Testing Conditional Dynamics in Asymmetry.

A Residual-Based Approach

Philippe Lambert
Université de Liège

Sébastien Laurent
Facultés Universitaires Notre-Dame de la Paix, Namur and CORE, Université Catholique de Louvain

David Veredas
ECARES, Université Libre de Bruxelles and CORE, Université Catholique de Louvain

ECARES working paper 2008_009
TESTING CONDITIONAL ASYMMETRY.

A RESIDUAL-BASED APPROACH

PHILIPPE LAMBERT *
SÉBASTIEN LAURENT †
DAVID VEREDAS ‡

Abstract

We propose three residual-based tests for conditional dynamic asymmetry. Estimation is performed under the null hypothesis of constant asymmetry of the innovations and, in a second step, the tests are performed either through a parametric model or a nonparametric method (runs). The working distribution is assumed to fall into the class of skewed distributions of Fernández and Steel (1998) for which asymmetry is measured by the ratio between the probabilities of being larger and smaller than the mode. 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 skewness, asymmetry, residuals.

JEL Classification: C32, G14, E44

* Méthodes quantitatives en sciences sociales, Université de Liège. Boulevard du Rectorat 7, B4000 Lige, Belgium. Tel: +3243665990; Fax: +3243664751; p.lambert@ulg.ac.be
† CeReFim, University of Namur, 8, rempart de la virge 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, Laura Coroneo, James Davidson, Marc Hallin, Niels Haldrup, Nikolaus Hautsch, David Hendry, Franz Palm, Masanobu Taniguchi, Giovanni Urga and the participants at the Humboldt and Namur econometrics seminars, 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 dynamics in the asymmetry: estimation is done under the null hypothesis of constant skewness and the test is performed on the residuals. The working distribution is assumed to be a member of the class of skewed distributions of Fernández and Steel (1998). These tests are similar to the Breusch-Godfrey test and Engle’s ARCH test for the mean and the variance. There are however two important differences. First of all, the tests presented here are not moments-based. While both the Breusch-Godfrey test and Engle’s ARCH test are based on autoregressions for the first and second conditional moments, we do not test for autocorrelation on the third moment. Rather, we test for dynamics in the asymmetry of the distribution, as measured by the ratio between the probabilities of being larger or smaller than the mode. Second of all, and most importantly, we take into account the uncertainty of the estimated parameters. Indeed, as the test is residual-based, it depends on the estimation of the distribution derived under the hypothesis of constant asymmetry. We rely on Pierce (1982) to derive the correct asymptotic distribution of the test. Neglecting the uncertainty coming from the first step estimators would severely distort the size of the test, and, hence, the reliability of the corresponding decisions.

This article fits into the empirical and theoretical area of financial research that incorporates asymmetry and kurtosis into, among others, portfolio allocation and risk management. Until recently, the analysis of financial returns has been indivisibly, and often exclusively, linked to the analysis of volatility. If markets are efficient and investors make choices consistent with a quadratic utility function, then returns are uncorrelated and Gaussian, which implies both symmetry and that all even orders are a function of the volatility, the only source of risk aversion. This paradigm has been well established since the sixties and forms the basis for all modern portfolio analysis. Ironically, the first study on the effect of higher order moments in portfolio analysis comes from Samuelson (1970). On the first page of this seminal paper, he states that In economics, the relevant probability distributions are not nearly Gaussian, and quadratic utility in the large leads to well-known absurdities. This statement seems surprising when we consider that it was written at a time when portfolio analysis was still in its infancy and high frequency data was not yet available. Samuelson (1970) characterizes the risk aversion of any investor by the preference scheme that the investor assigns to the moments of the distribution, thus distancing himself from the mean-variance approach.¹

More recently, Harvey and Siddique (2000) have proposed a modified mean-variance asset pricing model where conditional skewness is priced. The assumption that investors prefer left

¹However, he states that this approach is valid if portfolio returns are compact, that is if the risk of a position declines as the holding period becomes shorter and shorter.
skewed portfolios to right skewed portfolios justifies including the third moment. In their empirical application, they find conditional left skewness to be economically significant, increasing the risk premia by 3.6%. Dittmar (2002) expands the standard CAPM to a four-moment CAPM, including skewness and kurtosis, and shows that its explanatory power for cross-section US stock returns raises considerably. Similarly, Guidolin and Timmermann (2006) explain the home country bias in investments using a modified international CAPM where the preferences of investors are a function of the skewness and kurtosis, as in Dittmar (2002), furthermore allowing the quantity and price of risk to follow a regime switching process. In their empirical analysis of international stocks, they conclude that, while the standard mean-variance approach assigns 30% of stocks to domestic markets, adding both skewness and kurtosis, and the regime switching mechanism, raises this percentage to 70%. As in Harvey and Siddique (2000), 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, a new portfolio which is a function of the skewness, and analyze the risk compensation in the presence of asymmetry.

Risk management, and in particular Value at Risk, also considers asymmetry in its analysis. From a distributional viewpoint, taking skewness into account moves us away from the Gaussian paradigm and towards a skewed distribution instead. Among all existing unimodal skewed distributions, two have received particular attention. The family of stable distributions, of which the normal is a special case, represents a natural generalization of the Gaussian framework. Numerous empirical studies (e.g. Fama, 1965, Mittnik, Paolella, and Rachev, 2002, McCulloch, 1996, Simkowitz and Beedles, 1980, and So, 1987, among others and references therein) have found the stable distribution to be more appropriate for modelling asset returns, particularly when fitting the skewness. Notwithstanding its remarkable properties, the stable distribution has a major drawback. Estimating the parameters becomes quite challenging as the density function does not have a closed form, which renders estimation by maximum likelihood difficult (though not impossible, see, for instance, Lambert and Lindsey, 1999).

Our approach follows Fernández and Steel (1998), who present a general method for transforming symmetric distributions into asymmetric distributions. Conditions are rather weak and, in practice, only require the symmetric distribution to be continuous and unimodal. The method they propose consists of introducing skewness by using the inverse scaling of the density function on both sides of the mode; as the inverse scaling is a positive scalar, it captures the asymmetry. The density is left (right) skewed for values smaller (larger) than one, and symmetrical if the inverse scaling is exactly equal to one. They focus on applying their method to the Student-t, leading to the skewed-t distribution. The skewed-t distribution, featured in Giot and Laurent

Hansen (1994) proposes a similar distribution, though the derivation is different.
(2003) for evaluating Value at Risk and extended to the multivariate case in Bauwens and Laurent (2002), has a tractable density function, which facilitates maximum likelihood, and gives us parameters that have an intuitive explanation and are closely related to the first four moments.

The inverse scalar mentioned above captures the asymmetry, our object of interest. Contrary to Hansen (1994), who estimates conditional asymmetry and kurtosis by extending the ARMA and GARCH model to the third and fourth order moments, we estimate the model parameters under the null hypothesis of constant skewness (i.e. unconditional and conditional asymmetry are equal). Under this hypothesis, the sequence of the ratios of the conditional probabilities should be serially uncorrelated. Following Engle and Manganelli (2004), these ratios are equivalent to a sequence of indicator functions which take value one if the ith observation is larger than the (conditional) mode and zero otherwise. We test for uncorrelation of this sequence of binary variables both parametrically and nonparametrically. We first consider a Wald-type test based on generalized linear models, though detailed derivations are provided for both 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 Wald tests under the null.

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. Nour and Bontemps (2005) and Nour and Bontemps (2006) propose a GMM approach to test the distributional assumption. Both these 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.

On the other hand, a series of tests for conditional symmetry can already be found in the literature. 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 high conditional moments with unknown functional forms. 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.

A similar idea is found in Harvey and Siddique (2000), who use a non-central-t distribution and condition the third moment on past residuals to the power one and three.

Tse (2002) also uses Pierce’s correction in a residual-based test for conditional heteroskedasticity.
Finally, 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.

It is worth emphasizing that symmetry does not entail uncorrelation. In this framework, symmetry can be seen as a special case of asymmetry. When the estimated inverse scalar mentioned previously equals one, then there is an equal proportion of observations on either side of the mode. However, this unconditional symmetry does not imply uncorrelation as the conditional probability that the observation at time $t$ is larger than the mode could depend on where past observations lied. A similar argument can be made for the conditional mean: the conditional mean of a series of random variables can be different from zero while their unconditional mean can be equal to zero.

In a comprehensive Monte Carlo study, we evaluate the size and power properties of the test 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. Using daily log returns of several stocks, indexes and bonds for a period of 11 years, we detect the presence of dynamics in the conditional asymmetry of the standardized residuals of 6 different series. These findings are confirmed when estimating a dynamic conditional asymmetry ARCH-type model 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 laws. Since we use several regression models, we take a bottom-up approach: we start with a linear setting, followed by a logit regression and finally, a general form of the regression function where the parameters are estimated by maximum likelihood. We conclude the section with runs test. Section 3 presents the Monte Carlo study and Section 4, the empirical application. Finally, we present our overall conclusions. Technical derivations and details are provided in the Appendix.
2 Testing conditional asymmetry

Let \( y_t (t = 1, \ldots, T) \) denote the (daily, say) return of a series (stock, index, exchange rate, etc.) at time \( t \) and assumed to be generated by the following location-scale dynamic model:

\[
y_t = \mu_t(\eta) + \varepsilon_t \tag{1}
\]

\[
\varepsilon_t = \sigma_t(\eta)z_t \tag{2}
\]

\[
\mu_t(\eta) = \mu(\eta|\Omega_{t-1}) \tag{3}
\]

\[
\sigma_t(\eta) = \sigma(\eta|\Omega_{t-1}), \tag{4}
\]

where \( \Omega_{t-1} \) is the information set at time \( t - 1 \) and \( \eta \) is a vector of unknown parameters. The functional forms \( \mu_t(\eta) \) and \( \sigma_t(\eta) \) are specified according to an ARMA-GARCH type of model.

To estimate the model by maximum likelihood, a distributional assumption on the innovation term \( z_t \) is required. Let \( h(z_t; \delta) \) be the pdf of \( z_t \) that has zero mean and unit variance and \( \delta \) are parameters present in higher order moments. In most financial applications, \( z_t \) is assumed to be i.i.d. \( N(0,1) \) or i.i.d. \( ST(0,1,\nu) \), \( \nu \) being the degrees of freedom of the Student-t distribution. These distributions are symmetric and cannot capture the skewness often found in the residuals.

Fernández and Steel (1998) propose a flexible family of skewed densities. Their procedure allows for skewness to be introduced in any continuous unimodal and symmetric (about 0) distribution \( h(z_t; \delta) \) by changing the scale on each side of the mode through a single parameter \( \xi \). The resulting density is given by

\[
f(z_t; \delta, \xi) = \begin{cases} 
\frac{2}{\xi + \frac{1}{\xi}} h\left(\frac{\xi z_t}{\xi + \frac{1}{\xi}}; \delta\right) & \text{if } z < 0 \\
\frac{2}{\xi + \frac{1}{\xi}} h\left(\frac{\frac{1}{\xi} z_t}{\xi + \frac{1}{\xi}}; \delta\right) & \text{if } z \geq 0,
\end{cases} \tag{5}
\]

where \( z_t \) has mean and variance that are not zero and one anymore but \( m \) and \( s^2 \), depending on \( \xi \), but has mode at zero. In their examples, Fernández and Steel (1998) consider \( h(z_t; \delta) \) to be a Student-t density.\(^5\) Then \( z_t \) is said to follow a skewed-t distribution. The mean and variances are a function of the asymmetry parameter:

\[
m = \frac{\Gamma\left(\frac{\nu-1}{2}\right) \sqrt{\nu - 2}}{\sqrt{\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(\xi - \frac{1}{\xi}\right), \tag{6}
\]

and

\[
s^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2, \tag{7}
\]

and hence \( \mu_t(\eta) \) and \( \sigma_t(\eta) \) are not the mean and variance but the conditional mode and conditional dispersion of \( y_t \).

Giot and Laurent (2003) propose to use a standardized version of the skewed-t distribution. The innovation process \( z_t \) follows a standardized skewed-t distributed, or \( z_t \sim SKST(0, 1, \xi, \nu) \).

\(^5\)Other choices for \( h(z_t; \delta) \) are possible. For instance, Gaussian, Generalized Error Distribution, or a symmetric stable distribution (see Lambert and Laurent, 2001).
if:
\[
f(z^*_t; \xi, \nu) = \begin{cases} \frac{2}{\xi + \frac{1}{\xi}} \sh(\xi z^*_t; \nu) & \text{if } z_t < -\frac{m}{s} \\ \frac{2}{\xi + \frac{1}{\xi}} \sh\left(\frac{z^*_t}{\xi}; \nu\right) & \text{if } z_t \geq -\frac{m}{s} \end{cases}
\]

where \( z^*_t = sz_t + m \) and \( h(\cdot; \nu) \) has a zero mean and unit variance Student-t density. Note that by standardizing, \( z^*_t \) has mean \( m \) but zero mode while \( z_t \) has mean 0 and variance 1 but mode \( -m/s \).

The scale parameter that causes the asymmetry is such that
\[
P(z^*_t \geq 0) = P(z^*_t < 0) = \xi^2.
\]

Equation (9) clearly identifies \( \xi \) as an asymmetry parameter. Following Lambert and Laurent (2001a) and Giot and Laurent (2003), we work with \( \ln(\xi) \) instead of \( \xi \) as a positive (negative) sign for it corresponds to a positive (negative) or right (left) skewed distribution, while a zero value indicates symmetry. The vector of unknown parameters, \( \theta = (\eta, \ln(\xi), \delta) \), is estimated by maximum likelihood and the asymptotic variance-covariance matrix of the MLE is the standard inverse of the Fisher information matrix \( I^{-1}_\theta \).

The asymmetry parameter \( \xi \) is assumed to be constant over time. If this is true, the allocation of mass on either side of the mode of the residuals \( z^*_t(\hat{\theta}) \) should be constant. Equation (9) evaluated in \( z^*_t(\hat{\theta}) \) can be written as
\[
P\left( z^*_t(\hat{\theta}) < 0 \right) = \frac{1}{1 + \xi^2} \overset{\text{def}}{=} g(\hat{\xi}),
\]

and, following Engle and Manganelli (2004), for each observation this probability ratio is equivalently represented by the sequence of indicator functions \( I_t(\hat{\theta}) \overset{\text{def}}{=} I\left( z^*_t(\hat{\theta}) < 0 \right) \) that takes value 1 if the argument is true.

Under the hypothesis of constant asymmetry, the sequence of indicator functions should be uncorrelated with any variables belonging to \( \Omega_{t-1}. \) To test this hypothesis, one might regress \( I_t(\hat{\theta}) \) on its past (i.e. \( I_{t-1}(\hat{\theta}), I_{t-2}(\hat{\theta}), \ldots \)), on other past model-based variables potentially related to asymmetry (like past residuals to the power 1, 2 or 3), on past volatilities, or on a set of exogenous variables (e.g. day of the week dummies). We denote by \( x_t(\hat{\theta}) \) the \( k \times 1 \) vector containing these quantities.

### 2.1 Wald test

The test presented here is parametric and based on a regression of \( I_t(\hat{\theta}) \) on \( x_t(\hat{\theta}) \). On this basis, we rely on Generalized Linear Models (GLM hereafter). Let
\[
E\left[ I_t(\hat{\theta}) | x_t(\hat{\theta}) ; \beta \right] = \iota(\eta_t), \tag{10}
\]
be the conditional mean, \( \eta_t = x_t(\tilde{\theta})' \beta \) be the linear predictor produced by \( x_t(\tilde{\theta}) \), and \( l(\cdot) \) be the link function. Many functional forms for \( l(\cdot) \) are possible but we only present detailed results for the linear and logit links: \( l(\eta_t) = \eta_t \) and \( \ln(l(\eta_t)/1 - l(\eta_t)) = \eta_t \) respectively. As a generalization, we also present the case of the unspecified, but known, form (10) with parameters estimated by quasi-maximum likelihood.

To compute the correct asymptotic distribution of the test, differentiability of (10) with respect to \( \hat{\theta} \) is required. Unfortunately, \( I_t(\hat{\theta}) \) and its lags (involved in \( x_t(\hat{\theta}) \)) are not differentiable as they are binary indicators. Hence, we follow Engle and Manganelli (2004), and approximate \( I_{t-j}(\hat{\theta}) \) (\( j = 0, 1, \ldots \)) by a logistic function:

\[
\hat{I}_{t-j}(\theta) = \left[ 1 + \exp\left(-\frac{z_t^*(\theta)}{\hat{c}}\right)\right]^{-1},
\]

where \( \hat{c} \) is a scalar satisfying certain conditions ensuring that \( \hat{I}_t(\hat{\theta}) \rightarrow_p I_t(\theta) \) (see Appendix). The derivative of \( \hat{I}_{t-j}(\theta) \) with respect to \( \hat{\theta} \) is \( h_{\hat{c}}(z_t^*(\hat{\theta})) \frac{\partial z_t^*(\hat{\theta})}{\partial \theta} \), where \( h_{\hat{c}}(z_t^*(\hat{\theta})) \) is a logistic density function with zero mean and scale \( \hat{c} \). Following Engle and Manganelli (2004), one can show that \( E[h_{\hat{c}}(z_t^*(\hat{\theta}))|\Omega_{t-1}] = f^*(0; \hat{\theta}) \), where \( f^*(0; \hat{\theta}) \) is the pdf of \( z_t^* \) evaluated at zero.

2.1.1 Linear probability model - \( \theta \) known

For ease of exposition, let us first consider the case where \( \theta \) is known. Assuming a linear probability model

\[
P(I_t(\theta) = 1|x_t(\theta)) = g(\xi) + x_t(\theta)' \beta,
\]

and, since the indicator function is a Bernoulli random variable, \( E(I_t(\theta)|x_t(\theta)) = g(\xi) + x_t(\theta)' \beta. \) Alternatively, \( E(Hit_t(\theta)|x_t(\theta)) = x_t(\theta)' \beta \) where \( Hit_t(\theta) = I_t(\theta) - g(\xi). \) The least squares estimator of \( \beta \) is

\[
\sqrt{T} \hat{\beta} = \left( \frac{1}{T} \sum_{t=1}^T x_t(\theta) x'_t(\theta) \right)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T x_t(\theta) Hit_t(\theta) \right).
\]

(12)

The first term in the right hand side of (12) converges to a nonstochastic \( k \times k \) matrix \( J^{-1} \). The second converges in distribution to a Gaussian and, by standard Gauss-Markov assumptions, its asymptotic variance equals \( \sigma^2 J \). Hence the Wald test with null hypothesis \( H_0 : \beta = 0 \) is

\[
W^\text{OLS}_T = \frac{T}{\sigma^2} \hat{\beta}' J \hat{\beta} \sim \chi^2_k.
\]

2.1.2 Linear probability model - \( \theta \) unknown

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

\[
P(I_t(\hat{\theta})|x_t(\hat{\theta})) = g(\hat{\xi}) + x_t(\hat{\theta})' \beta,
\]

(13)
and the estimated $\beta$ is a function of $\hat{\theta}$, i.e.

$$\sqrt{T}\hat{\beta}(\hat{\theta}) = \left(\frac{1}{T}\sum_{t=1}^{T} x_t(\hat{\theta})x_t'(\hat{\theta})\right)^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t(\hat{\theta})H_{\theta}t_{\theta}(\hat{\theta})\right).$$

(14)

This entails an extra source of uncertainty. The following theorem provides the Wald test with the correct asymptotic covariance, denoted $RBD - W^{{OLS}}_T$, where $RBD$ stands for Residual Based Diagnostic.

**Theorem 1** Under the conditional mean (13), the assumptions in the Appendix and the null hypothesis $H_0 : \beta = 0$,

$$RBD - W^{{OLS}}_T = T\hat{\beta}'J(D^2J - \hat{\Omega}_0^{-1}\hat{\Omega}'\hat{\beta})^{-1}\hat{\beta} \sim \chi^2_k,$$

where $\hat{\Omega} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t(\theta) \left(f^*(0; \theta)\frac{\partial z(\theta)}{\partial \theta} - \frac{\partial g(\xi)}{\partial \theta}\right)$ and $\frac{\partial g(\xi)}{\partial \theta} = 0$ except $\frac{\partial g(\xi)}{\partial \ln(\xi)} = -\frac{2\xi^2}{(1+\xi^2)^2}$.

**Proof** See Appendix.

To apply the test, some of the quantities given in Theorem 1 have to be replaced by consistent estimators. A consistent estimator for $J$ is given by the term inside the inverse on the right hand side of (14). Under the null, $\sigma^2$ can be consistently estimated by $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} H_{\theta}t_{\theta}(\hat{\theta})^2$. The $k \times s$ matrix $\hat{\Omega}$ is evaluated at $\hat{\theta}$ and the $1 \times s$ vector of derivatives $\frac{\partial z(\theta)}{\partial \theta}$ evaluated at $\hat{\theta}$ can be computed using numerical methods. Finally, the $s \times s$ variance-covariance matrix $\hat{\Omega}_0^{-1}$ can be estimated by the inverse of the outer product of the scores evaluated at $\hat{\theta}$.

### 2.1.3 Logit model - $\theta$ unknown

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 (13) is a logit link function

$$\ln \left[ \frac{P(I_t(\theta)|x_t(\theta))}{1 - P(I_t(\theta)|x_t(\theta))} \right] = -q(\xi) + x_t(\theta)'\beta,$$

equivalently expressed as

$$P(I_t(\theta)|x_t(\theta)) = \frac{1}{1 + \exp \left( q(\xi) - x_t(\theta)'\beta \right)}.$$

(15)

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

$$r(I_t(\theta)|x_t(\theta)) = P(I_t(\theta)|x_t(\theta))I_t(\theta)(1 - P(I_t(\theta)|x_t(\theta)))^{1-I_t(\theta)},$$

(16)
and estimation can be performed by quasi maximum-likelihood. Let $S(\beta, \hat{\theta})$, $\mathcal{H}(\beta, \hat{\theta})$ and $\mathcal{I}_\beta$ be the score, Hessian and the Fisher information matrix respectively that correspond to the conditional likelihood obtained from the data contributions in (16). And let $\tilde{S}(\beta, \hat{\theta})$ be the approximated score which is a function of $\tilde{I}_{t-j}(\hat{\theta}) \ (j = 0, 1, \ldots)$ instead of $I_{t-j}(\hat{\theta})$. Given the conditions in $\hat{c}$ (see Appendix) and the differentiability of $S(\beta, \hat{\theta})$ with respect to $I_{t-j}(\hat{\theta})$, $\tilde{S}(\beta, \hat{\theta}) \to p S(\beta, \hat{\theta})$.

The following Theorem provides the Wald test with the correct asymptotic covariance.

**Theorem 2** Under model (15), the assumptions in the Appendix and the null hypothesis $H_0 : \beta = 0$,

$$RBD - W_{T}^{Logit} = T \hat{\beta}' J \left( \mathcal{I}_\beta - \tilde{\beta} \mathcal{I}_\beta^{-1} \tilde{\beta}' \right)^{-1} J \hat{\beta} \sim \chi^2_k,$$

where

$$J = T^{-1} \mathcal{H}(\beta, \hat{\theta})$$

$$\tilde{\beta} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t(\theta) \left( f^*(0; \theta) \frac{\partial z_t^*(\theta)}{\partial \theta'} - \frac{\partial g(\xi)}{\partial \theta'} \right).$$

**Proof** See Appendix.

$\tilde{\beta}$ is computed in a similar way as in the linear case and consistent estimators for $\mathcal{I}_\beta$ and $J$ are readily available in the form of the inverse of the outer product of the scores of the log likelihood of the density (16) and the sample average Hessian evaluated at $\hat{\beta}$ and $\hat{\theta}$ respectively.

2.1.4 General case - $\theta$ unknown

The matrix $\mathcal{B}$ 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, $E(I_t(\hat{\theta})|x_t(\hat{\theta})) = P(I_t(\hat{\theta})|x_t(\hat{\theta})) = g(\hat{\xi})$.

This implies that, for a given link function $h(\cdot)$, we can always choose the intercept appropriately such that $\mathcal{B}$ is equal to those of the former Theorems. In fact, the linear probability model and the logistic regressions are not the only models which 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 $0 < F(g(\hat{\xi}) + x_t(\hat{\theta})'\beta) < 1$, the appropriate intercept $q(\hat{\xi})$ is $F^{-1}(P(I_t(\hat{\theta}))) = F^{-1}(g(\hat{\xi}))$ such that $E(I_t(\hat{\theta})|x_t(\hat{\theta})) = F(F^{-1}(g(\hat{\xi}))) = g(\hat{\xi})$.

We now consider the more general case where the link function is known but left unspecified and estimation is done by quasi-maximum likelihood by means of a conditional probability function $r(I_t(\hat{\theta})|x_t(\hat{\theta}))$ that belongs to the exponential family. The following theorem provides the Wald test with the correct asymptotic covariance.

**Theorem 3** Under the conditional mean (10), a density function $r(I_t(\hat{\theta})|x_t(\hat{\theta}))$ that belongs to the exponential family, the assumptions in the Appendix and the null hypothesis $H_0 : \beta = 0$,

$$RBD - W_{T}^{M} = T \hat{\beta}' J \left( \mathcal{I}_\beta - \tilde{\beta} \mathcal{I}_\beta^{-1} \tilde{\beta}' \right)^{-1} J \hat{\beta} \sim \chi^2_k,$$
where
\[
J = T^{-1} \mathcal{H}(\beta, \hat{\theta})
\]
\[
\hat{B} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \hat{S}_t(\beta, \theta)}{\partial \theta'}.
\]

where \(\frac{\partial \hat{S}_t(\beta, \theta)}{\partial \theta'}\) is the contribution to the score of the \(t\)th observation.

Proof See Appendix.

Consistent estimators for \(\hat{B}, J, \mathcal{I}_\beta^{-1}\) and \(\mathcal{I}_\theta^{-1}\) are computed similarly to the previous theorems.

2.2 Runs test

The runs test (also called Wald-Wolfowitz test) is a nonparametric test that checks for nonrandomness in a two-valued data sequence. A run is a sequence of adjacent equal terms. For example, 1111001110011110000 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, \(R_T(\theta)\), is obtained as follows:

\[
R_T(\theta) = 1 + \sum_{t=2}^{T} (I_t(\theta) - I_{t-1}(\theta))^2.
\]

If 1’s and 0’s alternate randomly and \(\theta\) is known, the number of runs is a random variable whose asymptotic distribution is \(\mathcal{N}(a(\theta), b^2(\theta))\) where \(a(\theta) = 1 + 2T_1(\theta)T_2(\theta)T^{-1}, b^2(\theta) = (a(\theta) - 1)(a(\theta) - 2)/(T - 1), T_2(\theta) = \sum_{t=1}^{T} I_t(\theta)\) and \(T_1(\theta) = T - T_2(\theta)\). Premaratne and Tay (2002) use the runs test in an equivalent setting to ours for testing conditional asymmetry. However, our approach differs from Premaratne and Tay (2002) in several respects. First, they do not take into account the uncertainty from substituting \(\theta\) by \(\hat{\theta}\). 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.

The following Theorem provides the runs test with the correct asymptotic covariance.

**Theorem 4** Under the assumptions in the Appendix and under the null hypothesis of no conditional asymmetry in \(\hat{z}_t^*\),

\[
RBD - W_T^{\text{Runs}} = \left( \frac{R_T(\hat{\theta}) - a(\hat{\theta})}{b^2(\hat{\theta}) - \hat{B} \mathcal{I}_\theta^{-1} \hat{B}'} \right)^2 \sim \chi^2_1.
\]

where
\[
\hat{B} = \lim_{T \to \infty} \frac{2}{T} \sum_{t=1}^{T} \frac{\partial z^*_t(\theta)}{\partial \theta'} f^*(0; \theta)(2g(\hat{\xi}) - 1).
\]

\(^6\text{An alternative runs test for asymmetry is McWilliams (1990).}\)
Proof See Appendix.

Consistent estimators for $\tilde{B}$ and $\mathcal{I}_\theta^{-1}$ are computed in a similar way as in the previous Theorems. Notice however that $\tilde{B}$ is now a row vector of dimension $1 \times s$.

3 Monte Carlo study

Different data generating processes (DGP) were considered in order to study the size and power of the tests. All models belong to the general specification given in (1) but differ in the way the conditional mean and the conditional variance are specified. We consider i) an AR(0) and an AR(1) without intercept, i.e. $\mu_t = 0$ and $\mu_t = 0.1y_{t-1}$, ii) a homoscedastic and a GARCH(1,1) model for the conditional variance, i.e. $\sigma_t^2 = 0.4$ or $\sigma_t^2 = 0.4+0.1\epsilon_{t-1}+0.8\sigma_{t-1}^2$, and iii) standardized skewed-t innovations with 7 degrees of freedom and right skewness, i.e. $z_t \sim \text{SKST}(0, 1, \exp(0.08), 7)$.

Given a DGP, $S(= 1000)$ samples of $T (=1000, 2000$ or $3000)$ observations are generated and a model specification is estimated for each sample. The correct GARCH model is always estimated, i.e. a GARCH(0,0) in the homoscedastic case and a GARCH(1,1) otherwise. We allow however for some source of misspecification (overparameterisation) for the conditional mean in the sense that, in addition to the correct specifications, we also consider the case where an AR(1) model is estimated while the DGP has no dynamics in the conditional mean. We apply the three tests presented earlier and consider both the corrected (accounting for the uncertainty of the first step) and uncorrected versions of the Wald tests. Note that the corrected versions of the tests are labelled ‘RBD’ in the tables. For example, $RBD - W^{QLS}_T$ indicates that the correction was made in the test based on the linear probability model while $W^{QLS}_T$ corresponds to the Wald test that ignores the uncertainty in $\hat{\theta}$.

For the parametric tests we consider two sets of covariates. The first set consists of the first two lags of the residuals, i.e. $x_t(\hat{\theta}) = (\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2})'$, and the second consists of the first lags of the residuals and the indicator variable, i.e. $x_t(\hat{\theta}) = (\hat{\epsilon}_{t-1}, I(\hat{z}_{t-1}^* < 0))'$. If the test is properly sized, the alternative hypothesis should be wrongly selected with an empirical frequency closed to $\alpha \times 100\%$ (where $\alpha$ is the nominal size of the test) when $S$ is sufficiently large. When $S = 1000$ and when the true size of the test is exactly 5% or 10%, one expects 99% of the observed sizes to lie within $(3.3\%, 6.8\%)$ or $(7.6\%, 12.5\%)$ respectively. This information was used to point out the undersized (in bold) and the slightly oversized (in italic) tests in Tables 1 and 2 when $\alpha$ is equal to 0.05 and 0.10 respectively.

When an AR(1) is assumed for the conditional mean, the uncorrected tests are systematically undersized, whatever the alternative hypothesis. When the conditional mean is (correctly) assumed constant, the runs test always has the appropriate size, while $W^{QLS}_T$ and $W^{Logit}_T$ are undersized when the parametric tests consider $x_t(\hat{\theta}) = (\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2})'$. Whatever the setting, the
tests combined with the RBD correction always have the appropriate sizes (at the exception of 3 reported slightly oversized tests for the configurations which correspond to the values in italic in Table 1). The dynamics in the variance does not seem to affect the reported test sizes. The results of a complementary simulation study which introduces dynamics in the degrees of freedom, also revealed that this does not affect the size of the tests.\footnote{Not reported to save space but available under request.}

The (empirical) power of the tests is also investigated for some of the previous configurations where the conditional asymmetry parameter $\ln(\xi)$ is replaced by $\ln(\xi_t) = 0.02 + 0.2z_{t-1}$. Note that in this case, $m$ and $s$ in (8) must be replaced by $m_t$ and $s_t$ as they depend on $\xi_t$. The results reported in Tables 3 and 4 indicate that the empirical test sizes are compatible with the targetted nominal size. The powers of the $RBD - W_{T}^{OLS}$ and of the $RBD - W_{T}^{Logit}$ statistics are very similar. When no RBD correction is required to reach the nominal size (i.e. when the conditional mean is appropriately assumed constant), the OLS and the Logit tests have powers similar to those of the corrected tests. The powers are smaller when the alternative hypothesis inappropriately suggests the dependence of the conditional asymmetry. In this case, the power of the $RBD - W_{T}^{Runs}$ statistics is less severely affected, especially when $T$ grows, than the powers of the $RBD - W_{T}^{OLS}$ and $RBD - W_{T}^{Logit}$ statistics. For an appropriate alternative hypothesis, the parametric tests are not surprisingly more powerful than the runs tests. Finally, as to size, the dynamics in the variance do not seem to affect the reported test powers.

4 Empirical Application

In this section we apply our tests to real data. We consider the daily returns of several stocks, indices and bonds, namely the NIKKEI 225 index (NIKKEI), the FTSE 100 index (FTSE), AT&T INC. (AT), the HANG SENG index (HSI), Du Pont De Nemours (DUPONT), the Swiss Market Index (SMI), the Trade Weighted Exchange Index of major currencies (TWEI) and the 30-Years US TREASURY Bond (TBIL30). The data have been obtained from Yahoo Finance, except for TWEI (downloaded from the Federal Reserve Economic Data website) and cover the period 1995-2006, giving us a sample of around 3,000 observations.

The model we consider is a SKST-AR($m$)-APARCH(1,1):

$$y_t = \mu + \sum_{i=1}^{m} \rho_m(y_{t-m} - \mu) + \varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t$$

$$z_t \sim SKST(0, 1, \ln(\xi), \nu)$$

$$\sigma_t^2 = \omega + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\zeta + \beta_1 \sigma_{t-1}^2,$$

where $\theta = (\mu, \rho_1, \ldots, \rho_m, \omega, \alpha_1, \beta_1, \gamma, \zeta, \ln(\xi), \nu)$ are parameters to be estimated.
The AR orders have been selected in such a way that the innovation term does not present any sign of serial correlation. The first part of Table 5 contains the MLE estimates of the APARCH model with constant asymmetry parameter \( \ln(\xi) \) (standard errors in parenthesis). 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 correctly model both the conditional mean and the conditional variance. We first compute the Box-Pierce statistics with 10 lags on the standardized residuals \( (z_t) \) to test for the presence of remaining serial correlation. The statistic \( Q(10) \) is approximately distributed as a \( \chi^2(10 - m) \) under the null of no serial correlation. Second, we compute the same statistics on the squared standardized residuals, denoted \( Q^2(10) \), to test for the presence of heteroscedasticity in the innovation term. 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 \( \chi^2 \) distribution, many authors, nonetheless, apply the \( \chi^2 \) distribution as an approximation (the problem lies in the fact that estimated residuals are used to calculate the portmanteau statistics). Finally, we apply the Residual-Based Diagnostic for conditional heteroscedasticity of Tse (2002). This test involves estimating the following model by OLS: 
\[
\hat{z}_t^2 - 1 = \delta_1 \hat{z}_{t-1}^2 + \ldots + \delta_M \hat{z}_{t-M}^2 + u_t \quad (\text{we set } M = 10).
\]
Since the regressors are not observed (but indeed estimated), the standard inference procedure of OLS 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 second part of Table 5. As a whole, there is no evidence of misspecification of the first two conditional moments.

Out of 8 series, 5 (all except NIKKEI, AT and HSI) appear to be skewed in the sense that \( \ln(\xi) \) is significantly different from 0. We nest apply the three tests for conditional asymmetry presented in Section 2. Unlike the runs test, the parametric tests require the choice of a set of variables which we suspect possess information for forecasting the conditional asymmetry. In the application, we consider various candidates: \( I_{t-1}, \varepsilon_{t-1}, \varepsilon_{t-1}^2, \varepsilon_{t-1}^3, \sigma_{t-1}^2 \). We also report the outcome of the test which includes more lags of some of these variables. Since this test is based on an auxiliary regression, we also include more than one variable at a time in the test. Finally, because of the similarity between the results regarding \( RBD - W^{OLS}_T \) and \( RBD - W^{Logit}_T \), and also in order to save space, we choose only to report the p-values associated with \( RBD - W^{OLS}_T \) statistics.

Two series do not present any evidence of predictability of their asymmetry, i.e. SMI and TBIL30. The non-parametric runs test is only significant at the 5% level for 3 of the remaining 6 series (i.e. NIKKEI, HSI and TWEI). Using some relevant predictors, the \( RBD - W^{OLS}_T \) statistics detects the presence of dynamics in 3 additional series (i.e. FTSE, AT and DUPONT). The most
interesting case is probably the HANG SENG index. The estimated degree of asymmetry is not significantly different from zero, which indicates that the index has an unconditional symmetric distribution. However, the tests detect dependencies, which implies conditional asymmetry. 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 is indeed unconditionally symmetric.

The usefulness of the parametric tests is thus twofold: it has more power than the runs test in certain situations (see Section 3) and it gives us an idea of the source of the dynamics in the 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 test suggest the following for \( \ln(\xi_t) \):

\[
\begin{align*}
\ln(\xi_t) &= \ln(\xi) + \tau_1 \epsilon_{t-1} & \text{for NIKKEI} \\
\ln(\xi_t) &= \ln(\xi) + \tau_2 \epsilon_{t-2} & \text{for FTSE} \\
\ln(\xi_t) &= \ln(\xi) + \tau_3 \epsilon_{t-1} + \tau_4 \sigma_{t-1}^2 & \text{for AT, HSI and TWEI} \\
\ln(\xi_t) &= \ln(\xi) + \tau_3 \epsilon_{t-1} + \tau_5 \epsilon_{t-1}^2 + \tau_4 \sigma_{t-1}^2 & \text{for DUPONT.}
\end{align*}
\]

For the remaining two series (SMI and TBIL30), we propose to estimate Hansen’s (1994) specification, i.e.

\[
\ln(\xi_t) = \ln(\xi) + \tau_1 \epsilon_{t-1} + \tau_5 \epsilon_{t-1}^2.
\]

Since there was no evidence of dynamics for these series, we expect \( \tau_1 \) and \( \tau_5 \) not to be significantly different from 0.

The Dynamic Conditional Asymmetry SKST-AR-APARCH model estimation results are reported in Table 6. The first part of the table pertains to the MLE while the second part contains the loglikelihood and a likelihood ratio test (LRT) of the null assumption of no dynamics in the conditional asymmetry.

Results of the test for conditional asymmetry are compatible with the estimated parameters of the model. Dynamics are detected in all but 2 series, SMI and TBIL30. The specification choice suggested by the parametric test is relevant, as estimated parameters are significantly different from zero. For instance, \( \epsilon_{t-1} \) was found to be a good predictor of the conditional asymmetry for the NIKKEI and its coefficient is indeed highly significant. Similarly, \( \epsilon_{t-2} \) appears to be significant while \( \epsilon_{t-1}^2 \) is insignificant. The same comments apply to all the series, except for AT, where \( \epsilon_{t-1}^3 \)

---

8We have estimated various specifications for the conditional asymmetry of the model. We found that the best model (in terms of goodness-of-fit) was compatible with our prior.
is not found to explain the conditional asymmetry.\(^9\)

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 proposed three tests – two parametric and one nonparametric – based on the residuals. The tests are therefore 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 test 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 which 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

Assumptions

A.1 The joint distribution of \( \hat{\theta} \) and the score \( S(\beta, \theta) \) is asymptotically Gaussian:

\[
\left( \begin{array}{c}
\sqrt{T}(\hat{\theta} - \theta_0)
\sqrt{T}S(\beta, \theta)
\end{array} \right) \sim \mathcal{N} \left( \begin{array}{c}
0
0
\end{array} , \begin{pmatrix}
\mathcal{I}_\theta^{-1} & V_{12} \\
V_{12} & \mathcal{I}_\beta
\end{pmatrix} \right),
\]

where \( \mathcal{I}_\theta \) and \( \mathcal{I}_\beta \) are \( s \times s \) and \( k \times k \) Fisher information matrices.

A.2 The score \( S(\beta, \hat{\theta}) \) can be approximated by

\[
\sqrt{T}S(\beta, \hat{\theta}) = \sqrt{T}S(\beta, \theta) + B\sqrt{T}(\hat{\theta} - \theta) + o_p(1),
\]

where \( B = \lim_{T \to \infty} E \left( \frac{\partial S(\beta, \theta)}{\partial \theta} \right) \) is a \( k \times s \) matrix.

\(^9\)Note that concerning HSI, while \( \tau_3 \) (coefficient of \( \varepsilon_{t-1}^{3} \)) is only significant at the 10% level using a t-test, it is found to be significant at the 1% level through a LR T. This difference is due to the correlation between \( \varepsilon_{t-1} \) and \( \varepsilon_{t-1}^{3} \).
A.3 The approximated score $\tilde{S}(\beta, \hat{\theta})$ is differentiable with respect to $\beta, \hat{\theta}$ and $\hat{I}_{t-j} \hat{\theta}$, $j = 0, 1, \ldots$, and can be approximated by

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

where $\tilde{B} = \lim_{T \to \infty} E \left( \frac{\partial \tilde{S}(\theta)}{\partial \theta} \right)$ is a $k \times s$ matrix.

A.4 The joint distribution of $\hat{\theta}$ and the statistic $S(\hat{\theta}) = (R_T(\hat{\theta}) - a(\hat{\theta}))$ is asymptotically Gaussian:

$$\begin{pmatrix} \sqrt{T} (\hat{\theta} - \theta_0) \\ \sqrt{T} S(\theta) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_\theta^{-1} & V_{12} \\ V_{12} & V_{22} \end{pmatrix} \right),$$

where $I_\theta$ is a $s \times s$ Fisher information matrix and $V_{22}$ is strictly positive scalar.

A.5 The statistic $S(\hat{\theta}) = (R_T(\hat{\theta}) - a(\hat{\theta}))$ can be approximated by

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

where $B = \lim_{T \to \infty} E \left( \frac{\partial S(\theta)}{\partial \theta} \right)$ is a $1 \times s$ vector.

A.6 The approximated statistic $\tilde{S}(\hat{\theta})$ is differentiable with respect to $\hat{\theta}$ and $\hat{I}_{t-j} \hat{\theta}$, $j = 0, 1, \ldots$, and can be approximated by

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

where $\tilde{B} = \lim_{T \to \infty} E \left( \frac{\partial \tilde{S}(\theta)}{\partial \theta} \right)$ is a $1 \times s$ vector.

A.7 There exist a nonstochastic sequence $c = o(1)$ and $c^{-1} = o(\sqrt{T})$ such that $\hat{c}/c \to_p 1$ and

$$I_t(\hat{\theta}) \approx \hat{I}_t(\hat{\theta}) = \frac{1}{1 + \exp \left( \frac{z^*_t(\hat{\theta})}{c} \right)}.$$

Likewise, $\hat{H}t_t(\hat{\theta}) = \hat{I}_t(\hat{\theta}) - g(\hat{\xi}).$

A.8 $h_t(z^*_t)$ and $\left( x_t(\hat{\theta}), \frac{\partial x_t(\hat{\theta})}{\partial \hat{\theta}} \right)$, $r > 0$ are uncorrelated.

Proofs

Theorem 1

This proof partially relies on Tse (2002). The first term in the right hand side of (12) converges to a nonstochastic $k \times k$ matrix $J^{-1}$. We need to show that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t(\hat{\theta}) Ht_t(\hat{\theta}),$$

17
As for the second term, following Engle and Manganelli (2004) and using \[A.7\] which implies, by \[A.1\] and \[A.3\], we can replace the expectation operator by the sample mean. Grouping terms:

To compute the matrix \( \tilde{\Sigma} \), \( \tilde{\Pi} \) and by \[A.6\] null, zero:

Taking expectations in one of the summands, the first term of the right hand side is, under the null, zero:

As for the second term, following Engle and Manganelli (2004) and using \[A.7\]

By the law of iterated expectations and \[A.8\]

where \( f^*(0; \hat{\theta}) \) is Equation (8) evaluated at zero. Under some suitable form of the law of large number, we can replace the expectation operator by the sample mean. Grouping terms:

and by \[A.6\]

which implies, by \[A.1\] and \[A.3\],

A direct application of the Wald test yields the result.

**Theorem 2**

The mean log-likelihood is given by

and the score

\[
S(\beta, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left( -(1 - I_t(\hat{\theta})) x_t(\hat{\theta}) \beta - \ln \left( 1 + \exp(q(\hat{\xi}) - x_t(\hat{\theta})' \beta) \right) \right) \]
has, by [A.1], variance-covariance matrix $\mathbf{I}_\beta$. After some simplifications

$$S(\beta, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^{T} x_t(\hat{\theta}) \left( I_t(\hat{\theta}) - \frac{1}{1 + \exp(q(\xi) - x_t(\theta)^T \beta)} \right).$$

By [A.3], the score is approximated, substituting $I_{t-j}(\hat{\theta})$ by $\tilde{I}_{t-j}(\hat{\theta})$, $j = 0, \ldots$ whenever derivatives with respect to $\theta$ are taken. The matrix $\tilde{\mathbf{B}}$ is given in [A.3] and under the null,$^{10}$

$$\frac{\partial S(\beta, \hat{\theta})}{\partial \theta^t} = \frac{1}{T} \sum_{t=1}^{T} \left( x_t(\hat{\theta}) \frac{\partial I_t(\hat{\theta})}{\partial \theta^t} + I_t(\hat{\theta}) \frac{\partial x_t(\hat{\theta})}{\partial \theta^t} - g(\xi) \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta^t} - x_t(\hat{\theta}) \frac{\partial g(\xi)}{\partial \theta^t} \right),$$

where the second equality come from $H \tilde{I}_t(\hat{\theta}) = I_t(\hat{\theta}) - g(\xi)$ and $\frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta^t} = \frac{\partial I_t(\hat{\theta})}{\partial \theta^t} - \frac{\partial g(\xi)}{\partial \theta^t}$. Since this derivative is equivalent to (17),

$$\sqrt{T} \tilde{\beta}(\theta) \sim \mathcal{N} \left( \beta(\theta), J^{-1} \left( \sigma^2 J - \tilde{\mathbf{B}} \tilde{\mathbf{B}}^T \right) J^{-1} \right).$$

A direct application of the Wald test yields the result.

**Theorem 3**

It follows immediately from [A.1] and [A.2] and making use of [A.3].

**Theorem 4**

Let $S(\hat{\theta}) = (R_T(\hat{\theta}) - a(\theta))$ so that $\sqrt{T} S(\hat{\theta}) \sim \mathcal{N}(0, b^2(\theta)/T)$. Developing $S(\hat{\theta})$:

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

As $I_t^2(\hat{\theta}) = I_t(\hat{\theta})$, the above expression simplifies to

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

The matrix $\tilde{\mathbf{B}}$ is given in [A.6] and

$$\frac{\partial S(\hat{\theta})}{\partial \theta^t} = \sum_{t=2}^{T} \left( \frac{\partial I_t(\hat{\theta})}{\partial \theta^t} + \frac{\partial \tilde{I}_{t-1}(\hat{\theta})}{\partial \theta^t} - 2 \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta^t} I_{t-1}(\hat{\theta}) - 2 I_t(\hat{\theta}) \frac{\partial \tilde{I}_{t-1}(\hat{\theta})}{\partial \theta^t} \right)$$

$$- 2 \sum_{t=1}^{T} \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta^t} + 4 g(\xi) \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta^t}.$$  

$^{10}$Note that, under the null, $\frac{\partial}{\partial \theta^t} \exp(q(\xi) - x_t(\theta)^T \beta) = g(\xi)$.
To compute $\hat{B}$ we take expectations under the null and using [A.4] and [A.9]:

$$
\sum_{t=2}^{T} E \left( \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta'} \right) = (T-1)f^*(0; \hat{\theta})E \left( \frac{\partial z_1^*(\hat{\theta})}{\partial \theta'} \right)
$$

$$
\sum_{t=2}^{T} E \left( \frac{\partial \tilde{I}_{t-1}(\hat{\theta})}{\partial \theta'} \right) = (T-1)f^*(0; \hat{\theta})E \left( \frac{\partial z_{t-1}^*(\hat{\theta})}{\partial \theta'} \right)
$$

$$
-2 \sum_{t=2}^{T} E \left( \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta'} I_{t-1}(\hat{\theta}) \right) = -2(T-1)g(\xi)f^*(0; \hat{\theta})E \left( \frac{\partial z_1^*(\hat{\theta})}{\partial \theta'} \right)
$$

$$
-2 \sum_{t=2}^{T} E \left( \frac{\partial \tilde{I}_{t-1}(\hat{\theta})}{\partial \theta'} I_t(\hat{\theta}) \right) = -2(T-1)g(\xi)f^*(0; \hat{\theta})E \left( \frac{\partial z_{t-1}^*(\hat{\theta})}{\partial \theta'} \right)
$$

$$
-2 \sum_{t=1}^{T} E \left( \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta'} \right) = -2Tf^*(0; \hat{\theta})E \left( \frac{\partial z_1^*(\hat{\theta})}{\partial \theta'} \right)
$$

$$
4g(\xi) \sum_{t=1}^{T} E \left( \frac{\partial \tilde{I}_t(\hat{\theta})}{\partial \theta'} \right) = 4Tg(\xi)f^*(0; \hat{\theta})E \left( \frac{\partial z_1^*(\hat{\theta})}{\partial \theta'} \right)
$$

so that

$$
\frac{\partial S(\theta)}{\partial \theta'} = 2f^*(0; \hat{\theta})(2g(\xi) - 1)E \left( \frac{\partial z_1^*(\hat{\theta})}{\partial \theta'} \right).
$$

Under some suitable form of the law of large numbers and grouping terms

$$
\hat{B} = \lim_{T \to \infty} \frac{2}{T} \sum_{t=1}^{T} \frac{\partial z_1^*(\theta)}{\partial \theta'} f^*(0; \theta)(2g(\xi) - 1),
$$

which yields the result.

**References**


<table>
<thead>
<tr>
<th>DGP Model</th>
<th>Model</th>
<th>Variance</th>
<th>Mean</th>
<th>Mean</th>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$W_{\text{OLS}}^{T}$</td>
</tr>
<tr>
<td>GARCH(0,0)</td>
<td>AR(0)</td>
<td>AR(0)</td>
<td>1000</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AR(1)</td>
<td>1000</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AR(1)</td>
<td>1000</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>0.0</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>AR(0)</td>
<td>AR(0)</td>
<td>1000</td>
<td>2.9</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AR(1)</td>
<td>1000</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AR(1)</td>
<td>1000</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 1: Empirical frequency (in percentages) of rejection of the null hypothesis of constant skewness when the nominal size is $\alpha = 0.05$. Bold (italic) values correspond to an undersized (oversized) test.
### Test statistics

<table>
<thead>
<tr>
<th>Variance Mean</th>
<th>DGP Model</th>
<th>Mean T</th>
<th>OLS</th>
<th>$W_T^{OLS}$</th>
<th>$W_T^{Logit}$</th>
<th>$W_T^{Logit}$</th>
<th>$W_T^{Runs}$</th>
<th>$W_T^{Runs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}(\mathbf{t}) = (\hat{\epsilon}<em>{t-1}, \hat{\epsilon}</em>{t-2})'$</td>
<td>GARCH(0,0) AR(0) AR(0) 1000</td>
<td>5.8</td>
<td>11.7</td>
<td>5.5</td>
<td>12.2</td>
<td>9.4</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>4.5</td>
<td>9.6</td>
<td>4.5</td>
<td>10.0</td>
<td>9.2</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>5.7</td>
<td>11.2</td>
<td>5.5</td>
<td>11.5</td>
<td>10.6</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>AR(1) 1000</td>
<td>0.5</td>
<td>11.8</td>
<td>0.5</td>
<td>11.6</td>
<td>3.4</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.2</td>
<td>11.3</td>
<td>0.2</td>
<td>11.1</td>
<td>2.9</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>0.1</td>
<td>9.4</td>
<td>0.1</td>
<td>8.6</td>
<td>2.9</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>AR(1) AR(1) 1000</td>
<td>0.5</td>
<td>11.1</td>
<td>0.5</td>
<td>11.2</td>
<td>3.5</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.0</td>
<td>11.0</td>
<td>0.0</td>
<td>10.6</td>
<td>2.9</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>0.1</td>
<td>9.3</td>
<td>0.1</td>
<td>9.0</td>
<td>3.0</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>GARCH(1,1) AR(0) AR(0) 1000</td>
<td>5.7</td>
<td>10.5</td>
<td>5.7</td>
<td>10.7</td>
<td>9.4</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>4.8</td>
<td>8.9</td>
<td>4.6</td>
<td>9.0</td>
<td>9.1</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>5.6</td>
<td>11.1</td>
<td>5.6</td>
<td>11.7</td>
<td>10.4</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>AR(1) 1000</td>
<td>0.9</td>
<td>11.2</td>
<td>0.8</td>
<td>10.7</td>
<td>2.4</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.4</td>
<td>10.8</td>
<td>0.4</td>
<td>10.5</td>
<td>2.7</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>0.1</td>
<td>9.8</td>
<td>0.1</td>
<td>9.5</td>
<td>2.6</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>AR(1) AR(1) 1000</td>
<td>0.6</td>
<td>11.5</td>
<td>0.5</td>
<td>11.6</td>
<td>2.5</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.3</td>
<td>10.1</td>
<td>0.3</td>
<td>10.0</td>
<td>2.7</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>0.2</td>
<td>10.1</td>
<td>0.2</td>
<td>9.8</td>
<td>2.6</td>
<td>9.6</td>
</tr>
<tr>
<td>$\hat{\theta}(\mathbf{t}) = (\hat{\epsilon}<em>{t-1}, I(\hat{z}</em>{t-1}^*) &lt; 0))'$</td>
<td>GARCH(0,0) AR(0) AR(0) 1000</td>
<td>10.3</td>
<td>10.4</td>
<td>10.3</td>
<td>11.0</td>
<td>9.4</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>10.2</td>
<td>10.2</td>
<td>10.1</td>
<td>10.9</td>
<td>10.6</td>
<td>10.7</td>
</tr>
<tr>
<td></td>
<td>AR(1) 1000</td>
<td>3.3</td>
<td>10.1</td>
<td>3.2</td>
<td>9.7</td>
<td>3.4</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>2.9</td>
<td>9.2</td>
<td>2.9</td>
<td>9.2</td>
<td>2.9</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>3.0</td>
<td>10.5</td>
<td>3.0</td>
<td>9.9</td>
<td>2.9</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>AR(1) AR(1) 1000</td>
<td>3.0</td>
<td>10.1</td>
<td>3.0</td>
<td>10.1</td>
<td>3.5</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>2.8</td>
<td>9.2</td>
<td>2.8</td>
<td>8.9</td>
<td>2.9</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>3.0</td>
<td>9.6</td>
<td>2.7</td>
<td>9.6</td>
<td>3.0</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>GARCH(1,1) AR(0) AR(0) 1000</td>
<td>9.8</td>
<td>10.0</td>
<td>9.7</td>
<td>10.4</td>
<td>9.4</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>8.7</td>
<td>8.8</td>
<td>8.5</td>
<td>9.6</td>
<td>9.1</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>9.5</td>
<td>9.5</td>
<td>9.5</td>
<td>10.0</td>
<td>10.4</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>AR(1) 1000</td>
<td>3.0</td>
<td>10.2</td>
<td>2.8</td>
<td>10.0</td>
<td>2.4</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>3.3</td>
<td>8.6</td>
<td>3.3</td>
<td>8.5</td>
<td>2.7</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>3.1</td>
<td>9.4</td>
<td>3.0</td>
<td>9.1</td>
<td>2.6</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>AR(1) AR(1) 1000</td>
<td>3.2</td>
<td>9.4</td>
<td>3.2</td>
<td>9.9</td>
<td>2.5</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>3.1</td>
<td>9.2</td>
<td>3.0</td>
<td>9.0</td>
<td>2.7</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3000</td>
<td>3.2</td>
<td>9.7</td>
<td>3.2</td>
<td>9.6</td>
<td>2.6</td>
<td>9.6</td>
</tr>
</tbody>
</table>

Table 2: Empirical frequency (in percentages) of rejection of the null hypothesis of constant skewness when the nominal size is $\alpha = 0.10$. Bold (italic) values correspond to an undersized (oversized) test.
DGP: $\ln(\xi_t) = 0.08 + 0.2z_{t-1}$

<table>
<thead>
<tr>
<th>Variance</th>
<th>DGP Model</th>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\text{RBD-}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{RBD-}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{OLS}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{OLS}$</td>
</tr>
</tbody>
</table>

\[ x_t(\hat{\theta}) = (\hat{\epsilon}_{t-1}, \hat{\epsilon}_{t-2})' \]

| GARCH(0,0) | AR(0) | AR(1) | 1000 | 14.8 | 14.8 | 18.9 |
| 2000 | 25.0 | 24.8 | 29.2 |
| 3000 | 33.4 | 33.2 | 41.2 |

| GARCH(1,1) | AR(0) | AR(1) | 1000 | 14.7 | 14.4 | 18.1 |
| 2000 | 21.6 | 21.4 | 27.5 |
| 3000 | 28.6 | 28.2 | 38.9 |

\[ x_t(\hat{\theta}) = (\hat{\epsilon}_{t-1}, I(\hat{z}_{t-1} < 0))' \]

| GARCH(0,0) | AR(0) | AR(0) | 1000 | 38.9 | 44.6 | 38.1 | 46.1 | 38.1 | 38.3 |
| 2000 | 71.4 | 76.0 | 71.1 | 76.6 | 64.9 | 64.9 |
| 3000 | 89.4 | 91.2 | 89.1 | 91.6 | 82.3 | 82.4 |

| GARCH(1,1) | AR(0) | AR(1) | 1000 | 38.7 | 36.0 | 18.9 |
| 2000 | 64.5 | 62.2 | 29.2 |
| 3000 | 80.2 | 78.7 | 41.2 |

| GARCH(0,0) | AR(0) | AR(0) | 1000 | 39.7 | 46.7 | 39.2 | 47.5 | 38.0 | 38.4 |
| 2000 | 72.2 | 75.9 | 71.9 | 76.8 | 64.9 | 64.9 |
| 3000 | 89.4 | 90.9 | 89.3 | 91.3 | 82.4 | 82.5 |

| GARCH(1,1) | AR(0) | AR(1) | 1000 | 34.3 | 33.3 | 18.1 |
| 2000 | 58.2 | 56.5 | 27.5 |
| 3000 | 75.7 | 74.4 | 38.9 |

Table 3: Empirical frequency (in percentages) of (correct) rejection of the null hypothesis of constant skewness when the nominal size is $\alpha = 0.05$ and $\ln(\xi_t) = 0.02 + 0.2z_{t-1}$. No percentage is reported when the test is undersized.
\[ \ln(\xi_t) = 0.08 + 0.2z_{t-1} \]

<table>
<thead>
<tr>
<th>Variance</th>
<th>DGP Mean</th>
<th>Model Mean</th>
<th>T</th>
<th>Test statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( W_{T,\text{OLS}}^{\text{RBD}} )</td>
</tr>
<tr>
<td>GARCH(0,0)</td>
<td>AR(0)</td>
<td>AR(1)</td>
<td>1000</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>–</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>AR(0)</td>
<td>AR(1)</td>
<td>1000</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2000</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3000</td>
<td>–</td>
</tr>
</tbody>
</table>

\[ x_t(\hat{\theta}) = (\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2})' \]

\[ x_t(\hat{\theta}) = (\hat{\varepsilon}_{t-1}, I(\hat{z}_{t-1} < 0))' \]

Table 4: Empirical frequency (in percentages) of (correct) rejection of the null hypothesis of constant skewness when the nominal size is \( \alpha = 0.10 \) and \( \ln(\xi_t) = 0.02 + 0.2z_{t-1} \). No percentage is reported when the test is undersized.
Table 5: Estimation results

<table>
<thead>
<tr>
<th></th>
<th>NIKKEI</th>
<th>FTSE</th>
<th>AT</th>
<th>HSI</th>
<th>DUPONT</th>
<th>SMI</th>
<th>TWEI</th>
<th>TBIL30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>-0.001</td>
<td>0.017</td>
<td>0.052</td>
<td>0.050</td>
<td>0.051</td>
<td>0.043</td>
<td>0.003</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.013)</td>
<td>(0.025)</td>
<td>(0.022)</td>
<td>(0.026)</td>
<td>(0.013)</td>
<td>(0.007)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.049</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_2$</td>
<td></td>
<td>0.039</td>
<td>0.018</td>
<td>0.018</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.030</td>
<td>0.011</td>
<td>0.007</td>
<td>0.016</td>
<td>0.009</td>
<td>0.022</td>
<td>0.002</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.001)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.069</td>
<td>0.057</td>
<td>0.055</td>
<td>0.058</td>
<td>0.048</td>
<td>0.087</td>
<td>0.040</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.925</td>
<td>0.943</td>
<td>0.954</td>
<td>0.941</td>
<td>0.958</td>
<td>0.911</td>
<td>0.962</td>
<td>0.958</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.011)</td>
<td>(0.007)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.584</td>
<td>0.879</td>
<td>0.142</td>
<td>0.530</td>
<td>0.244</td>
<td>0.573</td>
<td>0.118</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.136)</td>
<td>(0.090)</td>
<td>(0.107)</td>
<td>(0.106)</td>
<td>(0.087)</td>
<td>(0.112)</td>
<td>(0.118)</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>1.107</td>
<td>1.006</td>
<td>1.191</td>
<td>1.307</td>
<td>1.033</td>
<td>1.024</td>
<td>1.150</td>
<td>1.479</td>
</tr>
<tr>
<td></td>
<td>(0.197)</td>
<td>(0.181)</td>
<td>(0.255)</td>
<td>(0.189)</td>
<td>(0.236)</td>
<td>(0.184)</td>
<td>(0.343)</td>
<td>(0.362)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>9.772</td>
<td>24.177</td>
<td>7.584</td>
<td>7.208</td>
<td>7.689</td>
<td>11.969</td>
<td>8.517</td>
<td>8.743</td>
</tr>
<tr>
<td></td>
<td>(1.565)</td>
<td>(8.624)</td>
<td>(0.999)</td>
<td>(0.959)</td>
<td>(0.966)</td>
<td>(2.118)</td>
<td>(1.276)</td>
<td>(1.241)</td>
</tr>
<tr>
<td>$\ln(\xi)$</td>
<td>-0.021</td>
<td>-0.129</td>
<td>0.036</td>
<td>0.004</td>
<td>0.071</td>
<td>-0.122</td>
<td>-0.053</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.024)</td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.027)</td>
</tr>
</tbody>
</table>

Sample Size 2955 3030 3022 2968 3022 3020 3017 3004
Loglikelihood -5020.790 -3933.878 -5758.071 -4983.409 -5732.859 -4163.528 -1382.580 -3790.851
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
RBD(5) 0.100 0.122 1.000 0.227 0.357 0.999 0.959

$RBD - W_{T}^{RBD}$

$RBD - W_{T}^{OLS}$

Standard errors in parenthesis.

Q(10), Q2(10) and RBD(5) correspond respectively to the p-value of the Box-Pierce statistics with 10 lags on the standardized and squared standardized residuals and the Residual-Based Diagnostic test for conditional heteroscedasticity of TSE (2002) with 5 lags.

$RBD - W_{T}^{RBD}$ is the p-value of the corresponding statistics.
The values reported in the last part of the Table (below $RBD - W_{T}^{OLS}$) are to the p-values of the corresponding test, using the associated covariates.
## Table 6: Estimation results

<table>
<thead>
<tr>
<th></th>
<th>NIKKEI</th>
<th>FTSE</th>
<th>AT</th>
<th>HSI</th>
<th>DUPONT</th>
<th>SMI</th>
<th>TWEI</th>
<th>TBIL30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0026</td>
<td>0.0177</td>
<td>0.0519</td>
<td>0.0539</td>
<td>0.0467</td>
<td>0.0450</td>
<td>0.0022</td>
<td>-0.0315</td>
</tr>
<tr>
<td>(0.0235)</td>
<td>(0.0142)</td>
<td>(0.0256)</td>
<td>(0.0249)</td>
<td>(0.0256)</td>
<td>(0.0157)</td>
<td>(0.0066)</td>
<td>(0.0152)</td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.0843</td>
<td>0.0046</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0213)</td>
<td>(0.0196)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_2$</td>
<td></td>
<td></td>
<td>-0.0353</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0173)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.0302</td>
<td>0.0115</td>
<td>0.0057</td>
<td>0.0170</td>
<td>0.0077</td>
<td>0.0217</td>
<td>0.0018</td>
<td>0.0094</td>
</tr>
<tr>
<td>(0.0077)</td>
<td>(0.0027)</td>
<td>(0.0054)</td>
<td>(0.0053)</td>
<td>(0.0049)</td>
<td>(0.0047)</td>
<td>(0.0013)</td>
<td>(0.0046)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0718</td>
<td>0.0590</td>
<td>0.0535</td>
<td>0.0602</td>
<td>0.0467</td>
<td>0.0867</td>
<td>0.0438</td>
<td>0.0381</td>
</tr>
<tr>
<td>(0.0090)</td>
<td>(0.0069)</td>
<td>(0.0197)</td>
<td>(0.0101)</td>
<td>(0.0120)</td>
<td>(0.0108)</td>
<td>(0.0100)</td>
<td>(0.0071)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.9217</td>
<td>0.9411</td>
<td>0.9558</td>
<td>0.9378</td>
<td>0.9596</td>
<td>0.9107</td>
<td>0.9596</td>
<td>0.9565</td>
</tr>
<tr>
<td>(0.0098)</td>
<td>(0.0073)</td>
<td>(0.0175)</td>
<td>(0.0095)</td>
<td>(0.0113)</td>
<td>(0.0113)</td>
<td>(0.0097)</td>
<td>(0.0118)</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.5626</td>
<td>0.8982</td>
<td>0.1266</td>
<td>0.5548</td>
<td>0.2201</td>
<td>0.5703</td>
<td>0.1167</td>
<td>0.1699</td>
</tr>
<tr>
<td>(0.1277)</td>
<td>(0.1213)</td>
<td>(0.0938)</td>
<td>(0.1031)</td>
<td>(0.1157)</td>
<td>(0.0780)</td>
<td>(0.1251)</td>
<td>(0.1298)</td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>1.1437</td>
<td>1.0267</td>
<td>1.1973</td>
<td>1.3007</td>
<td>1.0404</td>
<td>1.0286</td>
<td>1.2626</td>
<td>1.5211</td>
</tr>
<tr>
<td>(0.1894)</td>
<td>(0.1904)</td>
<td>(0.2314)</td>
<td>(0.1970)</td>
<td>(0.2318)</td>
<td>(0.1734)</td>
<td>(0.3775)</td>
<td>(0.4640)</td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>9.9919</td>
<td>23.7613</td>
<td>7.4802</td>
<td>7.3836</td>
<td>7.6851</td>
<td>11.9550</td>
<td>8.7309</td>
<td>8.8661</td>
</tr>
<tr>
<td>(1.7410)</td>
<td>(9.2969)</td>
<td>(0.9820)</td>
<td>(1.0364)</td>
<td>(1.0675)</td>
<td>(2.9181)</td>
<td>(1.2805)</td>
<td>(1.4780)</td>
<td></td>
</tr>
</tbody>
</table>

### Conditional Asymmetry

<table>
<thead>
<tr>
<th></th>
<th>$\ln(\xi)$</th>
<th>$\tau_1(\varepsilon_{t-1})$</th>
<th>$\tau_2(\varepsilon_{t-2})$</th>
<th>$\tau_3(\varepsilon_{t-1}^3)$</th>
<th>$\tau_4(\varepsilon_{t-1}^2)$</th>
<th>$\tau_5(\varepsilon_{t-1}^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.0236</td>
<td>0.0605</td>
<td>0.0731</td>
<td>-0.0002</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0322)</td>
<td>(0.0269)</td>
<td>(0.0263)</td>
<td>(0.0226)</td>
<td>(0.0003)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>-0.1328</td>
<td>0.0530</td>
<td>0.0731</td>
<td>-0.8470</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0249)</td>
<td>(0.0233)</td>
<td>(0.0263)</td>
<td>(0.0207)</td>
<td>(0.4834)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>0.0381</td>
<td>0.0925</td>
<td>0.0731</td>
<td>-0.1016</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0249)</td>
<td>(0.0476)</td>
<td>(0.0207)</td>
<td>(0.0111)</td>
<td>(0.0278)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>0.0103</td>
<td>0.0282</td>
<td>0.0731</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0233)</td>
<td>(0.0476)</td>
<td>(0.0207)</td>
<td>(0.0111)</td>
<td>(0.0278)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>-0.0273</td>
<td>0.0340</td>
<td>0.0731</td>
<td>-0.1016</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0249)</td>
<td>(0.0246)</td>
<td>(0.0263)</td>
<td>(0.0111)</td>
<td>(0.0278)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>-0.1296</td>
<td>0.3438</td>
<td>0.0731</td>
<td>-0.1016</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0579)</td>
<td>(0.0242)</td>
<td>(0.0263)</td>
<td>(0.0111)</td>
<td>(0.0278)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td></td>
<td>-0.0579</td>
<td>0.4472</td>
<td>0.0731</td>
<td>-0.1016</td>
<td>0.0153</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.0547)</td>
<td>(0.0242)</td>
<td>(0.0263)</td>
<td>(0.0111)</td>
<td>(0.0278)</td>
<td>(0.0140)</td>
<td>(0.0003)</td>
</tr>
</tbody>
</table>

### Loglikelihood

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>0.0005</td>
<td>0.0061</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0167</td>
<td>0.2866</td>
<td>0.0000</td>
<td>0.2378</td>
</tr>
</tbody>
</table>

Standard errors in parenthesis.
LRT corresponds to the p-value of the LRT of the null assumption of no dynamics in the conditional asymmetry.