Testing Conditional Asymmetry. A Residual-Based Approach

Philippe Lambert *

Sébastien Laurent †

David Veredas ‡

Abstract

We propose three residual-based tests for conditional asymmetry. The distribution is assumed to fall into the class of skewed distributions of Fernández and Steel (1998). In this class, asymmetry is measured by the ratio between the probabilities of being larger and smaller than the mode. Estimation is performed under the null hypothesis of constant asymmetry of the innovations and, in a second step, tests for conditional asymmetry are performed on generalized residuals through parametric and nonparametric methods. We derive the asymptotic distribution of the tests that incorporates the uncertainty of the estimated parameters in the first step. A Monte Carlo study shows that neglecting this uncertainty severely biases the tests and an empirical application on a basket of daily returns reveals that financial data often present dynamics in the conditional skewness.

Keywords: Conditional asymmetry, residuals, Wald, runs.

JEL Classification: C32, G14, E44

*Méthodes quantitatives en sciences sociales, Université de Liège. Boulevard du Rectorat 7, B4000 Liège, Belgium. Tel: +3243665990; Fax: +3243664751; p.lambert@ulg.ac.be

†CeReFim, University of Namur, 8, rempart de la vierge B5000, Namur, Belgium. Tel: +3281724869; Fax: +3281724840; sebastien.laurent@fundp.ac.be

‡ECARES, Université Libre de Bruxelles. 50, Av Roosevelt, CP114, B1050 Brussels, Belgium. Tel: +3226503075; Fax: +3226504475; david.veredas@ulb.ac.be

We are grateful to Luc Bauwens, Kris Boudt, Laura Coroneo, James Davidson, Niels Haldrup, Marc Hallin, Nikolaus Hautsch, David Hendry, Franz Palm, Mark Steel, Masanobu Taniguchi, Giovanni Urga and the participants at the Humbolt and Namur econometrics seminars, the workshop on the Methods in International Finance Network in Maastricht, the 2007 Brussels-Waseda workshop in statistics in Tokyo, and the 5th OxMetrics User Conference at CASS Business School for insightful remarks. The second and third authors gratefully acknowledges financial support from the Belgian National Bank and the IAP P6/07 contract, from the IAP programme (Belgian Scientific Policy), 'Economic policy and finance in the global economy'. The last two authors are members of ECORE, the recently created association between CORE and ECARES. Any remaining errors and inaccuracies are ours.

1 Introduction

We propose parametric and non parametric residual-based tests for conditional asymmetry. We assume that observations follow a dynamic location-scale model with error term following a law in the class of skewed distributions of Fernández and Steel (1998). Estimation is done under the null hypothesis of constant skewness and the tests are performed on the residuals. These tests are, in spirit, similar to the Breusch-Godfrey test and Engle's ARCH test for the mean and the variance. There are however two important differences. First, the tests presented here are not moment-based. While both the Breusch-Godfrey test and Engle's ARCH test are based on (auto)regressions for the first and second conditional moments, we do not test for (auto)correlation on the third moment. Rather, we test for (auto)correlation in the asymmetry, defined as the ratio between the probabilities of being larger or smaller than the mode. Second, we take into account the uncertainty of the estimated parameters. Indeed, as the tests are residual-based, their distributions depend on the estimated parameters. We rely on Pierce (1982) to derive the correct asymptotic distributions of the tests. Neglecting the uncertainty coming from the first step estimators severely distorts the size of the test, and, hence, the reliability of the corresponding decisions.

From a financial perspective, this article fits within the research that incorporates asymmetry and kurtosis into, among others, asset pricing and portfolio allocation. Harvey and Siddique (2000) propose a modified mean-variance asset pricing model where conditional skewness is priced. The assumption that investors prefer left skewed portfolios to right skewed portfolios justifies the inclusion of the third moment. In their empirical application, they find conditional left skewness to be economically significant, increasing the risk premia by 3.6%. Chabi-Yo, Leisen, and Renault (2006) study the importance of conditional skewness when estimating a skewness premium. Based on Samuelson (1970), they study the optimal asset allocation in a mean-variance-skewness framework. They also introduce a separation theorem based on a fund of the market portfolio, which is a function of the skewness, and analyze the risk compensation in the presence of conditional asymmetry. Dittmar (2002) and Guidolin and Timmermann (2006) expand the standard CAPM to a four-moment CAPM, including skewness and kurtosis. The former shows that its explanatory power for cross-section US stock returns raises considerably. The later explains the home country bias in investments, concluding that, while the standard mean-variance approach assigns 30% of stocks to domestic markets, adding both skewness and kurtosis raises this percentage to 70%.

From an econometric perspective, this article fits into a branch of the literature that proposes tests and models for conditional symmetry. Zheng (1998) presents a nonparametric kernel-based test for conditional symmetry. Bai and Ng (2001) propose a distribution free conditional symmetry test that is valid for non-stationary and non-*i.i.d.* observations. In a similar vein, Delgado and

Escanciano (2007) propose an omnibus test for conditional symmetry in dynamic models. The main feature of this test is that it allows for unknown functional forms of higher conditional moments. Hong and Li (2005) and Egorov, Hong, and Li (2006) present an omnibus nonparametric evaluation test for conditional density models that explicitly takes into account the impact of parameter estimation uncertainty. Hansen (1994) estimates conditional asymmetry and kurtosis by extending the ARMA and GARCH model to the third and fourth order moments. Jondeau and Rockinger (2003) extend the generalized Student-t distribution of Hansen (1994) and investigate the presence of conditional skewness and kurtosis, which are modelled as a function of lagged innovations. In their application to FX markets, they detect the presence and persistence of conditional asymmetry, while the mass on the tails is relatively constant.

As the last two references, our approach is also parametric. We assume that observations follow a dynamic location-scale model with error term following a skewed distribution on the class of Fernández and Steel (1998). Since our focus are financial applications, in the Monte Carlo study and in the empirical application we illustrate our main results on the skewed-t distribution, a skewed version of the Student-t density. This law has a tractable density function, which facilitates maximum likelihood, and four parameters that explain the main features of a probability law for financial returns: location, scale, asymmetry and tail thickness.[1] The skewing mechanism is based on a scale parameter that is introduced inversely on each side of the mode. Asymmetry is therefore defined in terms of the ratio of probabilities of being to the left and right of the mode.

This skewing parameter is our main object of interest. We estimate the model under the null hypothesis of constant asymmetry, i.e. time invariant skewing parameter. Under this hypothesis, the sequence of ratios of the conditional probabilities of being to the left and right of the mode should be independent. Following Engle and Manganelli (2004), these ratios are equivalent to a sequence of indicator functions that take value one if the t-th observation is lower than the mode and zero otherwise. We test for independence of this sequence of binary variables both parametrically and nonparametrically. We first consider Wald-type tests based on generalized linear models. Detailed derivations are provided for two important cases: the linear probability and the logistic regression models. Second, we consider a nonparametric Wald-Wolfowitz runs test. In both cases, we take into account the uncertainty from the estimation of the parameters of the skewed distribution. Using the results from Pierce (1982), we derive the correct asymptotic distribution of the tests under the null.[2]

The method proposed in Pierce (1982) is not the only possible way to account for parameter uncertainty. Engle and Manganelli (2004) propose the in-sample dynamic quantile test, which takes into account the parameter uncertainty from the conditional quantile model. Meddahi and

[1] Hansen (1994) proposes a similar distribution, though the derivation is different.

[2] Tse (2002) also uses Pierce's correction in a residual-based test for conditional heteroskedasticity.

Bontemps (2005) and Meddahi and Bontemps (2006) propose a GMM approach to test the distributional assumption. Both articles are based on moment conditions that are robust to parameter uncertainty. Hong and Lee (2003) also propose a diagnostic test for linear and nonlinear time series models, where the parameter uncertainty has no impact on the limit distribution of the test statistic. Accuracy of out-of-sample predictions can also be affected by parameter uncertainty. Escanciano and Olmo (2008), MacCracken (2000) and West (1996) explore the problem and propose test statistics with a distribution that incorporates the parameter uncertainty.

In a comprehensive Monte Carlo study, we evaluate the properties of the tests for different specifications of the conditional mean and variance, and for different sample sizes. We show that the tests are correctly sized provided that the uncertainty in the parameter estimation is taken into account. We also show that the power of the test is acceptable for a reasonable sample size. As a robustness check, we show the properties of the tests under different misspecifications that the researcher may encounter: overparametrisation of the conditional mean, light misspecification of the conditional variance, and misspecification of the tail index (i.e. assuming it to be constant when the true is time varying). Results show that the size of the tests does not suffer significantly from these departures of the DGP.

Last, using daily log returns of several stocks, indexes and bonds for a period of eleven years, we detect the presence of dynamics in the conditional asymmetry of the standardized residuals of six different series. These findings are confirmed when estimating a dynamic conditional asymmetry such as in Hansen (1994).

The paper is structured as follows. Section 2 introduces the family of skewed distributions of Fernández and Steel (1998) and presents the tests and their asymptotic properties. Since several functional forms for the regression model used for the test are at hand, we start with a linear setting, followed by a logit regression, and we finish with a general form of the regression function where the parameters are estimated by maximum likelihood. We conclude the section with the runs test. Section 3 presents the Monte Carlo study and Section 4 the empirical application. Finally, we present our overall conclusions. Proofs and preparatory results are provided in the Appendix.

2 Testing conditional asymmetry

Let y_t ($t = 1, \ldots, T$) denote a time series of financial returns of a given instrument (e.g. a stock, index or exchange rate) at time t. We assume that this vector of realizations is generated by the following dynamic location-scale model:

$$y_t = \mu_t(\boldsymbol{\phi}|\Omega_{t-1}) + \sigma_t(\boldsymbol{\rho}|\Omega_{t-1}) z_t^*, \tag{1}$$

where Ω_{t-1} is the information set at time $t-1$, $\boldsymbol{\eta} = (\boldsymbol{\phi}, \boldsymbol{\rho})$ is a vector of unknown parameters, and z_t^* is a zero mean and unit variance random variable. The functional forms $\mu_t = \mu_t(\boldsymbol{\phi}|\Omega_{t-1})$ and $\sigma_t = \sigma_t(\boldsymbol{\rho}|\Omega_{t-1})$ are specified according to an ARMA-GARCH type of model.[3] To estimate the model by maximum likelihood, a distributional assumption on the innovation term z_t^* is required. In most financial applications, z_t^* is assumed to be *i.i.d.* $\mathcal{N}(0,1)$ or *i.i.d.* $ST(0,1,\nu)$, $\nu \in \mathbb{R}$ being the degrees of freedom of the Student-t distribution. These distributions are symmetric and cannot capture the skewness that is often observed.

Let $h(z_t;\boldsymbol{\delta})$ be a continuous unimodal and symmetric distribution of a zero mean and unit variance random variable z_t; $\boldsymbol{\delta}$ are parameters present higher-order moments. Fernández and Steel (1998) introduce skewness by changing the scale on each side of the mode through a skewing parameter $\xi \in \mathbb{R}^+$:

$$f_z(z_t;\boldsymbol{\delta},\xi) = \begin{cases} \frac{2}{\xi+\frac{1}{\xi}} h_z\left(\xi z_t;\boldsymbol{\delta}\right) & \text{if} \quad z_t < 0 \\ \frac{2}{\xi+\frac{1}{\xi}} h_z\left(\frac{z_t}{\xi};\boldsymbol{\delta}\right) & \text{if} \quad z_t \geq 0, \end{cases} \tag{2}$$

where z_t has now mean m and variance s^2, depending on (among others) ξ, but has mode at zero. The distribution $f(z_t;\boldsymbol{\delta},\xi)$ is left (right) skewed if $0 < \xi < 1$ ($\xi > 1$). Indeed, by definition (see Fernández and Steel (1998), eq. 3):

$$\frac{P(z_t \geq 0;\boldsymbol{\delta},\xi)}{P(z_t < 0;\boldsymbol{\delta},\xi)} = \xi^2. \tag{3}$$

Clearly, ξ controls the allocation of mass to each side of the mode. In order to obtain z_t^* in (1) we standardize z_t: $z_t^* = (z_t - m)/s$. Note that z_t^* has zero mean, unit variance and density

$$f_{z^*}(z_t^*;\xi,\nu) = \begin{cases} \frac{2s}{\xi+\frac{1}{\xi}} h_{z^*}\left(\xi(z_t^* s + m);\boldsymbol{\delta}\right) & \text{if} \quad z_t^* < -\frac{m}{s} \\ \frac{2s}{\xi+\frac{1}{\xi}} h_{z^*}\left(\frac{(z_t^* s+m)}{\xi};\boldsymbol{\delta}\right) & \text{if} \quad z_t^* \geq -\frac{m}{s}. \end{cases} \tag{4}$$

Additionally, we work with $\ln\xi \in \mathbb{R}$ instead of ξ. A positive (negative) sign for $\ln\xi$ corresponds to a right (left) skewed distribution, while zero indicates symmetry.

In the Monte Carlo study and in the empirical application we consider the particular case of a skewed-t distribution. If $h(z_t;\boldsymbol{\delta})$ is a Student-t density in (2) with $\boldsymbol{\delta} = \nu$, z_t follows a skewed-t distribution or $SKST(m, s^2, \xi, \nu)$ where

$$\begin{aligned} m &= \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\xi - \frac{1}{\xi}\right) \text{ and} \\ s^2 &= \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2 \end{aligned} \tag{5}$$

are the mean and variance. Hence $z_t^* \sim SKST(0,1,\xi,\nu)$ and $y_t \sim SKST(\mu_t, \sigma_t^2, \xi, \nu)$. The skewed-t density has been used a distribution for a GARCH model (Lambert and Laurent (2001b)

[3]More flexibility in the specification of the first two conditional moments could be achieved by including e.g. thresholds, smooth transitions, Markov switching regimes, and stochastic volatility models. This extra flexibility would be in at the expense of estimation.

and Lambert and Laurent (2001a)), for computation of Value at Risk (Giot and Laurent (2003)), indirect estimation of stable densities (Lombardi and Calzolari (2008), Garcia, Renault, and Veredas (2006), and Lombardi and Veredas (2008)) and its multivariate extension has been proposed by Bauwens and Laurent (2005). This is a flexible distribution with four parameters that accounts for the four main features of a distribution: location, scale, asymmetry and tail thickness. It nests important densities that are commonly used in the analysis of financial data, such as the Student-t and the Gaussian.

The $s \times 1$ vector of unknown parameters, $\boldsymbol{\theta} = (\boldsymbol{\eta}', \ln(\xi), \boldsymbol{\delta}')'$, is estimated by maximum likelihood and the asymptotic variance-covariance matrix of the MLE is the inverse of the Fisher information matrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}$. If ξ, assumed to be constant, is truly time invariant, the conditional allocation of mass on either side of the mode of the density of the residuals $z_t(y_t, \hat{\boldsymbol{\theta}}) = \hat{s} z_t^*(y_t, \hat{\boldsymbol{\eta}}) + \hat{m}$, where $z_t^*(y_t, \hat{\boldsymbol{\eta}}) = (y_t - \hat{\mu_t})/\hat{\sigma}_t$, should be independent. Equation (3) evaluated at $z_t(y_t, \hat{\boldsymbol{\theta}})$ can be written as

$$P\left(z_t(y_t, \hat{\boldsymbol{\theta}}) < 0\right) = \frac{1}{1+\hat{\xi}^2} \stackrel{\text{def}}{=} g(\hat{\xi}), \tag{6}$$

and, following Engle and Manganelli (2004), for each observation the probability ratio of being to the left and right of the mode is equivalently represented by the sequence of indicator functions $I\left(z_t(y_t, \hat{\boldsymbol{\theta}}) < 0\right)$ that takes value 1 if the argument is true. From now on we skip the argument y_t in the residual and the indicator function is denoted by $I_t(\hat{\boldsymbol{\theta}})$. The reader, however, should keep in mind the dependence of both functions on y_t and the information set Ω_{t-1} through the estimated parameters.

Under the null hypothesis of constant asymmetry, the sequence of indicator functions should be independent. To test this hypothesis, we take two approaches. The first, parametric, consists on regressing $I_t(\hat{\boldsymbol{\theta}})$ on its past (i.e. $I_{t-1}(\hat{\boldsymbol{\theta}}), I_{t-2}(\hat{\boldsymbol{\theta}}), \ldots$), other past model-based variables potentially related to asymmetry (like past residuals and functions of it), past volatilities, and a set of exogenous variables (e.g. day of the week dummies). We denote by $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$ the $k \times 1$ vector containing these quantities.

The indicator function $I_t(\hat{\boldsymbol{\theta}})$ follows a Bernouilli distribution. That has two implications: (a) Testing for uncorrelation is equivalent to testing for independence, as all moments solely depend on the conditional mean and the number of observations. And (b) several regression functions are available. We rely on Generalized Linear Models, where the conditional expectation of $I_t(\hat{\boldsymbol{\theta}})$ depends on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$ through a link function. We analyze in detail two important cases, linear and logistic link functions (leading to the linear probability and logit models), and we relegate to the end the most general case of a known, but unspecified, link function. The second approach is nonparametric: runs test. A run is a sequence of adjacent equal terms (zeros and ones in our case) and the test checks for non-randomness in this sequence. In all cases, the distributions of the test-statistics incorporate the uncertainty due to $\hat{\boldsymbol{\theta}}$.

Several remarks are in order before moving forward to the next section.

First, our framework is model (1) with z_t^* following a skewed density (4). One may argue that, in the practical applications, rejection of the null hypothesis of independence of $I_t(\hat{\boldsymbol{\theta}})$ could be interpreted as being due to either: (a) a genuine rejection of constant asymmetry, (b) misspecification of μ_t, σ_t or both, or (c) misspecification of the distribution of z_t^*. Thus, to have interpretation (a) as the appropriate one, correct specification of μ_t, σ_t and of the distribution of z_t^* are fundamental assumptions underlying the tests. As for μ_t and σ_t, the tests are built such that these functions can be fairly general. But in practical applications they should be such that $z_t^*(\hat{\boldsymbol{\theta}})$ does not present any dependence in the first two moments. There are available in the literature a sizeable amount of tests for this purpose, which explanation is out of the scope of this article. Notwithstanding, in the Monte Carlo study we analyze the sensitivity of the size and power of the tests to misspecification of μ_t and σ_t. As for (c), and as mentioned previously, the class of Fernández and Steel (1998) is very flexible and nests some relevant distributions for financial data, such as the skewed-t, the skewed-Gaussian and the skewed-stable (Lambert and Laurent (2001b)) distributions. Yet, in the Monte Carlo study we analyze the sensitivity of the tests to misspecification of the tail index (estimation of constant tail index when the true is time varying).

Second, unconditional symmetry does not entail conditional symmetry. If $\ln \hat{\xi} = 0$, unconditionally there is an equal proportion of observations on either side of the mode. However, the conditional probability that the observation at time t is lower than the mode could depend on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$.

Third, one can argue that working with $I_t(\hat{\boldsymbol{\theta}})$ instead of $z_t(\hat{\boldsymbol{\theta}})$ induces a loss of information that could be used to increase the power of the tests. The use of $I_t(\hat{\boldsymbol{\theta}})$ is motivated by the following reasons. (a) In this family of distributions, asymmetry is defined in terms of the ratio of probabilities (3) that can be written as (6). And, following Engle and Manganelli (2004), constant asymmetry is equivalent to requiring that $I_t(\hat{\boldsymbol{\theta}})$ is independent. (b) Tests on $I_t(\hat{\boldsymbol{\theta}})$ do not require any condition on the existence of moments of $z_t(\hat{\boldsymbol{\theta}})$. Testing conditional asymmetry with the later would naturally lead to a regression of its cube on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$, requiring the existence of its first six moments. And (c) the correction in the asymptotic distribution of the tests we propose is fairly simple.

Fourth, besides the potential non desirable causes of rejection of the null hypothesis of the first point above, another potential reason is misspecification of the link function on the regression of $I_t(\hat{\boldsymbol{\theta}})$ on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$. Since $I_t(\hat{\boldsymbol{\theta}})$ is a binary process, many link functions are at hand (linear, logit, probit, etc). In the Monte Carlo study we analyze the sensitivity of the size and power of the parametric tests to this choice. We conclude that differences are insignificant.

2.1 Wald tests

Let

$$E\left(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}});\boldsymbol{\beta}\right) = l(\lambda_t), \tag{7}$$

be the conditional mean of $I_t(\hat{\boldsymbol{\theta}})$. $l(\cdot)$ is the link function and $\lambda_t = m(\hat{\xi}) + \mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta}$ is the linear predictor produced by $m(\hat{\xi})$ -an offset to ensure that under the null (7) equals $g(\hat{\xi})$- and $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$. Many functional forms for $l(\cdot)$ are at hand but we only present detailed results for the linear and logit links: $l(\lambda_t) = \lambda_t$ and $l(\lambda_t) = 1/(1 + \exp(-\lambda_t))$ respectively. As a generalization, we present, at the end of the section, the case of the unspecified, but known, form (7). Since our parameters of interest are only present in the conditional mean, we estimate (7) by quasi-maximum likelihood with the following assumption:

Assumption W1 The conditional probability density function of $I_t(\hat{\boldsymbol{\theta}})$, $r(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))$, belongs to the exponential family.

This assumption ensures consistency of the estimators of $\boldsymbol{\beta}$. Although it could be replaced by an exact knowledge of $r(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))$, in order to gain efficiency, our main results remain unchanged. Under maximum likelihood the so-called sandwich form simplifies to the inverse of the Fisher information but this does not hold in our case. The inclusion of the uncertainty on $\hat{\boldsymbol{\theta}}$ consists of a correction of the Fisher information matrix, the middle part of the sandwich form, making the simplification unfeasible. The specific functional form of $r(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))$ depends on $l(\lambda_t)$, which implies a different support for $I_t(\hat{\boldsymbol{\theta}})$. For instance, for the linear probability model, the conditional probability function is Gaussian while for the logistic conditional mean $r(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))$ is binomial.

Let $\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ be a zero mean statistic which is a function of parameters $\boldsymbol{\beta}$, to be estimated, and of the estimators $\hat{\boldsymbol{\theta}}$ in an explicit way. An example of such a statistic is the score function. Likewise, let $\mathcal{H}_{\boldsymbol{\beta},\hat{\boldsymbol{\theta}}}$ and $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta},\boldsymbol{\theta}}$ be the hessian and the Fisher information matrix respectively that correspond to the conditional likelihood obtained from the data contributions. Note that in the former $\boldsymbol{\theta}$ is estimated while in the later $\boldsymbol{\theta}$ is known. Following Pierce (1982), we assume that the joint distribution of $\hat{\boldsymbol{\theta}}$ and $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is asymptotically Gaussian:

Assumption W2

$$\begin{pmatrix} \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \sqrt{T}\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1} & \mathbf{V}_{12} \\ \mathbf{V}_{12} & \mathbf{V}_{22} \end{pmatrix} \right),$$

This assumption is needed to compute the asymptotic distribution of the statistic $\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ conditional to $\hat{\boldsymbol{\theta}}$. If, for instance, $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the score, $\mathbf{V}_{22} = \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta},\boldsymbol{\theta}}$, the Fisher information matrix. Gaussianity is a mild assumption, justified by some appropriate central limit theorem. The following assumption explicitly establishes the link between $\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$:

Assumption W3 The statistic $\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ can be approximated by

$$\sqrt{T}\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \sqrt{T}\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta}) + \mathbf{B}\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1),$$

where $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is differentiable in $\boldsymbol{\theta}$ and $\mathbf{B} = \lim_{T\to\infty} E\left(\frac{\partial \mathcal{S}(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)$ is a $k \times s$ matrix.

This assumption allows us to compute the variance-covariance matrix of $\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})$ as a combination of the variance-covariance of the statistic as if $\boldsymbol{\theta}$ would be known plus an additional term that captures the uncertainty in $\hat{\boldsymbol{\theta}}$.

For ease of exposition, let us first consider the case where $\boldsymbol{\theta}$ is known. Assuming a linear probability model:

$$P\left(I_t(\boldsymbol{\theta}) = 1 | \, \mathbf{x}_t(\boldsymbol{\theta})\right) = g(\xi) + \mathbf{x}_t(\boldsymbol{\theta})'\boldsymbol{\beta},$$

and, since the indicator function is a Bernoulli random variable, $E(I_t(\boldsymbol{\theta})|\mathbf{x}_t(\boldsymbol{\theta})) = P\left(I_t(\boldsymbol{\theta}) = 1 | \, \mathbf{x}_t(\boldsymbol{\theta})\right)$. Alternatively, $E(Hit_t(\boldsymbol{\theta})|\mathbf{x}_t(\boldsymbol{\theta})) = \mathbf{x}_t(\boldsymbol{\theta})'\boldsymbol{\beta}$ where $Hit_t(\boldsymbol{\theta}) = I_t(\boldsymbol{\theta}) - g(\xi)$. The least squares estimator of β is

$$\sqrt{T}\hat{\boldsymbol{\beta}} = \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t(\boldsymbol{\theta})\mathbf{x}_t'(\boldsymbol{\theta})\right)^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{x}_t(\boldsymbol{\theta})Hit_t(\boldsymbol{\theta})\right). \tag{8}$$

The first term in the right hand side of (8) converges to a nonstochastic $k \times k$ matrix $\mathbf{J}^{-1}$. The second converges in distribution to a Gaussian and, by standard Gauss-Markov assumptions, its asymptotic variance equals $\sigma^2\mathbf{J}$. Hence the Wald test for the null hypothesis $H_0 : \boldsymbol{\beta} = 0$ is

$$W_T^{OLS} = \frac{T}{\sigma^2}\hat{\boldsymbol{\beta}}'\mathbf{J}\hat{\boldsymbol{\beta}} \sim \chi^2_k.$$

In reality, $\boldsymbol{\theta}$ is unknown and must be replaced by an estimator $\hat{\boldsymbol{\theta}}$:

$$P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})) = g(\hat{\xi}) + \mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta}, \tag{9}$$

and the estimated $\boldsymbol{\beta}$ is a function of $\hat{\boldsymbol{\theta}}$, i.e.

$$\sqrt{T}\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t(\hat{\boldsymbol{\theta}})\mathbf{x}_t'(\hat{\boldsymbol{\theta}})\right)^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{x}_t(\hat{\boldsymbol{\theta}})Hit_t(\hat{\boldsymbol{\theta}})\right). \tag{10}$$

This entails an extra source of uncertainty. The following theorem provides the Wald test with the correct asymptotic variance-covariance matrix.

Theorem 1 *Under (1) and (4), the conditional mean (9), assumptions W1-W3, and the null hypothesis* $H_0 : \boldsymbol{\beta} = 0$,

$$W^{OLS} = T\hat{\boldsymbol{\beta}}'\mathbf{J}\left(\sigma^2\mathbf{J} - \mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)^{-1}\mathbf{J}\hat{\boldsymbol{\beta}} \sim \chi^2_k,$$

where

$$\mathbf{B} = -\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\boldsymbol{\theta})\left(b(\xi)h_z(0;\boldsymbol{\delta})\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{y_t=\mu_t - m\frac{\sigma_t}{s}} + \frac{\partial g(\xi)}{\partial \boldsymbol{\theta}'}\right),$$

$b(\xi) = \frac{2}{\xi + \frac{1}{\xi}}$, $h_z(0;\boldsymbol{\delta})$ *is the symmetric density in (2), and* $\frac{\partial g(\xi)}{\partial \boldsymbol{\theta}} = 0$ *except* $\frac{\partial g(\xi)}{\partial \ln \xi} = -\frac{2\xi^2}{(1+\xi^2)^2}$.

To apply the test, some of the quantities have to be replaced by consistent estimators. A consistent estimator for $\mathbf{J}$ is given by the term inside the inverse on the right hand side of (10). Under the null, σ^2 can be estimated by $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} Hit_t(\hat{\boldsymbol{\theta}})^2$. The $k \times s$ matrix $\mathbf{B}$ is evaluated at $\hat{\boldsymbol{\theta}}$. The $s \times s$ variance-covariance matrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}$ can be estimated by the inverse of the outer product of the scores evaluated at $\hat{\boldsymbol{\theta}}$.

The linear probability model has the advantage of simplicity and estimation can be performed by least squares. But it does not constitute an adequate description of the probability of being larger or smaller than the mode as it does not exclude quantities outside the interval $(0, 1)$. An alternative and more natural approach than (9) is a logit link function

$$\ln\left(\frac{P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))}{1 - P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))}\right) = -q(\hat{\xi}) + \mathbf{x}_t(\hat{\boldsymbol{\theta}})'\hat{\boldsymbol{\beta}},$$

equivalently expressed as

$$P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})) = \frac{1}{1 + \exp\left(q(\hat{\xi}) - \mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta}\right)}. \tag{11}$$

The function $q(\hat{\xi}) = \ln\left(\frac{1}{g(\hat{\xi})} - 1\right) = \ln \hat{\xi}^2$ is an intercept that accounts for the fact that, when $\boldsymbol{\beta} = 0$, $P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))$ is not 0.5 but $g(\hat{\xi})$. The conditional probability function involved is Bernoulli:

$$r(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})) = P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}}))^{I_t(\hat{\boldsymbol{\theta}})}(1 - P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})))^{1-I_t(\hat{\boldsymbol{\theta}})}. \tag{12}$$

The following theorem provides the Wald test with the correct asymptotic variance-covariance matrix.

Theorem 2 *Under (1) and (4), the conditional mean (11), assumptions W1-W3, and the null hypothesis* $H_0 : \boldsymbol{\beta} = 0$,

$$W^{Logit} = T\hat{\boldsymbol{\beta}}'\mathbf{J}\left(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)^{-1}\mathbf{J}\hat{\boldsymbol{\beta}} \sim \chi_k^2,$$

where

$$\begin{aligned}
\mathbf{J} &= \frac{1}{T}\mathcal{H}_{\boldsymbol{\beta},\boldsymbol{\theta}}, \\
\mathbf{B} &= -\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\boldsymbol{\theta})\left(b(\xi)h_z(0;\boldsymbol{\delta})\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{y_t=\mu_t-m\frac{\sigma_t}{s}} + \frac{\partial g(\xi)}{\partial\boldsymbol{\theta}'}\right),
\end{aligned}$$

$b(\xi) = \frac{2}{\xi+\frac{1}{\xi}}$, *$h_z(0;\boldsymbol{\delta})$ is the symmetric density in (2), and* $\frac{\partial g(\xi)}{\partial\boldsymbol{\theta}} = 0$ *except* $\frac{\partial g(\xi)}{\partial \ln \xi} = -\frac{2\xi^2}{(1+\xi^2)^2}$.

$\mathbf{B}$ is computed in a similar way as in the linear case and consistent estimators for $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta},\boldsymbol{\theta}}$ and $\mathbf{J}$ are readily available in the form of the inverse of the outer product of the scores of the log likelihood of the density (12) and the sample average Hessian evaluated at $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ respectively.

Finally, we consider the more general case where the link function is known but left unspecified. In this case the statistic $\mathcal{S}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the score function with variance-covariance matrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta},\boldsymbol{\theta}}$.

Theorem 3 *Under (1) and (4), the conditional mean (7), assumptions W1-W3, and the null hypothesis* $H_0 : \boldsymbol{\beta} = 0$,

$$W^M = T\hat{\boldsymbol{\beta}}'\mathbf{J}\left(\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta},\boldsymbol{\theta}} - \mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)^{-1}\mathbf{J}\hat{\boldsymbol{\beta}} \sim \chi^2_k,$$

where

$$\mathbf{J} = \frac{1}{T}\mathcal{H}_{\boldsymbol{\beta},\boldsymbol{\theta}} \ , \ \mathbf{B} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\mathcal{S}_t(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'},$$

and where $\frac{\partial\mathcal{S}_t(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$ *is the contribution to the score of the* $t-th$ *observation.*

These three theorems lead to four remarks.

First, the matrix $\mathbf{B}$ in Theorems 1 and 2 takes the same form. This is due to the treatment of the intercepts. They are chosen such that, under the null, $E(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})) = P(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})) = g(\hat{\xi})$. For a given link function $l(\cdot)$, we can always choose the intercept appropriately such that $\mathbf{B}$ is equal to those of the former theorems. In fact, the linear probability model and the logistic regressions are not the only models that can be used. The probit regression, in which the link function is the inverse of the standard normal cumulative distribution function, is also a good candidate. For any function $0 < F(q(\hat{\xi}) + \mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta}) < 1$, the appropriate intercept $q(\hat{\xi})$ is $F^{-1}(P(I_t(\hat{\boldsymbol{\theta}}))) = F^{-1}(g(\hat{\xi}))$ such that, under the null, $E(I_t(\hat{\boldsymbol{\theta}})|\mathbf{x}_t(\hat{\boldsymbol{\theta}})) = F(F^{-1}(g(\hat{\xi}))) = g(\hat{\xi})$.

Second, the distribution of the tests can be easily adapted to the assumptions about (4) just replacing $h_z(0;\boldsymbol{\delta})$ by the corresponding symmetric distribution. The Monte Carlo study and the empirical application are illustrated assuming $h_z(0;\boldsymbol{\delta})$ a Student-t so (4) is a standardized skewed-t distribution, as in Giot and Laurent (2003) and Lambert and Laurent (2001a). The following Corollary provides the Wald tests for the linear and logistic link functions if $h_z(\cdot;\boldsymbol{\delta})$ is a Student-t pdf.

Corollary 1 *Under (1) and (4),* $h_z(\cdot)$ *a standardized Student-t probability density function, the conditional mean (9) or (11), assumptions W1-W3, and the null hypothesis* $H_0 : \boldsymbol{\beta} = 0$*, the test statistics* W^{OLS} *and* W^{Logit} *remain unchanged but* $h_z(0;\boldsymbol{\delta}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi(\nu-2)}}$.

The Student-t distribution is not the only candidate for $h_z(0;\boldsymbol{\delta})$ in financial applications. The Gaussian, Cauchy, Laplace or GED distributions are also good candidates for which $h_z(0;\boldsymbol{\delta})$ equals $2^{-1/2}$, π^{-1}, $(2\pi)^{-1/2}$ and $\frac{\nu}{\lambda 2^{1+1/\nu}\Gamma(1/\nu)}$ where $\lambda = \sqrt{\Gamma(1/\nu)2^{-2/\nu}/\Gamma(3/\nu)}$ respectively.

Third, if the distribution of y_t is unconditionally symmetric, then the matrix $\mathbf{B}$ simplifies to

$$\mathbf{B} = -\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\boldsymbol{\theta})\left(h_z(0;\boldsymbol{\delta})\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{y_t=\mu_t} + \frac{\partial g(\xi)}{\partial\boldsymbol{\theta}'}\right),$$

where $\frac{\partial g(\xi)}{\partial\boldsymbol{\theta}} = 0$ except $\frac{\partial g(\xi)}{\partial\ln\xi} = -\frac{1}{2}$.

Fourth, the asymptotic variance-covariance matrix of the tests is simple to compute and does not necessarily require numerical methods. The only apparent complication is the derivative of

$z_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ but it has a tractable expression. For instance, if $h_z(\cdot;\boldsymbol{\delta})$ is the skewed-t distribution, $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\rho}, \ln \xi, \nu)$ and

$$\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\frac{s}{\sigma_t}\frac{\partial \mu_t}{\partial \boldsymbol{\phi}} \\ -\frac{s(y_t-\mu_t)}{\sigma_t^2}\frac{\partial \sigma_t}{\partial \boldsymbol{\rho}} \\ \frac{\partial s}{\partial \ln \xi}\frac{y_t-\mu_t}{\sigma_t} + \frac{\partial m}{\partial \ln \xi} \\ \frac{\partial s}{\partial \nu}\frac{y_t-\mu_t}{\sigma_t} + \frac{\partial m}{\partial \nu} \end{pmatrix}.$$

The first two components depend on the specifications for the conditional mean and variance.[4] Analytical formulas of the partial derivatives of m and s with respect to $\ln \xi$ and ν depend on the assumptions about (2). If $h(z_t;\boldsymbol{\delta})$ is a Student-t density:

$$\begin{aligned} \frac{\partial s}{\partial \ln \xi} &= \xi^2 \left(1 - \xi^{-4} - m\frac{\partial m}{\partial \ln \xi}\right) s^{-1} \\ \frac{\partial m}{\partial \ln \xi} &= \left(\xi + \frac{1}{\xi}\right) \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)} \\ \frac{\partial s}{\partial \nu} &= -\frac{m}{s}\frac{\partial m}{\partial \nu} \\ \frac{\partial m}{\partial \nu} &= 0.5\left(\xi - \frac{1}{\xi}\right) \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \left[\frac{1}{\nu-2} + F\left(\frac{\nu+1}{2}\right) - F\left(\frac{\nu}{2}\right)\right], \end{aligned}$$

where $F(x) = \frac{\partial \ln \Gamma(x)}{\partial x}$ is the di-gamma function. Evaluating $\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ at $y_t = \mu_t - m\frac{\sigma_t}{s}$:

$$\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{y_t=\mu_t-m\frac{\sigma_t}{s}} = \begin{pmatrix} -\frac{s}{\sigma_t}\frac{\partial \mu_t}{\partial \boldsymbol{\phi}} \\ \frac{m}{\sigma_t}\frac{\partial \sigma_t}{\partial \boldsymbol{\rho}} \\ -\frac{\partial s}{\partial \ln \xi}\frac{m}{s} + \frac{\partial m}{\partial \ln \xi} \\ -\frac{\partial s}{\partial \nu}\frac{m}{s} + \frac{\partial m}{\partial \nu} \end{pmatrix}.$$

Pre-multiplying by $b(\xi)c(\nu)$ and adding $\frac{\partial g(\xi)}{\partial \theta}$ gives a $s \times 1$ vector that has a simple expression and it is straightforward to program. Transposing it and pre-multiplying by the $k \times 1$ vector $\mathbf{x}_t(\boldsymbol{\theta})$ gives the t-*th* summand of the $k \times s$ matrix $\mathbf{B}$.

2.2 Runs test

The Wald-Wolfowitz runs test (runs test in short) is a nonparametric test that checks for non-randomness in a two-valued data sequence. A run is a sequence of adjacent equal terms. For example, 1111000111001111110000 is divided in six runs, three of which consist of 1's and the others of 0's. In our setting, a run is defined as a sequence of positive (resp. negative) innovations z_t^*. The number of runs, $R_T(\boldsymbol{\theta})$, is obtained as follows:

$$R_T(\boldsymbol{\theta}) = 1 + \sum_{t=2}^{T} \left(I_t(\boldsymbol{\theta}) - I_{t-1}(\boldsymbol{\theta})\right)^2.$$

[4]See Engle (1982), Fiorentini, Calzolari, and Panattoni (1996) and Chung (1999) and Laurent (2004) for a derivation of analytical scores for, respectively, the ARCH, GARCH, ARFIMA-FIGARCH and APARCH models.

If 1's and 0's alternate randomly and $\boldsymbol{\theta}$ is known, the number of runs is a random variable which asymptotic distribution is $\mathcal{N}(a(\boldsymbol{\theta}), b^2(\boldsymbol{\theta}))$ where $a(\boldsymbol{\theta}) = 1+2T_1(\boldsymbol{\theta})T_2(\boldsymbol{\theta})T^{-1}$, $b^2(\boldsymbol{\theta}) = (a(\boldsymbol{\theta})-1)(a(\boldsymbol{\theta})-2)/(T-1)$, $T_2(\boldsymbol{\theta}) = \sum_{t=1}^{T} I_t(\boldsymbol{\theta})$ and $T_1(\boldsymbol{\theta}) = T - T_2(\boldsymbol{\theta})$. Premaratne and Tay (2002) use the runs test in an equivalent setting to ours for testing conditional asymmetry.[5] However, our approach differs from Premaratne and Tay (2002) in several respects. First, they do not take into account the uncertainty involved by the substitution of $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$. Second, they apply the runs test on $\hat{z}_t$ instead of on $\hat{z}_t^*$. Finally, they consider the skewed Student distribution of Hansen (1994) for which the link between the asymmetry parameter and the runs is not immediate.

Let $\mathcal{S}(\hat{\boldsymbol{\theta}}) = \frac{1}{T-1}(R_T(\hat{\boldsymbol{\theta}}) - a(\hat{\boldsymbol{\theta}}))$ be a zero mean statistic that is a function of $\hat{\boldsymbol{\theta}}$ in an explicit way. Similarly to the Wald tests, we assume that the joint distribution of $\hat{\boldsymbol{\theta}}$ and $\mathcal{S}(\boldsymbol{\theta})$ is asymptotically Gaussian:

Assumption R1

$$\begin{pmatrix} \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \sqrt{T}\mathcal{S}(\boldsymbol{\theta}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1} & \mathbf{V}_{12} \\ \mathbf{V}_{12} & \mathbf{V}_{22} \end{pmatrix}\right),$$

where $V_{22} = \frac{T}{(T-1)^2} b^2(\boldsymbol{\theta})$. This assumption is needed to compute the asymptotic distribution of the statistic $\mathcal{S}(\hat{\boldsymbol{\theta}})$ conditional to $\hat{\boldsymbol{\theta}}$. We also need an equivalent assumption to W2 that allows us to compute the variance-covariance matrix of $\mathcal{S}(\hat{\boldsymbol{\theta}})$ as a combination of the variance-covariance matrix of the statistic as if $\boldsymbol{\theta}$ were known and an additional term that captures the uncertainty in $\hat{\boldsymbol{\theta}}$.

Assumption R2 The statistic $\mathcal{S}(\hat{\boldsymbol{\theta}}) = (R_T(\hat{\boldsymbol{\theta}}) - a(\hat{\boldsymbol{\theta}}))$ can be approximated by

$$\sqrt{T}\mathcal{S}(\hat{\boldsymbol{\theta}}) = \sqrt{T}\mathcal{S}(\boldsymbol{\theta}) + \mathbf{B}\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1),$$

where $\mathcal{S}(\boldsymbol{\theta})$ is differentiable in $\boldsymbol{\theta}$, and $\mathbf{B} = \lim_{T\to\infty} E\left(\frac{\partial \mathcal{S}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)$ is a $1 \times s$ vector.

The following theorem provides the runs test with the correct asymptotic variance-covariance matrix:

Theorem 4 *Under (1) and (4), assumptions R1 and R2, and the null hypothesis of no conditional asymmetry in* $z_t(\hat{\boldsymbol{\theta}})$*,*

$$Runs = \frac{R_T(\hat{\boldsymbol{\theta}}) - a(\hat{\boldsymbol{\theta}})}{\sqrt{T\left(\frac{b^2(\hat{\boldsymbol{\theta}})}{T} - \mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)}} \sim \mathcal{N}(0,1),$$

where

$$\mathbf{B} \quad = \quad -\lim_{T\to\infty} \frac{2}{T-1} b(\xi) h_z(0;\boldsymbol{\delta}) \sum_{t=2}^{T} (g(\xi) - I_{t-1}) \left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{y_t = \mu_t - m\frac{\sigma_t}{s}} - g(\xi) \left.\frac{\partial z_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{y_1 = \mu_1 - m\frac{\sigma_1}{s}},$$

[5] An alternative runs test for asymmetry is McWilliams (1990).

Consistent estimators for $\mathbf{B}$ and $\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}$ are computed in a similar way as in the previous theorems. Notice however that $\mathbf{B}$ is now a row vector of dimension $1 \times s$. Last, as in the Wald tests, the distribution of the test can be easily adapted to the assumptions about (4) just replacing $h_z(0; \boldsymbol{\delta})$ by the corresponding symmetric distribution.

3 Monte Carlo study

In this section we first study the size and power of the three tests and compare the results with those of the test without the corrections (by setting $\mathbf{B}$ to $\mathbf{0}$). Second we assess the robustness of the tests to mean, variance and tail index misspecification.

In the first part of the Monte Carlo analysis, all data generating processes (DGP) belong to the general specification given in (1) but differ in the way the conditional mean and variance are specified. For the conditional mean we consider two scenarios: an AR(0) and an AR(1) (with AR parameter 0.1); both without intercept. We also consider two scenarios for the conditional variance: an homocedastic model ($\sigma_t^2 = 1$) and a GARCH(1,1) model ($\sigma_t^2 = 0.1 + 0.1\varepsilon_{t-1} + 0.8\sigma_{t-1}^2$). The distribution of the error term z_t^* is a standardized skewed-t with skewness parameter exp(0.08) and 8 degrees of freedom: $z_t^* \sim SKST(0, 1, \exp(0.08), 8)$.

Given a DGP, $S(= 1000)$ samples of T (=500, 1000 or 2000) observations are generated and a model specification is estimated for each sample. The DGP is always estimated. We apply the three tests presented earlier and consider both the corrected and uncorrected versions of the Wald tests. The uncorrected versions of the tests are denoted with a subindex U in the tables. For example, W^{OLS} indicates that the correction was made in the test based on the linear probability model while W_U^{OLS} corresponds to the Wald test that ignores the uncertainty in $\hat{\boldsymbol{\theta}}$. For the parametric tests we consider two alternative hypotheses corresponding to two sets of covariates. The first consists of the first two lags of the residuals, i.e. $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}}), \hat{\varepsilon}_{t-2}(\hat{\boldsymbol{\theta}}))'$, and the second consists of the first lags of the residuals and the Hit variable, i.e. $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}}), Hit_{t-1}(\hat{\boldsymbol{\theta}}))'$. If the tests are properly sized, the alternative hypothesis should be wrongly selected with an empirical frequency close to $\alpha \times 100\%$ (where α is the nominal size of the test) when S is sufficiently large. In fact, when $S = 1000$, one expects 99% of the empirical sizes to lie within $(3.3\%, 6.8\%)$ or $(7.6\%, 12.5\%)$ when $\alpha = 5\%$ and 10% respectively.

Tables 1 and 2 show the empirical sizes for $\alpha = 5\%$ and 10% respectively and for different combinations of the specifications of the conditional mean and variance. Numbers in bold indicate that the test is statistically undersized while numbers in italic indicate that the test is statistically oversized. The first two columns show the conditional mean and variance while the remaining columns show the tests with and without the correction. When the conditional mean is constant, all tests have the appropriate size. However, if the conditional mean is time varying, the uncorrected tests are systematically undersized, whatever the alternative hypothesis. For small sample sizes, corrected tests do not always provide satisfactory results (as suggested for instance by the cases AR(1)-GARCH(0,0); due to the fact that the degrees of freedom have a smaller rate of convergence than the other parameters) but for any other sample size the corrected tests display no significant size distortion. The presence of conditional variance does not affect the size of the tests, either corrected or not. For instance, the sizes for all tests with constant conditional mean are correct regardless of whether the process is homocedastic or not. Last, the sizes for W^{OLS} and W^{Logit} are very similar, indicating that the test is not sensitive to the parametric assumption of the link function

Table 3 shows the empirical powers for $\alpha = 5\%$ (top panel) and 10% (bottom panel). The true conditional mean and conditional variance are the same as above but z_t^* now follows a SKST(0,1,ξ_t,8) with $\ln \xi_t = 0.08 + 0.2\varepsilon_{t-1}$ (top sub panel) or $\ln \xi_t = 0.08 + 0.2 z_{t-1}^*$ (bottom sub panel).[6] To compute the power, the correct AR-GARCH model is estimated but constant asymmetry is assumed. The first two columns of the table indicate the true conditional mean and variance. The third column indicates the sample size. The next four columns show the power for the W^{OLS} and W^{Logit} tests with two different regressors under the alternative: $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = \hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}})$ and $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = Hit_{t-1}(\hat{\boldsymbol{\theta}})$). The last column is the power of the *Runs* test. Note that results for AR(1)-GARCH(0,0) under the two specifications for $\ln \xi_t$ are the same: Under $\sigma_t = 1$ ε_{t-1} equals z_{t-1}^*. Note as well that we only report the power of the unbiased tests (cfr. Tables 1 and 2). Results for the case with constant mean are not reported either.

As expected, the powers of W^{OLS} and W^{Logit} are very similar (meaning that the choice of the link function is not crucial) and they increase with the sample size. For an appropriate alternative hypothesis, the parametric tests have good power and it is higher than the runs test. In fact, the power of the runs test is comparable to the power of the parametric test when the alternative hypothesis inappropriately speculates on the form of the dynamics in the asymmetry (for instance when $\ln \xi_t$ depends on ε_{t-1} while the parametric tests consider $\mathbf{x_t} = Hit_{t-1}$); yet they increase with the sample size. Finally, the presence of GARCH effects slightly decreases the power of the tests.

In the second part of the Monte Carlo study we provide some robustness checks to mean, variance and tail index misspecifications. As for the conditional mean, we study the effect of overparameterisation in the tests: an AR(1) model is estimated while the DGP has no dynamics in the conditional mean and the conditional variance is either homocesdastic or a GARCH(1,1). Results are reported in the rows AR(1) of Table 4. As for the conditional variance, we study the case where the true conditional variance follows a APARCH(1,1) model: $\sigma_t^{1.2} = 0.1 + 0.05\left(|\varepsilon_{t-1}| - 0.1\varepsilon_{t-1}\right)^{1.2} + 0.8\sigma_{t-1}^{1.2}$ but the estimated model in a GARCH(1,1). Note that although

[6] Note that in this case, m and s in (4) must be replaced by m_t and s_t as they depend on ξ_t.

both models capture the volatility clustering, the GARCH model does not explain the long memory and the leverage effect of the APARCH model.[7] Results are reported in the rows GARCH(1,1) of Table 4. We also consider the case where the tail index is time-varying, $\nu_t = 8 + 0.1 z^*_{t-1}$, but wrongly asumed constant (ν). Results are reported in rows ν_t of Table 4. In all cases, the alternative for the parametric tests is $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}}), \hat{\varepsilon}_{t-2}(\hat{\boldsymbol{\theta}}))'$ and α equals 5% (top panel) and 10% (bottom panel).

Top panel shows that the uncorrected tests are systematically undersized if the estimated model for the mean is conditional, even if the DGP has constant mean. This dovetails with results in Tables 1 and 2, which show that the uncorrected tests are undersized whenever the estimated mean follows an AR model. By contrast, misspecification of the conditional variance does not seem to affect the tests, which also dovetails with the results of Tables 1 and 2, for which the sizes for the tests, for a given conditional mean, are similar regardless of whether the process is homocedastic or not. Last, the sizes of the tests are not affected if the tail index presents conditional behaviour but it is wrongly assumed constant.

4 Empirical Application

We apply our tests to a basket of daily returns consisting of several stocks, indices and bonds, namely AT&T INC. (AT), Du Pont De Nemours (DUPONT), the NIKKEI 225 index (NIKKEI), the FTSE 100 index (FTSE), the HANG SENG index (HSI), the Swiss Market Index (SMI), the Trade Weighted Exchange Index of major currencies (TWEI) and the 30-Years US TREASURY Bond (TBIL30). Data have been obtained from Yahoo Finance, except for TWEI, downloaded from the Federal Reserve Economic Data website, and cover the period 1995-2006; roughly 3000 observations.

The model we consider is an AR(m)-APARCH(1,1) with skewed-t distribution:

$$\begin{aligned}
y_t &= \mu + \sum_{i=1}^{m} \phi_i (y_{t-i} - \mu) + \varepsilon_t \\
\varepsilon_t &= \sigma_t z_t^* \\
z_t^* &\sim SKST(0, 1, \ln(\xi), \nu) \\
\sigma_t^\zeta &= \omega + \alpha\left(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1}\right)^\zeta + \beta\sigma_{t-1}^\zeta,
\end{aligned}$$

where $\boldsymbol{\theta} = (\mu, \rho_1, \ldots, \rho_m, \omega, \alpha_1, \beta_1, \gamma,\ \zeta,\ \ln(\xi),\ \nu)$ are the unknown parameters to be estimated. The AR orders have been selected in such a way that the innovation term does not present any sign

[7] More serious departures from the true volatility process are possible, such as simulating from the APARCH but assuming constant variance. We believe, however, that this is not a realistic case. While the leverage or the long memory of the volatility may be unnoticed at first, the presence of volatility clustering is evident from visual inspection of data.

of serial correlation. The APARCH model is flexible enough and captures the volatility clustering, long memory and leverage effect, three stylized facts that often found in financial returns.

Top panel of Table 5 contains the MLE estimates with standard errors in parenthesis. Results, in terms of the conditional mean and variance, are expected: returns are little, if any, autocorrelated, the conditional volatility presents clustering, leverage and long memory. Moreover, the estimated degrees of freedom are low (except for FTSE) indicating the presence of fat tails in the innovations.

Since the tests are based on the assumption of correct location and scale parametrizations, we first test the ability of the AR-APARCH model to fit the conditional mean and variance. We compute the Box-Pierce statistics with 10 lags on the standardized residuals to test for serial correlation, denoted by $Q(10)$ in the middle panel. We also test for autocorrelation in the squared residuals computing the same statistic, denoted by $Q2(10)$.[8] Finally, we apply the test for conditional heteroscedasticity of Tse (2002). This test involves regressing the demeaned squared standardized residuals on its past: $\hat{z}_t^{*2} - 1 = \delta_1 \hat{z}_{t-1}^{*2} + \ldots + \delta_M \hat{z}_{t-M}^{*2} + u_t$ (we set $M = 10$), in the same spirit as the Engle's ARCH test. But, since the regressors are estimated, standard inference is invalid. Tse (2002) derives the asymptotic distribution of the estimated parameters and shows that a joint test of significance of the $\delta_1, \ldots, \delta_M$ is $\chi^2(M)$ distributed. The p-values of the three tests are reported in the middle panel. As a whole, $Q(10)$, $Q2(10)$ and Tse's test show that there is no evidence of misspecification of the first two conditional moments.

Out of eight series, five (all except NIKKEI, AT and HSI) appear to be unconditionally skewed as $\ln(\xi)$ is significantly different from 0. We apply the three tests for conditional asymmetry presented in Section 2. The first line of bottom panel of Table 5 shows the p-value of W^{Runs}. Second to last rows show the p-values of the W^{OLS} test for a set of educated guess of regressors in the residual equation (Because of the similarity between W^{OLS} and W^{Logit} we only report the p-values associated W^{OLS}). Unlike W^{Runs}, W^{OLS} and W^{Logit} require the choice of a set of variables which we suspect possess information for explaining conditional asymmetry. A set of sensitive candidates are $\varepsilon_{t-1}(\hat{\boldsymbol{\theta}})$, $\varepsilon_{t-1}^2(\hat{\boldsymbol{\theta}})$, $\varepsilon_{t-1}^3(\hat{\boldsymbol{\theta}})$, $\sigma_{t-1}^2(\hat{\boldsymbol{\theta}})$ and its lags.

Three series do not present any evidence of conditional asymmetry: FTSE, SMI and TBIL30, meaning that the p-values of W^{Runs} and W^{OLS} (for any set of regressors) are larger than 0.05. The p-value for W^{Runs} is lower than 0.05 for NIKKEI, HSI and TWEI, while W^{OLS} detects the presence of conditional asymmetry in AT and DUPONT. The most interesting case is HSI.

[8] It is standard practice in the literature to approximate the distribution of these statistics under the null of homoscedasticity by a $\chi^2(10-2)$, where the value 2 is related to the number of ARCH and GARCH terms of the ARCH-type model. Although it has been noted that the portmanteau statistics do not have an asymptotic χ^2 distribution, many authors, nonetheless, apply the χ^2 distribution as an approximation (the problem lies in the fact that estimated residuals are used to calculate the portmanteau statistics).

Unconditionally the index is symmetric but conditionally it is not.[9] This effect was already emphasized in the introduction: symmetry has to be seen, in this context, as a special case of asymmetry. But it does not imply that conditional and unconditional (a)symmetry are equal. In fact, HSI shows that, conditional to past information, the distribution is not symmetric, but it is unconditionally symmetric.

The usefulness of the parametric tests is twofold: they have more power than the runs test in certain situations (see Section 3) and they give us an idea of the source conditional asymmetry. This information is particularly useful if one wants to model the conditional asymmetry in a similar way than Hansen (1994) or Harvey and Siddique (2000), i.e. by making the conditional asymmetry parameter ($\ln(\xi)$ in our case) time-varying. A deeper inspection of the W^{OLS} suggests the following models for $\ln(\xi_t)$:

$$\begin{aligned}
\ln(\xi_t) &= \ln(\xi) + \tau_1 \varepsilon_{t-1} \text{ for NIKKEI,} \\
\ln(\xi_t) &= \ln(\xi) + \tau_2 \varepsilon_{t-2} \text{ for FTSE,} \\
\ln(\xi_t) &= \ln(\xi) + \tau_1 \varepsilon_{t-1} + \tau_3 \varepsilon_{t-1}^3 \text{ for AT, HSI and TWEI, and} \\
\ln(\xi_t) &= \ln(\xi) + \tau_1 \varepsilon_{t-1} + \tau_4 \sigma_{t-1}^2 \text{ for DUPONT.}
\end{aligned}$$

For the remaining two series (SMI and TBIL30), we estimate Hansen's (1994) specification:

$$\ln(\xi_t) = \ln(\xi) + \tau_1 \varepsilon_{t-1} + \tau_5 \varepsilon_{t-1}^2.$$

Since there was no evidence of conditional asymmetry for these series, we expect τ_1 and τ_5 not to be significantly different from 0. Estimation results are in Table 6.[10] Top panel pertains to the MLE while the bottom panel contains the log-likelihood of the models under the null and the alternatives, and a likelihood ratio test (LRT) of the null hyphotesis of no conditional asymmetry.

Results are compatible with the estimated parameters of the model. Conditional asymmetry is detected in all but two series, SMI and TBIL30. The specification choice suggested by the parametric test is relevant, as estimated parameters are significantly different from zero. For instance, ε_{t-1} was found to be a good predictor of the conditional asymmetry for the NIKKEI and its coefficient is indeed highly significant. Similarly, ε_{t-2} appears to be significant while ε_{t-1}^2 is insignificant. The same comments apply to all the series, except for AT, where ε_{t-1}^3 does not explain the conditional asymmetry.[11]

[9] The use of the terms conditionally and unconditionally may be seen by the reader like a bit lousy. We allow for this lack of precision for the sake exposition. By unconditionally we refer, in the context of asymmetry, to a model where the skewness parameter is constant over time but the mean and variance can be conditional.

[10] We have estimated various specifications for the conditional asymmetry. We found that the best model (in terms of goodness-of-fit) was compatible with our prior.

[11] Note that concerning HSI, while τ_3 (coefficient of ε_{t-1}^3) is only significant at the 10% level using a t-test, it is found to be significant at the 1% level through a LRT. This difference is due to the correlation between ε_{t-1} and ε_{t-1}^3.

5 Conclusion

While the bulk of quantitative financial research has focused on the location-scale, other features present in the data are meaningful and deserve study. Among these, skewness, measured either as the third moment or as the degree of asymmetry around the mode, has received increasing attention. However, a rigorous test for conditional asymmetry, comparable to the Breusch-Godfrey test and Engles ARCH test for the first two moments, was missing. The present article fills this gap and presents a residual-based test for conditional asymmetry. Assuming that the true density function falls within the class of skewed distributions of Fernández and Steel (1998), we propose three tests – two parametric and one nonparametric – based on the residuals. The tests are estimated in a second step after initial estimation of the distribution under the null hypothesis of constant asymmetry. The fact that there is a prior estimation to the computation of the tests creates additional uncertainty that must be taken into account. Using the general results of Pierce (1982) -which others have drawn on, such as Tse (2002)- we compute the asymptotic distribution of the tests that includes a variance correction for the uncertainty of the prior estimation. An application to a basket of daily financial time series confirms the presence of dynamics in the conditional asymmetry.

Appendix

In this final Section we proof Theorems 1 to 4. We start with a preparatory lemma that makes use of one property of the Dirac function: the integral of a Dirac function times a function equals the function evaluated at zero.

Lemma A.1 *Let*

$$\frac{\partial I_t(\boldsymbol{\theta})}{\partial z_t(\boldsymbol{\theta})}\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f_z\left(z_t;\boldsymbol{\theta}\right)$$

where $\frac{\partial I_t(\boldsymbol{\theta})}{\partial z_t(\boldsymbol{\theta})}$ *is the negative value of a Dirac function,* $z_t(\boldsymbol{\theta}) = s\frac{y_t-\mu_t}{\sigma_t} + m$, y_t *is defined in (1), and* $\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f_z\left(z;\boldsymbol{\theta}\right)$ *is a continuous function in* $\boldsymbol{\theta}$*. Then*

$$\int_{-\infty}^{+\infty} \frac{\partial I_t(\boldsymbol{\theta})}{\partial z_t(\boldsymbol{\theta})}\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f_z\left(z_t;\boldsymbol{\theta}\right) dz_t = -\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{y_t=\mu_t-m\frac{\sigma_t}{s}} f_z\left(0;\boldsymbol{\theta}\right)$$

Proof of Lemma A.1 It follows by the properties of the Dirac function. We need the value of the residual that makes $\frac{\partial I_t(\boldsymbol{\theta})}{\partial z_t(\boldsymbol{\theta})}$ equal to $-\infty$: $z_t(\boldsymbol{\theta}) = 0$. Since $z_t(\boldsymbol{\theta}) = s\frac{y_t-\mu_t}{\sigma_t} + m$ and $y_t = \mu_t + \sigma_t\frac{z_t-m}{s}$, $z_t(\boldsymbol{\theta}) = 0$ is equivalent to $z_t = 0$ or $y_t = \mu_t - m\frac{\sigma_t}{s}$.

Proof of Theorem 1 This proof partially relies on Tse (2002). The first term in the right hand side of (8) converges to a nonstochastic $k \times k$ matrix $\mathbf{J}^{-1}$. We need to show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t(\hat{\boldsymbol{\theta}})Hit_t(\hat{\boldsymbol{\theta}}),$$

in (10) converges in distribution to a Gaussian distribution with zero mean and variance-covariance matrix $\sigma^2\mathbf{J} - \mathbf{B}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\theta})\mathbf{B}'$. Let $\sqrt{T}\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{x}_t(\hat{\boldsymbol{\theta}}) Hit_t(\hat{\boldsymbol{\theta}})$ so the statistic to apply Pierce's (1982) theorem is $\mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t(\hat{\boldsymbol{\theta}}) Hit_t(\hat{\boldsymbol{\theta}})$. Taking derivatives with respect to $\hat{\boldsymbol{\theta}}$:

$$\frac{\partial \mathcal{S}(\boldsymbol{\beta}, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\partial \mathbf{x}_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} Hit_t(\hat{\boldsymbol{\theta}}) + \mathbf{x}_t(\hat{\boldsymbol{\theta}}) \frac{\partial Hit_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right). \tag{13}$$

Taking expectations in one of the summands, the first term of the right hand side is, by the law of iterated expectations:

$$E\left(\frac{\partial \mathbf{x}_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} Hit_t(\hat{\boldsymbol{\theta}}) \right) = E_{\mathbf{x}_t}\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) E\left(Hit_t(\hat{\boldsymbol{\theta}}) \middle| \mathbf{x}_t(\hat{\boldsymbol{\theta}}) \right) \right) = 0$$

since $E(Hit_t(\hat{\boldsymbol{\theta}})) = 0$ as, under the null, it does not depend on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$. As for the second term, by the law of iterated expectations:

$$E\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) \frac{\partial Hit_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right) = E\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) \left(\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - \frac{\partial g(\hat{\xi})}{\partial \hat{\boldsymbol{\theta}}} \right) \right) = E_{\mathbf{x}_t}\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) E\left(\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - \frac{\partial g(\hat{\xi})}{\partial \hat{\boldsymbol{\theta}}} \middle| \mathbf{x}_t(\hat{\boldsymbol{\theta}}) \right) \right)$$

The derivative of $I_t(\hat{\boldsymbol{\theta}})$ with respect to $\hat{\boldsymbol{\theta}}$ equals,

$$\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial z_t(\hat{\boldsymbol{\theta}})} \frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}.$$

The first term in the right hand side is a Dirac function: It takes value zero for all $z_t(\hat{\boldsymbol{\theta}}) \neq 0$ and $-\infty$ for $z_t(\hat{\boldsymbol{\theta}}) = 0$. Taking conditional expectations:

$$E\left(\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \middle| \mathbf{x}_t(\hat{\boldsymbol{\theta}}) \right) = \int_{-\infty}^{+\infty} \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial z_t(\hat{\boldsymbol{\theta}})} \frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} f_z(z_t|\mathbf{x}_t(\hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}}) dz_t,$$

The dependence on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$ is made explicit because the Dirac function, under the alternative, depends on $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$. The expectation is taken with respect to z_t, the source of randomness.

To solve this integral we make use Lemma A.1:

$$\begin{aligned} E\left(\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \middle| \mathbf{x}_t(\hat{\boldsymbol{\theta}}) \right) &= -\left. \frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right|_{z_t(\hat{\boldsymbol{\theta}})=0} f_z\left(0|\mathbf{x}_t(\hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}} \right) \\ &= -\left. \frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi}) h_z\left(0|\mathbf{x}_t(\hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\delta}} \right) \end{aligned}$$

where $b(\hat{\xi}) = \frac{2}{\hat{\xi} + \frac{1}{\hat{\xi}}}$ and the last equality comes from (2). Grouping terms and under the null:

$$E\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) \frac{\partial Hit_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right) = -E_{\mathbf{x}_t}\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) \left. \frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \right|_{y_t = \hat{\mu}_t - \hat{m}\frac{\hat{\sigma}_t}{\hat{s}}} b(\hat{\xi}) h_z\left(0; \hat{\boldsymbol{\delta}} \right) \right) - E_{\mathbf{x}_t}\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}}) \frac{\partial g(\hat{\xi})}{\partial \hat{\boldsymbol{\theta}}} \right)$$

Under a suitable form of the law of large number, we can replace the expectation operator by the sample mean:

$$E\left(\frac{\partial \mathcal{S}(\boldsymbol{\beta},\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right) = -\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\hat{\boldsymbol{\theta}})\left(\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{y_t=\hat{\mu}_t-\hat{m}\frac{\sigma_t}{\hat{s}}} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)+\frac{\partial g(\hat{\xi})}{\partial \hat{\boldsymbol{\theta}}}\right),$$

and by W2

$$\mathbf{B} = -\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\boldsymbol{\theta})\left(\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{y_t=\mu_t-m\frac{\sigma_t}{s}} b(\xi)h_z\left(0;\boldsymbol{\delta}\right)+\frac{\partial g(\xi)}{\partial \boldsymbol{\theta}'}\right),$$

which implies, by W1,

$$\sqrt{T}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\beta}(\boldsymbol{\theta}), \mathbf{J}^{-1}\left(\sigma^2\mathbf{J}-\mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)\mathbf{J}^{-1}\right).$$

A direct application of the Wald test yields the result.

Proof of Theorem 2 The mean log-likelihood is given by

$$\mathcal{L}(\beta,\hat{\boldsymbol{\theta}}) = -\frac{1}{T}\sum_{t=1}^{T}\left((1-I_t(\hat{\boldsymbol{\theta}}))(-q(\hat{\xi})+\mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta}) - \ln\left(1+\exp(q(\hat{\xi})-\mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta})\right)\right),$$

and the score

$$\mathcal{S}(\beta,\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T}\left(-(1-I_t(\hat{\boldsymbol{\theta}}))\mathbf{x}_t(\hat{\boldsymbol{\theta}}) + \frac{\exp(q(\hat{\xi})-\mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta})}{1+\exp(q(\hat{\xi})-\mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta})}\mathbf{x}_t(\hat{\boldsymbol{\theta}})\right)$$

has variance-covariance matrix $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}}$. After some simplifications

$$\mathcal{S}(\beta,\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\hat{\boldsymbol{\theta}})\left(I_t(\hat{\boldsymbol{\theta}}) - \frac{1}{1+\exp(q(\hat{\xi})-\mathbf{x}_t(\hat{\boldsymbol{\theta}})'\boldsymbol{\beta})}\right).$$

Under the null,[12]

$$\begin{aligned}\frac{\partial \mathcal{S}(\beta,\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} &= \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_t(\hat{\boldsymbol{\theta}})\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} + I_t(\hat{\boldsymbol{\theta}})\frac{\partial \mathbf{x}_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - g(\hat{\xi})\frac{\partial \mathbf{x}_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - \mathbf{x}_t(\hat{\boldsymbol{\theta}})\frac{\partial g(\hat{\xi})}{\partial \hat{\boldsymbol{\theta}}}\right)\\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial \mathbf{x}_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}Hit_t(\hat{\boldsymbol{\theta}}) + \mathbf{x}_t(\hat{\boldsymbol{\theta}})\frac{\partial Hit_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right),\end{aligned}$$

where the second equality come from $Hit_t(\hat{\boldsymbol{\theta}}) = I_t(\hat{\boldsymbol{\theta}}) - g(\hat{\xi})$ and $\frac{\partial Hit_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - \frac{\partial g(\hat{\xi})}{\partial \hat{\boldsymbol{\theta}}}$. Since this derivative is equivalent to (13),

$$\mathbf{B} = -\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t(\boldsymbol{\theta})\left(\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{y_t=\mu_t-m\frac{\sigma_t}{s}} b(\xi)h_z\left(0;\boldsymbol{\delta}\right)+\frac{\partial g(\xi)}{\partial \boldsymbol{\theta}'}\right),$$

which implies, by W1,

$$\sqrt{T}\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \sim \mathcal{N}\left(\boldsymbol{\beta}(\boldsymbol{\theta}), \mathbf{J}^{-1}\left(\sigma^2\mathbf{J}-\mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)\mathbf{J}^{-1}\right).$$

A direct application of the Wald test yields the result.

[12] Recall that $\frac{1}{1+\exp(q(\hat{\xi}))} = g(\hat{\xi})$.

Proof of Theorem 3 It follows from W1 and W2 since $Var(\mathcal{S}(\beta,\hat{\boldsymbol{\theta}})) = \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta}} + \mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'$.

Proof of Theorem 4 Let $\mathcal{S}(\hat{\boldsymbol{\theta}}) = \frac{1}{T-1}\mathcal{S}^*(\hat{\boldsymbol{\theta}})$ where $\mathcal{S}^*(\hat{\boldsymbol{\theta}}) = R_T(\hat{\boldsymbol{\theta}}) - a^*(\hat{\boldsymbol{\theta}})$ where $a^*(\hat{\boldsymbol{\theta}}) = 1 + \frac{2}{T-1}\left((T-1) - \sum_{t=2}^T I_t(\boldsymbol{\theta})\right)\sum_{t=2}^T I_t(\boldsymbol{\theta})$, which is equivalent to $a(\hat{\boldsymbol{\theta}})$ as $T \to \infty$. Developing:

$$\mathcal{S}^*(\hat{\boldsymbol{\theta}}) = 1+\sum_{t=2}^T \left(I_t^2(\hat{\boldsymbol{\theta}}) - 2I_t(\hat{\boldsymbol{\theta}})I_{t-1}(\hat{\boldsymbol{\theta}}) + I_{t-1}^2(\hat{\boldsymbol{\theta}})\right) - 1 - \frac{2}{T-1}\left((T-1) - \sum_{t=2}^T I_t(\boldsymbol{\theta})\right)\sum_{t=2}^T I_t(\boldsymbol{\theta}).$$

As $I_t^2(\hat{\boldsymbol{\theta}}) = I_t(\hat{\boldsymbol{\theta}})$ and as $T \to \infty$ $\sum_{t=1}^T I_t(\hat{\boldsymbol{\theta}}) \approx \sum_{t=1}^T I_{t-1}(\hat{\boldsymbol{\theta}})$, $\mathcal{S}^*(\hat{\boldsymbol{\theta}})$ simplifies to

$$\mathcal{S}^*(\hat{\boldsymbol{\theta}}) = \frac{2}{T-1}\left(\sum_{t=2}^T I_t(\hat{\boldsymbol{\theta}})\right)^2 - 2\sum_{t=2}^T I_t(\hat{\boldsymbol{\theta}})I_{t-1}(\hat{\boldsymbol{\theta}}).$$

Taking derivatives with respect to $\hat{\boldsymbol{\theta}}$:

$$\begin{aligned}\frac{\partial \mathcal{S}^*(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} &= \frac{4}{T-1}\left(\sum_{t=2}^T I_t(\hat{\boldsymbol{\theta}})\right)\left(\sum_{t=2}^T \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right) - 2\sum_{t=2}^T \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} I_{t-1}(\hat{\boldsymbol{\theta}}) - I_t(\hat{\boldsymbol{\theta}})\frac{\partial I_{t-1}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} \\ &= 2\sum_{t=2}^T\left(2g(\hat{\xi})\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} - \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} I_{t-1}(\hat{\boldsymbol{\theta}}) - I_t(\hat{\boldsymbol{\theta}})\frac{\partial I_{t-1}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right). \qquad (14)\end{aligned}$$

We take expectations to each component of the sum in the LHS. The first term equals, by Lemma A.1,

$$E\left(2g(\hat{\xi})\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right) = -2g(\hat{\xi})\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)$$

since $\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = \frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial z_t(\hat{\boldsymbol{\theta}})}\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}$ and (2). The expectation of one of the summands in the second term in the LHS of (14) equals, by the law of iterated expectations and Lemma A.1:

$$\begin{aligned}E\left(\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} I_{t-1}(\hat{\boldsymbol{\theta}})\right) &= E_{I_{t-1}}\left(E\left(\frac{\partial I_t(\hat{\boldsymbol{\theta}})}{\partial z_t(\hat{\boldsymbol{\theta}})}\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} I_{t-1}(\hat{\boldsymbol{\theta}})\right| I_{t-1}\right)\right) \\ &= -E_{I_{t-1}}\left(I_{t-1}(\hat{\boldsymbol{\theta}})\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)\right).\end{aligned}$$

The last equality is by Lemma A.1. Equivalently, for one of the summands in the third term in the LHS of (14):

$$\begin{aligned}E\left(\frac{\partial I_{t-1}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} I_t(\hat{\boldsymbol{\theta}})\right) &= E_{I_{t-1}}\left(E\left(\frac{\partial I_{t-1}(\hat{\boldsymbol{\theta}})}{\partial z_t(\hat{\boldsymbol{\theta}})}\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} I_t(\hat{\boldsymbol{\theta}})\right| I_{t-1}\right)\right) \\ &= E_{I_{t-1}}\left(\frac{\partial I_{t-1}(\hat{\boldsymbol{\theta}})}{\partial z_t(\hat{\boldsymbol{\theta}})}\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} E(I_t(\hat{\boldsymbol{\theta}})|I_{t-1})\right) \\ &= -g(\hat{\xi})\left.\frac{\partial z_{t-1}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_{t-1}(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)\end{aligned}$$

where last equality is by Lemma A.1. Grouping terms:

$$
\begin{aligned}
E\left(\frac{\partial \mathcal{S}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right) &= -\frac{2}{T-1}\sum_{t=2}^{T} 2g(\hat{\xi})\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right) \\
&+ \frac{2}{T-1}\sum_{t=2}^{T} E_{I_{t-1}}\left(I_{t-1}(\hat{\boldsymbol{\theta}})\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)\right) \\
&+ \frac{2}{T-1}\sum_{t=2}^{T} g(\hat{\xi})\left.\frac{\partial z_{t-1}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_{t-1}(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)
\end{aligned}
$$

Under a suitable form of the law of large number, we can replace the expectation operator by the sample mean:

$$
\begin{aligned}
E\left(\frac{\partial \mathcal{S}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right) &= -\frac{2}{T-1}\sum_{t=2}^{T} 2g(\hat{\xi})\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right) \\
&+ I_{t-1}(\hat{\boldsymbol{\theta}})\left.\frac{\partial z_t(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_t(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right) \\
&+ g(\hat{\xi})\left.\frac{\partial z_{t-1}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right|_{z_{t-1}(\hat{\boldsymbol{\theta}})=0} b(\hat{\xi})h_z\left(0;\hat{\boldsymbol{\delta}}\right)
\end{aligned}
$$

and by R2

$$
\mathbf{B} = -\lim_{T\to\infty}\frac{2}{T-1}b(\xi)h_z\left(0;\boldsymbol{\delta}\right)\sum_{t=2}^{T}(g(\xi)-I_{t-1})\left.\frac{\partial z_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{y_t=\mu_t-m\frac{\sigma_t}{s}} - g(\xi)\left.\frac{\partial z_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{y_1=\mu_1-m\frac{\sigma_1}{s}}.
$$

Hence $\mathcal{S}(\hat{\boldsymbol{\theta}}) \sim \mathcal{N}\left(0, \frac{T}{(T-1)^2}b^2(\hat{\boldsymbol{\theta}}) - \mathbf{B}\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta}}^{-1}\mathbf{B}'\right)$, which yields the result.

References

Bai, J., and S. Ng (2001): "A Consistent Test for Conditional Symmetry in Time Series Models," *Journal of Econometrics*, 103, 225–158.

Bauwens, L., and S. Laurent (2005): "A New Class of Multivariate Skew Densities, with Application to GARCH Models," *Journal of Business and Economic Statistics*, 23, 3–30.

Chabi-Yo, F., D. Leisen, and E. Renault (2006): "Implications of Asymmetry Risk for Portfolio Analysis and Asset Pricing," Mimeo.

Chung, C.-F. (1999): "Estimating the Fractionally Integrated GARCH Model," National Taïwan University working paper.

Delgado, M., and J. Escanciano (2007): "Nonparametric Tests for Conditional Symmetry in Dynamic Models," *Journal of Econometrics*.

Dittmar, R. (2002): "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns," *Journal of Finance*, 57, 369–403.

Egorov, A., Y. Hong, and H. Li (2006): "Validating Forecasts of the Joint Probability Density of Bond Yields: Can Affine Models Beat Random Walk?," *Journal of Econometrics*, 135, 255–284.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R., and S. Manganelli (2004): "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22, 367–381.

Escanciano, J., and J. Olmo (2008): "Backtesting Parametric Value-at-Risk with Estimation Risk," Mimeo.

Fernández, C., and M. Steel (1998): "On Bayesian Modelling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

Fiorentini, G., G. Calzolari, and L. Panattoni (1996): "Analytic Derivatives and the Computation of GARCH Estimates," *Journal of Applied Econometrics*, 11, 399–417.

Garcia, R., E. Renault, and D. Veredas (2006): "Estimation of Stable Distributions by Indirect Inference," CORE DP 2006/112.

Giot, P., and S. Laurent (2003): "Value-at-Risk for Long and Short Positions," *Journal of Applied Econometrics*, 18, 641–664.

Guidolin, M., and A. Timmermann (2006): "International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences," Mimeo.

Hansen, B. (1994): "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

Harvey, C., and A. Siddique (2000): "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263–1295.

Hong, Y., and T. Lee (2003): "Diagnostic Checking for the Adequacy of Nonlinear Time Series Models," *Econometric Theory*, 19, 1065–1121.

Hong, Y., and H. Li (2005): "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates," *Review of Financial Studies*, 18, 37–84.

Jondeau, E., and M. Rockinger (2003): "Conditional volatility, skewness, and kurtosis: existence, persistence and comovements," *Journal of Economic Dynamics & Control*, 27, 1699 1737.

Lambert, P., and S. Laurent (2001a): "Modelling Financial Time Series Using GARCH-Type Models and a Skewed Student Density," Discussion Paper 01-25, Institut de Statistique, Louvain-la-Neuve.

——— (2001b): "Modelling Skewness Dynamics in Series of Financial Data Using Skewed Location-scale Distributions," Discussion Paper 01-19, Institut de Statistique, Louvain-la-Neuve.

Laurent, S. (2004): "Analytical derivates of the APARCH model," *Computational Economics*, 24, 51–57.

Lombardi, M., and G. Calzolari (2008): "Indirect Inference for α-Stable Distributions and Processes," *Econometrics Journal*, 11, 193–208.

Lombardi, M., and D. Veredas (2008): "Indirect Estimation of Elliptical Stable Distributions," *Computational Statistics and Data Analysis*.

MacCracken, M. (2000): "Robust Out-of-Sample Inference," *Journal of Econometrics*, 99, 195–223.

McWilliams, T. (1990): "A Distribution-Free Test for Symmetry Based on a Runs Statistic," *Journal of the American Statistical Association*, 85, 1130–1133.

Meddahi, N., and C. Bontemps (2005): "Testing Normality: A GMM Approach," *Journal of Econometrics*, 124, 149–186.

——— (2006): "Testing Distributional Assumptions: A GMM Approach," Mimeo.

Pierce, D. (1982): "The Asymptotic Effect of Substituting Estimator for Parameters in Certain Types of Statistics," *Annals of Statistics*, 10, 475–478.

Premaratne, G., and A. Tay (2002): "How Should We Interpret Evidence of Time Varying Conditional Skewness?," Mimeo.

Samuelson, P. (1970): "The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variance and Higher Moments," *The Review of Economic Studies*, 37, 537–542.

Tse, Y. (2002): "Residual-based Diagnostics for Conditional Heteroskedasticity Models," *Econometrics Journal*, 5, 358–373.

West, K. (1996): "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067–1084.

Zheng, J. (1998): "Consistent Specification Testing for Conditional Symmetry," *Econometric Theory*, 14, 139–149.

Table 1: Monte Carlo study: Size for $\alpha = 0.05$

Mean	Variance	T	W_U^{OLS}	W^{OLS}	W_U^{Logit}	W^{Logit}	$Runs_U$	$Runs$
			$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2})'$					
AR(0)	GARCH(0,0)	500	6.00	5.90	5.90	6.50	5.60	5.70
		1000	5.80	5.80	5.60	6.40	6.60	6.60
		2000	3.90	3.90	3.90	4.70	5.20	5.20
AR(1)	GARCH(0,0)	500	**1.70**	*8.40*	**1.60**	*7.90*	**0.60**	6.20
		1000	**1.40**	4.30	**1.20**	4.70	**0.40**	3.90
		2000	**2.20**	4.60	**2.00**	4.70	**0.60**	5.30
AR(0)	GARCH(1,1)	500	4.00	4.10	3.60	4.20	4.60	5.10
		1000	5.10	5.10	5.20	5.90	4.90	5.00
		2000	5.50	5.50	5.40	6.00	5.40	5.40
AR(1)	GARCH(1,1)	500	**1.90**	6.40	**1.70**	6.40	**0.90**	6.10
		1000	**1.20**	5.10	**1.20**	5.10	**0.30**	5.10
		2000	**2.00**	6.10	**2.10**	6.10	**1.20**	5.90
			$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}, Hit_{t-1}(\hat{\boldsymbol{\theta}}))'$					
AR(0)	GARCH(0,0)	500	4.70	4.70	4.60	5.70	5.60	5.70
		1000	5.10	5.10	5.00	6.00	6.60	6.60
		2000	4.30	4.30	4.30	4.70	5.20	5.20
AR(1)	GARCH(0,0)	500	**2.40**	*9.70*	**2.60**	*9.00*	**0.60**	6.20
		1000	**1.90**	4.40	**1.70**	4.20	**0.40**	3.90
		2000	**1.50**	4.60	**1.70**	4.80	**0.60**	5.30
AR(0)	GARCH(1,1)	500	4.40	4.40	4.40	4.90	4.60	5.10
		1000	4.70	4.60	5.00	5.60	4.90	5.00
		2000	5.60	5.60	5.20	5.90	5.40	5.40
AR(1)	GARCH(1,1)	500	**1.90**	6.30	**2.10**	6.30	**0.90**	6.10
		1000	**1.90**	5.20	**2.10**	5.40	**0.30**	5.10
		2000	**2.10**	6.10	**2.00**	6.30	**1.20**	5.90

Empirical frequency (in percentages) of incorrect rejection of the null hypothesis of constant skewness for nominal size $\alpha = 0.05$ and $\ln \xi = 0.08$. First two columns indicate the true conditional mean and variance. The distribution for z_t^* is $SKST(0, 1, \exp(0.08), 8)$. Third column indicates the sample size. The remaining six columns show the size for the different tests. The subindex U indicates that the test is uncorrected. For each horizontal block a different set of regressors under the alternative is considered, gathered in $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$, for the parametric tests. Bold (italic) values correspond to an undersized (oversized) test.

Table 2: Monte Carlo study: Size for $\alpha = 0.10$

Mean	Variance	T	W_U^{OLS}	W^{OLS}	W_U^{Logit}	W^{Logit}	$Runs_U$	$Runs$
			$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2})'$					
AR(0)	GARCH(0,0)	500	10.30	10.30	10.10	11.10	9.40	9.50
		1000	11.00	11.00	11.10	11.50	*13.10*	*13.30*
		2000	8.90	8.90	8.60	9.70	9.30	9.30
AR(1)	GARCH(0,0)	500	**4.70**	*15.50*	**4.30**	*12.30*	**3.40**	11.30
		1000	**4.20**	9.70	**4.10**	9.60	**2.10**	8.30
		2000	**4.40**	9.60	**4.30**	9.30	**3.20**	9.40
AR(0)	GARCH(1,1)	500	9.00	8.90	8.40	9.70	9.40	9.70
		1000	9.50	9.60	9.60	10.80	9.50	9.60
		2000	10.90	10.90	10.80	11.40	11.30	11.30
AR(1)	GARCH(1,1)	500	**4.60**	*12.60*	**4.50**	*12.50*	**2.80**	11.60
		1000	**3.00**	10.00	**3.10**	10.60	**1.90**	9.90
		2000	**4.20**	10.20	**4.30**	10.20	**3.30**	10.7
			$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}, Hit_{t-1}(\hat{\boldsymbol{\theta}}))'$					
AR(0)	GARCH(0,0)	500	9.00	9.20	8.60	9.80	9.40	9.50
		1000	11.00	11.00	11.10	12.10	*13.10*	*13.30*
		2000	8.90	8.90	8.50	9.60	9.30	9.30
AR(1)	GARCH(0,0)	500	**4.90**	*15.80*	**5.30**	*12.30*	**3.40**	11.30
		1000	**4.20**	8.70	**3.80**	9.40	**2.10**	8.30
		2000	**4.00**	10.10	**3.30**	9.50	**3.20**	9.40
AR(0)	GARCH(1,1)	500	8.80	8.90	8.50	9.70	9.40	9.70
		1000	10.30	10.30	10.20	10.70	9.50	9.60
		2000	11.10	11.10	11.60	12.40	11.30	11.30
AR(1)	GARCH(1,1)	500	**4.10**	10.80	**3.80**	10.90	**2.80**	11.60
		1000	**4.20**	10.50	**4.20**	10.50	**1.90**	9.90
		2000	**4.30**	10.60	**4.30**	10.90	**3.30**	10.7

Empirical frequency (in percentages) of incorrect rejection of the null hypothesis of constant skewness for nominal size $\alpha = 0.10$ and $\ln \xi = 0.08$. First two columns indicate the true conditional mean and variance. The distribution for z_t^* is $SKST(0, 1, \exp(0.08), 8)$. Third column indicates the sample size. The remaining six columns show the size for the different tests. The subindex U indicates that the test is uncorrected. For each horizontal block a different set of regressors under the alternative is considered, gathered in $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$, for the parametric tests. Bold (italic) values correspond to an undersized (oversized) test.

Table 3: Monte Carlo study: Power

Mean	Variance	T	W^{OLS}	W^{Logit}	W^{OLS}	W^{Logit}	*Runs*
			$\alpha = 0.05$				
			$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = \hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}})$		$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = Hit_{t-1}(\hat{\boldsymbol{\theta}})$		
$\ln \xi_t = 0.08 + 0.2\varepsilon_{t-1}$							
AR(1)	GARCH(0,0)	500	32.7	30.5	12.5	12.9	10.8
		1000	52.8	51.4	20.5	20.7	18.5
		2000	78.8	77.2	40.3	40.8	39.0
AR(1)	GARCH(1,1)	500	29.3	29.4	12.9	12.5	12.0
		1000	41.1	47.5	21.0	21.9	19.4
		2000	70.0	69.2	35.0	35.7	34.2
$\ln \xi_t = 0.08 + 0.2z^*_{t-1}$							
AR(1)	GARCH(0,0)	500	32.3	30.5	12.5	12.9	10.8
		1000	52.9	51.4	20.5	20.7	18.5
		2000	78.9	77.2	40.3	40.8	39.0
AR(1)	GARCH(1,1)	500	28.9	28.2	12.8	13.7	11.5
		1000	46.5	46.0	22.8	22.0	20.7
		2000	72.9	72.0	36.3	36.3	35.1
			$\alpha = 0.10$				
			$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = \hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}})$		$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = Hit_{t-1}(\hat{\boldsymbol{\theta}})$		
$\ln \xi_t = 0.08 + 0.2\varepsilon_{t-1}$							
AR(1)	GARCH(0,0)	500	42.0	39.1	21.4	21.8	20.1
		1000	63.4	62.1	30.5	31.2	29.1
		2000	86.9	86.1	50.7	51.2	49.5
AR(1)	GARCH(1,1)	500	39.3	37.6	20.6	20.5	19.1
		1000	61.4	60.8	31.7	31.5	30.3
		2000	81.0	80.0	47.6	47.7	46.9
$\ln \xi_t = 0.08 + 0.2z^*_{t-1}$							
AR(1)	GARCH(0,0)	500	42.0	39.1	21.4	21.8	20.1
		1000	63.4	62.1	30.5	31.2	29.1
		2000	86.9	86.1	50.7	51.2	49.5
AR(1)	GARCH(1,1)	500	38.1	38.2	19.6	20.6	18.3
		1000	59.0	58.3	30.4	30.6	28.8
		2000	80.6	80.3	49.6	49.6	48.0

Empirical frequency (in percentages) of correct rejection of the null hypothesis of constant skewness for nominal sizes $\alpha = 0.05$ (top panel) and $\alpha = 0.10$ (bottom panel). The distribution for z^*_t is $SKST(0, 1, \xi_t, 8)$. Each panel is divided in two sub panels. In the top sub panel the true conditional asymmetry is $\ln \xi_t = 0.08 + 0.2\varepsilon_{t-1}$ while in the bottom sub panel $\ln \xi_t = 0.08 + 0.2z^*_{t-1}$. First two columns indicate the true conditional mean and variance. Third column indicates the sample size. The next four columns show the power for the W^{OLS} and W^{Logit} tests with two different regressors under the alternative: $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = \hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}})$ and $\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = Hit_{t-1}(\hat{\boldsymbol{\theta}})$. The last column is the power of the *Runs* test.

Table 4: Monte Carlo study: Robustness check

	T	W_U^{OLS}	W^{OLS}	W_U^{Logit}	W^{Logit}	$Runs_U$	$Runs$
				$\alpha = 0.05$			
				$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}}), \hat{\varepsilon}_{t-2}(\hat{\boldsymbol{\theta}}))'$			
AR(1)	500	**1.80**	5.60	**1.70**	*5.30*	**0.50**	3.80
	1000	**1.60**	6.60	**1.60**	6.10	**1.70**	5.70
	2000	**1.60**	5.30	**1.60**	4.90	**1.30**	4.40
AR(1)	500	**1.90**	5.90	**2.00**	6.10	**0.70**	5.30
	1000	**1.40**	5.80	**1.40**	6.20	**1.40**	6.70
	2000	**2.60**	5.30	**2.60**	5.50	**1.00**	4.90
GARCH(1,1)	500	**1.60**	5.70	**1.70**	6.00	**0.50**	5.90
	1000	**1.40**	4.70	**1.40**	5.20	**0.40**	3.70
	2000	**1.30**	5.40	**1.40**	5.70	**0.50**	4.80
ν_t	500	**1.90**	4.80	**1.50**	5.30	**0.70**	5.00
	1000	**2.00**	5.20	**2.00**	5.80	**0.80**	4.90
	2000	**2.00**	5.50	**2.10**	5.80	**0.90**	6.00
				$\alpha = 0.10$			
				$\mathbf{x}_t(\hat{\boldsymbol{\theta}}) = (\hat{\varepsilon}_{t-1}(\hat{\boldsymbol{\theta}}), \hat{\varepsilon}_{t-2}(\hat{\boldsymbol{\theta}}))'$			
AR(1)	500	**3.40**	11.30	**3.30**	11.00	**2.70**	8.90
	1000	**4.40**	11.20	**4.30**	11.20	**3.80**	11.40
	2000	**3.60**	10.90	**3.50**	10.50	**2.90**	8.60
AR(1)	500	**3.20**	10.10	**3.60**	10.30	**2.50**	9.70
	1000	**3.60**	10.10	**3.60**	10.40	**1.70**	11.00
	2000	**3.90**	10.20	**3.70**	9.90	**2.80**	10.80
GARCH(1,1)	500	**3.20**	11.00	**3.20**	11.50	**2.20**	8.90
	1000	**3.90**	11.90	**4.30**	12.20	**3.10**	11.80
	2000	**3.40**	10.40	**3.70**	10.60	**2.80**	8.90
ν_t	500	**4.50**	11.80	**4.40**	11.50	**2.40**	9.40
	1000	**4.60**	10.50	**4.20**	10.70	**2.30**	9.40
	2000	**4.50**	11.20	**4.50**	11.00	**2.70**	10.80

Empirical frequency (in percentages) of incorrect rejection of the null hypothesis of constant skewness for nominal sizes $\alpha = 0.05$ (top panel) and $\alpha = 0.10$ (bottom panel). The distribution for z_t^* is $SKST(0, 1, \exp(0.08), 8)$. First column indicates the model component that is misspecified. In the first 3 rows of each panel the true conditional variance is $\sigma_t^2 = 1$ and the conditional mean is 0 but it is estimated as an AR(1). In the rows 4 to 6 the true conditional variance is the GARCH(1,1) used in previous tables and the conditional mean is 0 but it is estimated as an AR(1). In the rows 7 to 9 the true conditional mean is 0 and the conditional variance follows an APARCH model but it is estimated as a GARCH(1,1). Rows 10 to 12 present the case where the conditional mean is 0 and the conditional variances is $\sigma_t^2 = 1$. The true tail index is conditional: $\nu_t = 8 + 0.1 z_{t-1}^*$, but wrongly assumed constant: $\nu_t = \nu$. Second column indicates the sample size. The remaining six columns show the size for the different tests. The subindex U indicates that the test is uncorrected. For each horizontal block a different set of regressors under the alternative is considered, gathered in $\mathbf{x}_t(\hat{\boldsymbol{\theta}})$, for the parametric tests. Bold (italic) values correspond to an undersized (oversized) test.

Table 5: Estimation results, diagnosis and tests

	NIKKEI	FTSE	AT	HSI	DUPONT	SMI	TWEI	TBIL30
μ	-0.001	0.017	0.052	0.050	0.051	0.043	0.003	-0.030
	(0.023)	(0.013)	(0.025)	(0.022)	(0.026)	(0.013)	(0.007)	(0.015)
ϕ_1				0.049	0.006			
				(0.018)	(0.018)			
ϕ_2					-0.039			
					(0.018)			
ω	0.030	0.011	0.007	0.016	0.009	0.022	0.002	0.009
	(0.007)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)	(0.001)	(0.004)
α_1	0.069	0.057	0.055	0.058	0.048	0.087	0.040	0.037
	(0.010)	(0.008)	(0.012)	(0.010)	(0.010)	(0.011)	(0.008)	(0.008)
β_1	0.925	0.943	0.954	0.941	0.958	0.911	0.962	0.958
	(0.010)	(0.007)	(0.011)	(0.008)	(0.009)	(0.011)	(0.007)	(0.010)
γ	0.584	0.879	0.142	0.530	0.244	0.573	0.118	0.194
	(0.116)	(0.136)	(0.090)	(0.107)	(0.106)	(0.087)	(0.112)	(0.118)
ζ	1.107	1.006	1.191	1.307	1.033	1.024	1.150	1.479
	(0.197)	(0.181)	(0.255)	(0.189)	(0.236)	(0.184)	(0.343)	(0.362)
ν	9.772	24.177	7.584	7.208	7.689	11.969	8.517	8.743
	(1.565)	(8.624)	(0.999)	(0.959)	(0.966)	(2.118)	(1.276)	(1.241)
$\ln \xi$	-0.021	-0.129	0.036	0.004	0.071	-0.122	-0.053	0.062
	(0.026)	(0.027)	(0.025)	(0.024)	(0.026)	(0.027)	(0.025)	(0.027)
				Diagnosis				
$Q(10)$	0.785	0.336	0.478	0.194	0.811	0.755	0.434	0.115
$Q2(10)$	0.380	0.233	0.282	0.524	0.468	0.298	0.710	0.871
Tse	0.100	0.122	1.000	0.227	0.462	0.357	0.999	0.959
				Tests				
W^{Runs}	0.033	0.849	0.133	0.034	0.866	0.209	0.011	0.461
W^{OLS} :								
ε_{t-1}	0.038	0.678	0.016	0.004	0.168	0.678	0.059	0.354
ε_{t-1},ε_{t-2}	0.035	0.051	0.053	0.007	0.041	0.833	0.166	0.468
ε^2_{t-1}	0.545	0.610	0.735	0.699	0.325	0.164	0.847	0.328
ε^3_{t-1}	0.191	0.509	0.666	0.034	0.891	0.557	0.545	0.665
σ^2_{t-1}	0.656	0.739	0.606	0.523	0.010	0.333	0.640	0.704
ε_{t-1},ε^2_{t-1}	0.100	0.794	0.050	0.012	0.242	0.337	0.158	0.367
ε_{t-1},σ^2_{t-1}	0.106	0.865	0.048	0.013	0.017	0.574	0.148	0.596
ε_{t-1},ε^3_{t-1}	0.117	0.803	0.004	0.011	0.238	0.841	0.009	0.650

Top panel shows the estimation results of the model with constant asymmetry. Standard errors in parenthesis. Middle panel shows the diagnosis for the conditional mean and variance. $Q(10)$, $Q2(10)$ and Tse correspond respectively to the p-value of the Box-Pierce statistics with 10 lags on the standardized and squared standardized residuals, and the Tse test for conditional heteroscedasticity with 10 lags. Bottom panel shows the p-values of the W^{Runs} test and the W^{OLS} test under different alternatives. W^{Logit} tests are not shown because of the similarity with W^{OLS}.

Table 6: Estimation results with conditional asymmetry

	NIKKEI	FTSE	AT	HSI	DUPONT	SMI	TWEI	TBIL30
μ	0.002	0.017	0.051	0.053	0.046	0.045	0.002	-0.031
	(0.023)	(0.014)	(0.025)	(0.024)	(0.025)	(0.015)	(0.006)	(0.015)
ρ_1				0.084	0.004			
				(0.021)	(0.019)			
ρ_2					-0.035			
					(0.017)			
ω	0.030	0.011	0.006	0.017	0.007	0.021	0.001	0.009
	(0.007)	(0.002)	(0.005)	(0.005)	(0.004)	(0.004)	(0.001)	(0.004)
α_1	0.071	0.059	0.053	0.060	0.046	0.086	0.043	0.038
	(0.009)	(0.006)	(0.019)	(0.010)	(0.012)	(0.010)	(0.010)	(0.007)
β_1	0.921	0.941	0.955	0.937	0.959	0.910	0.959	0.956
	(0.009)	(0.007)	(0.017)	(0.009)	(0.011)	(0.011)	(0.009)	(0.011)
γ	0.562	0.898	0.126	0.554	0.220	0.570	0.116	0.169
	(0.127)	(0.121)	(0.093)	(0.103)	(0.115)	(0.078)	(0.125)	(0.129)
ζ	1.143	1.026	1.197	1.300	1.040	1.028	1.262	1.521
	(0.189)	(0.190)	(0.231)	(0.197)	(0.231)	(0.173)	(0.377)	(0.464)
ν	9.991	23.76	7.480	7.383	7.685	11.95	8.730	8.866
	(1.741)	(9.296)	(0.982)	(1.036)	(1.067)	(2.918)	(1.280)	(1.478)
$\ln(\xi)$	-0.023	-0.138	0.038	0.010	-0.027	-0.129	-0.057	0.054
	(0.032)	(0.026)	(0.024)	(0.023)	(0.047)	(0.029)	(0.024)	(0.042)
$\tau_1(\varepsilon_{t-1})$	0.060	0.026	0.053	0.092	0.028	0.034	0.343	0.042
	(0.017)	(0.022)	(0.014)	(0.020)	(0.011)	(0.024)	(0.062)	(0.056)
$\tau_2(\varepsilon_{t-2})$		0.073						
		(0.020)						
$\tau_3(\varepsilon_{t-1}^3)$			-0.001	-0.847			-0.101	
			(0.000)	(0.483)			(0.027)	
$\tau_4(\sigma_{t-1}^2)$					0.015			
					(0.012)			
$\tau_5(\varepsilon_{t-1}^2)$						0.008		0.006
						(0.006)		(0.037)
Log-lik H_0	-5020	-3933	-5758	-4983	-5732	-4163	-1382	-3790
Log-lik H_1	-5014	-3928	-5750	-4971	-5728	-4162	-1371	-3789
LR	0.000	0.006	0.000	0.000	0.016	0.286	0.000	0.237

Top panel shows the estimation results of the model with conditional asymmetry. Standard errors in parenthesis. Bottom panel shows the log-likelihoods for the model with constant asymmetry (Log-lik H_0), with conditional asymmetry (Log-lik H_1), and the p-value of the log-likelihood ratio test. The test is χ_k^2 under the null, where $k = 1$ or 2 depending on the model.