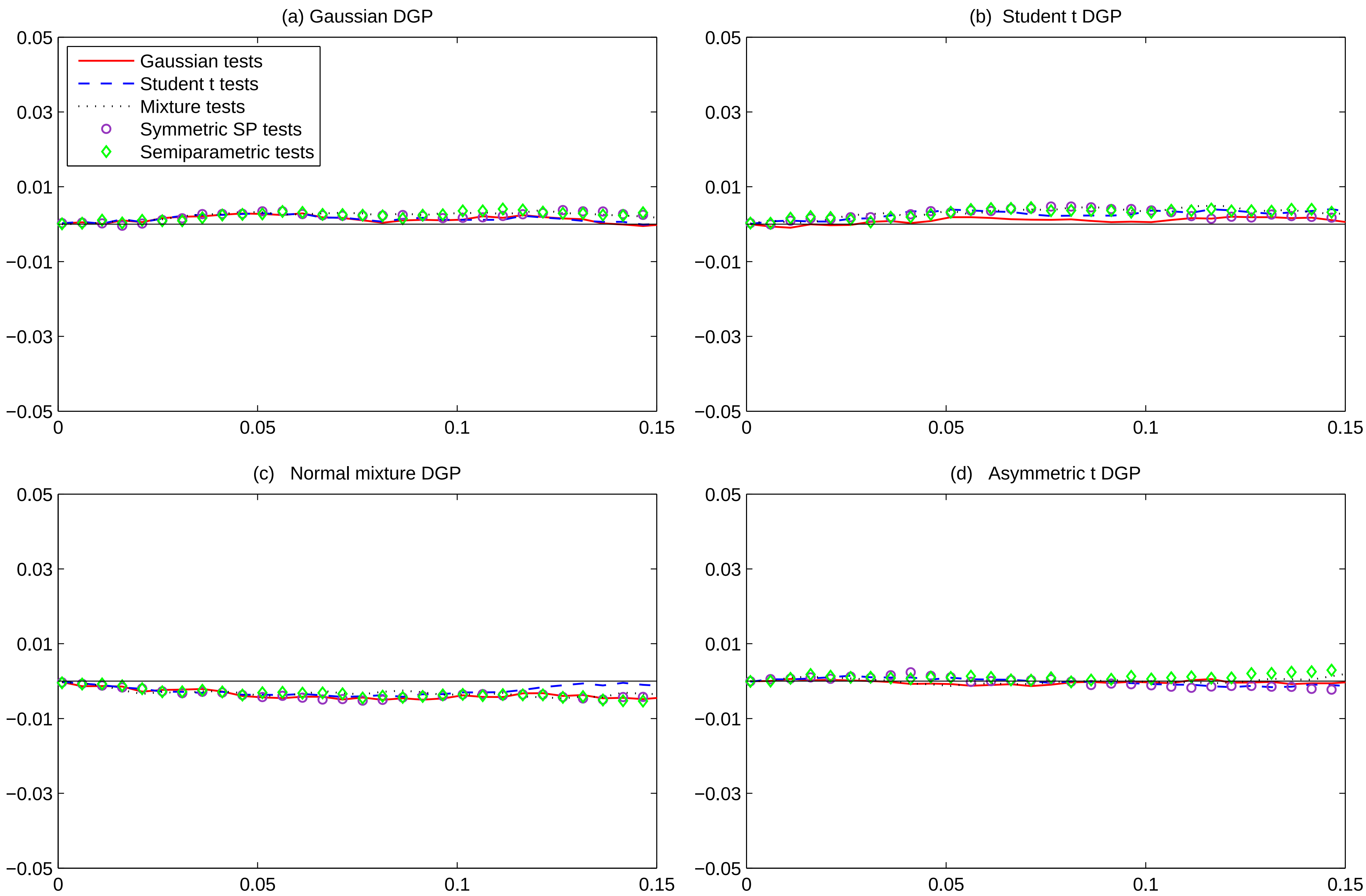

Figure 7: P−value discrepancy plots. Tests against AR(12) alternatives.

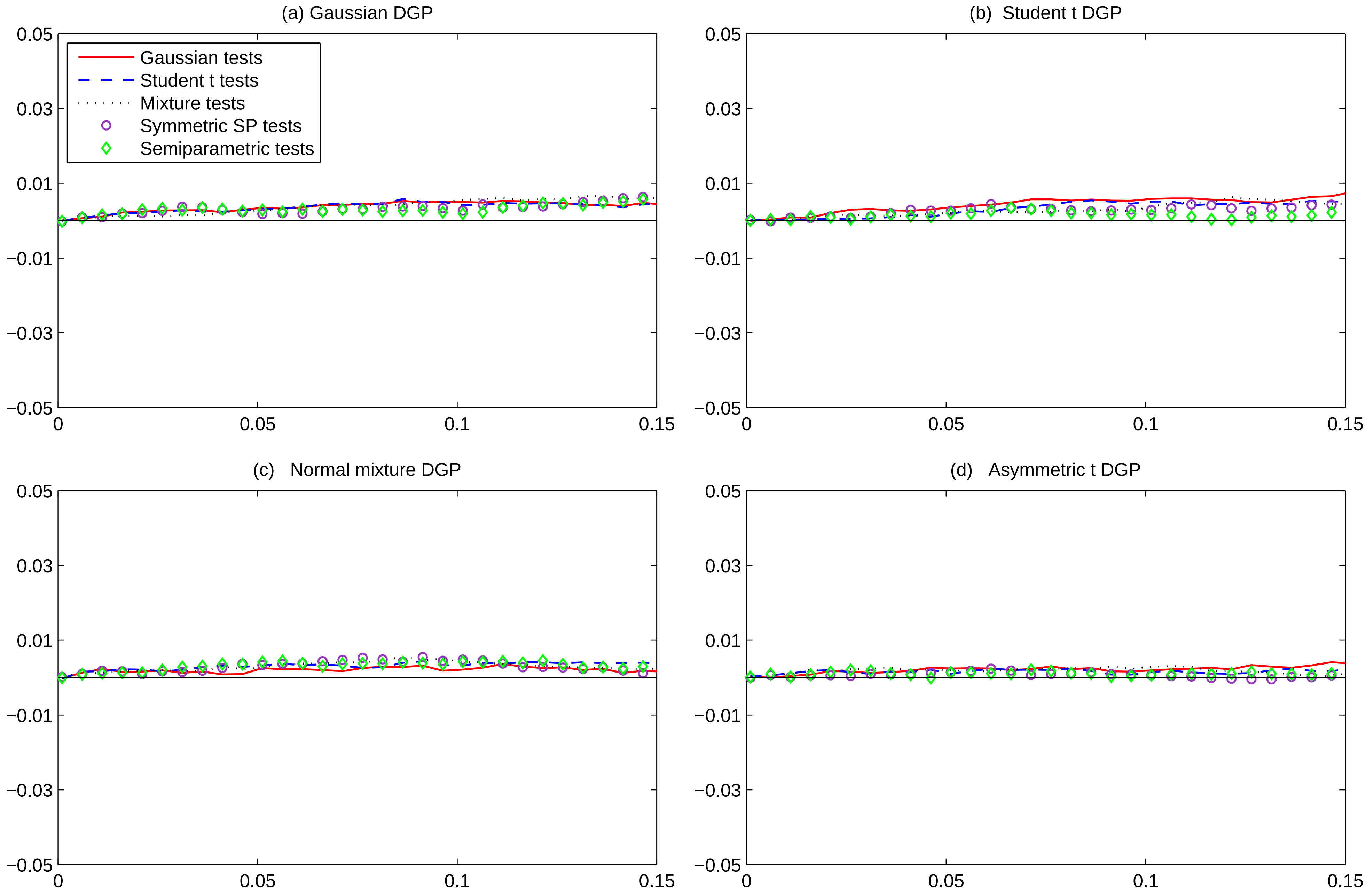

Figure 8: P−value discrepancy plots. Tests against ARCH(1) alternatives.

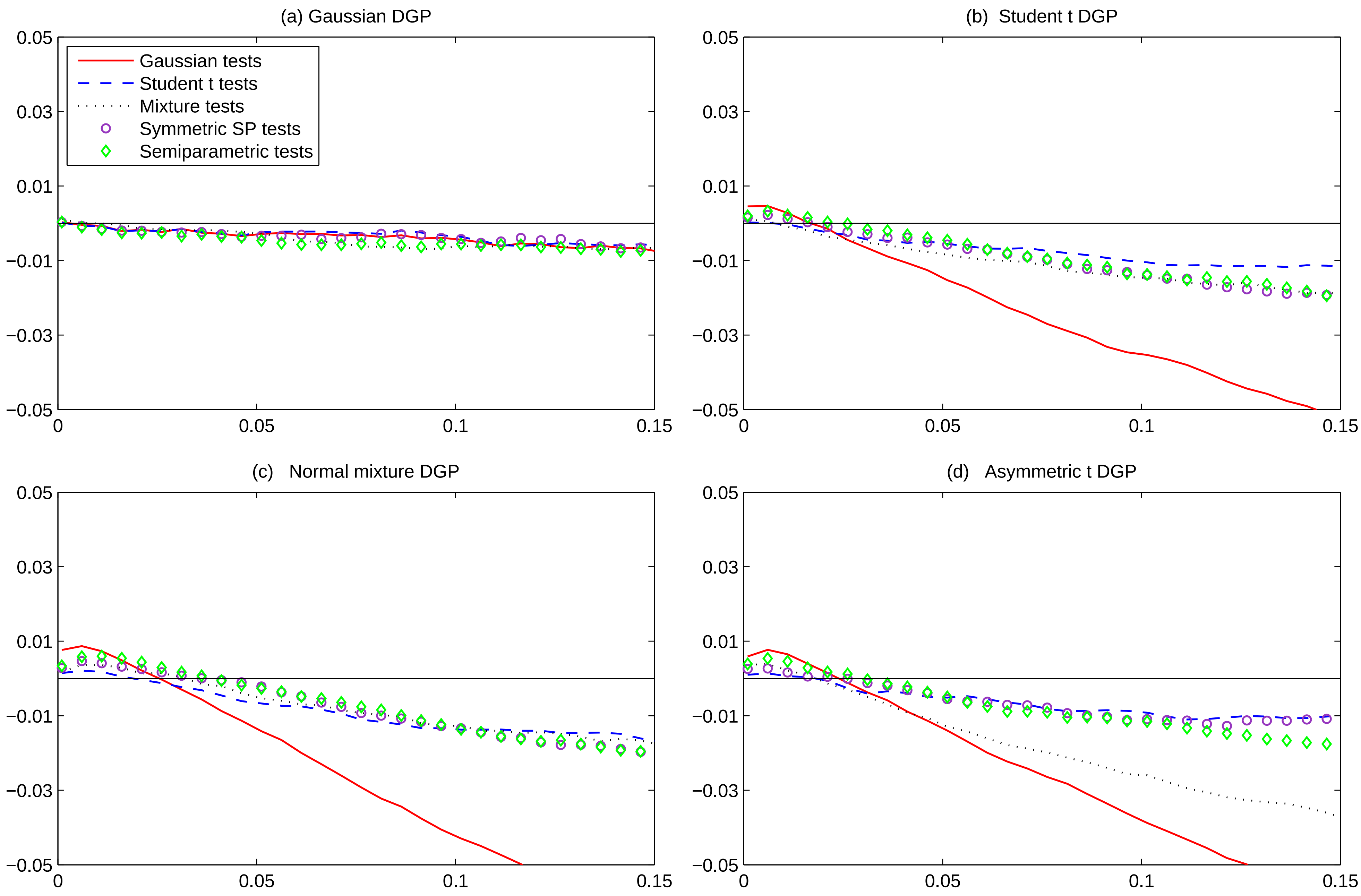

Figure 9: P−value discrepancy plots. Tests against GARCH(1,1) alternatives.