Testing Normality: A GMM Approach

Christian Bontemps[a] and Nour Meddahi[b*]

[a] LEEA-CENA, 7 avenue Edouard Belin, 31055 Toulouse Cedex, France.
[b] Département de sciences économiques, CIRANO, CIREQ, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), H3C 3J7, Canada.

First version: March 2001

This version: June 2003

Abstract

In this paper, we consider testing marginal normal distributional assumptions. More precisely, we propose tests based on moment conditions implied by normality. These moment conditions are known as the Stein (1972) equations. They coincide with the first class of moment conditions derived by Hansen and Scheinkman (1995) when the random variable of interest is a scalar diffusion. Among other examples, Stein equation implies that the mean of Hermite polynomials is zero. The GMM approach we adopt is well suited for two reasons. It allows us to study in detail the parameter uncertainty problem, i.e., when the tests depend on unknown parameters that have to be estimated. In particular, we characterize the moment conditions that are robust against parameter uncertainty and show that Hermite polynomials are special examples. This is the main contribution of the paper. The second reason for using GMM is that our tests are also valid for time series. In this case, we adopt a Heteroskedastic-Autocorrelation-Consistent approach to estimate the weighting matrix when the dependence of the data is unspecified. We also make a theoretical comparison of our tests with Jarque and Bera (1980) and OPG regression tests of Davidson and MacKinnon (1993). Finite sample properties of our tests are derived through a comprehensive Monte Carlo study. Finally, two applications to GARCH and realized volatility models are presented.

Key words: Stein equation; Hermite polynomials; parameter uncertainty; HAC.

---

*Corresponding author. Tel: 1-514-343-2399; fax: 1-514-343-7221. E-mail: nour.meddahi@umontreal.ca; Web-page: http://www.mapageweb.umontreal.ca/meddahin/.

1 Introduction

In many econometric models, distributional assumptions play an important role in the estimation, inference and forecasting procedures. Robust estimation methods against distributional assumption are available, such as the Quasi-Maximum-Likelihood (White, 1982; QML) and Generalized Method of Moments (Hansen, 1982; GMM). However, knowing the true distribution of the considered random variable may be useful for improving inference. Such is the case in stochastic volatility models where several studies have shown that simulation and Bayesian methods outperform the QML and GMM methods (Jacquier, Polson and Rossi, 1994; Kim, Shephard and Chib, 1998; Andersen, Chung and Sorensen, 1999; Gallant and Tauchen, 1999). Moreover, knowing the distribution is also crucial when one forecasts nonlinear variables like volatility in the EGARCH model of Nelson (1991), or the high frequency realized volatility model of Andersen, Bollerslev, Diebold and Labys (2001, ABDL). This is also important when one evaluates density forecasts as in Diebold, Gunther and Tay (1998). In continuous time modeling, Chen, Hansen and Scheinkman (2000) argue that an interesting approach is to first specify the unconditional distribution of the process, and then specify the diffusion term. Therefore, developing test procedures for distributional assumption diagnostics in both cross-sectional and time-series settings is of particular interest.

The main purpose of our paper is to provide a new approach for testing normality. We consider normality given its importance in the econometric literature. Moreover, econometricians are more familiar with testing normality. Finally, any continuous distribution may be transformed on a normal one.

There is an important literature on testing normality. This includes tests based on the cumulative distribution function (Kolmogorov, 1933; Smirnov, 1939), the characteristic function (Koutrouvelis, 1980; Koutrouvelis and Kellermeier, 1981; Epps and Pulley, 1983), the moment generating function (Epps, Singleton and Pulley, 1982), the third and fourth moment (Mardia, 1970; Bowman and Shenton, 1975; Jarque and Bera, 1980), and the Hermite polynomials (Kiefer and Salmon, 1983; Hall, 1990; van der Klaauw and Koning, 2001).[1]

Our approach is based on testing moment conditions. The conditions we consider are based on Stein (1972), where it is shown that the marginal distribution of a random variable is normal with zero mean and unit variance if and only if a particular set of moment conditions hold. Each moment condition is known as the Stein equation (see for instance Schoutens, 2000). We show that special examples of this equation correspond to the zero mean of any Hermite polynomial. Interestingly, the Stein equation coincides with the first class of moment conditions given by Hansen and Scheinkman (1995) for continuous time processes when one considers a normal process, that is the Ornstein-Uhlenbeck process.[2]

[1] Multivariate tests are also based on the third and fourth moments (Mardia, 1970; Bera and John, 1983; Richardson and Smith, 1993; Kilian and Demiroglu, 2000; Fiorentini, Sentana and Calzolari, 2003a).

[2] Hansen and Scheinkman (1995) present two classes of moment conditions related respectively to the

We used the GMM approach for testing the Stein equation. The GMM approach is very appealing for two reasons. It is well suited for correcting the test statistic distribution when one uses estimated parameters. Moreover, in the GMM setting, it is easy to take into account potential dependence in the data when one tests marginal moment conditions.

In general, the normality assumption is made for unobservable variables. Hence, one has to estimate the model parameters and then test normality on the fitted variables such as the residuals. As a consequence, one has to take into account the parameter uncertainty, since it is well-known that in general the distribution of the test statistic is not the same when one uses the true parameter and an estimator. This problem leads Lilliefors (1967) to tabulate the Kolmogorov-Smirnov test statistic when one estimates the mean and the variance of the distribution. In the linear homoskedastic model, White and MacDonald (1980) stated that various tests are robust against parameter uncertainty, particularly in tests based on moments that used standardized residuals. Dufour, Farhat, Gardiol and Khalaf (1998) developed Monte Carlo tests to take into account parameter uncertainty in the linear homoskedastic regression model in finite samples. More recently, several solutions have been proposed in the literature: Bai (2002) and Duan (2003) proposed transformations of their test statistics that are robust against parameter uncertainty; Thompson (2002) proposed upper bound critical values for his tests; Hong and Li (2002) used separate inference procedure by splitting the sample; while Corradi and Swanson (2002) used the bootstrap.

It turns out that the GMM setting is well suited for incorporating parameter uncertainty in testing procedures by using Newey (1985) and Tauchen (1985); see also Gallant (1987), Gallant and White (1988), and Wooldridge (1990). In this paper, we show that some testing functions are robust to the parameter uncertainty problem. That is, the asymptotic distribution of the feasible test statistic based on an estimated parameter is identical to that of the test statistic based on the true (unknown) parameter. Hermite polynomials are special examples of functions that have this robustness property. This result is a generalization of Kiefer and Salmon (1983) who showed that tests using Hermite polynomials are robust to parameter uncertainty when one considers a nonlinear homoskedastic regression estimated by the maximum likelihood method. In contrast, our result holds for more general models and for any estimation method. This property is very important when one uses advanced technical methods as in the stochastic volatility case. This result is the main contribution of the paper.

The second reason for using GMM is, when the variable of interest is serially correlated, the GMM setting is also well suited to take into account this dependence by using the Heteroskedastic-Autocorrelation-Consistent (HAC) method of Newey and West (1987) and Andrews (1991). Using a HAC procedure in testing marginal distributions was already adopted by Richardson and Smith (1993) and Bai and Ng (2002) for testing normality, Aït-Sahalia

marginal and conditional distributions of the process. Note, however, that while Hansen and Scheinkman (1995) derived these moment conditions in a Markovian case, we do not make this assumption.

(1996) and Conley, Hansen, Luttmer and Scheinkman (1997) for testing marginal distributions of nonlinear scalar diffusion processes.

The paper is organized as follows. In section 2, we introduce the Stein equation and characterize its relationships with Hermite polynomials and Hansen and Scheinkman (1995) moment conditions. In section 3, we derive the test statistics we consider in both cross-sectional and time series cases. Then, we study the parameter uncertainty problem in section 4. In section 5, we provide an extensive Monte Carlo study in order to assess the finite sample properties of the test statistics we consider and to compare them with the most popular methods, i.e. the Kolmogorov-Smirnov and Jarque-Bera tests. Section 6 applies our theory to two examples from the volatility literature while the last section concludes the paper. All the proofs are provided in the Appendix.

2 The Stein equation

In this section, we first introduce the Stein (1972) equation which will be the basis of the test functions we consider to test normality. Then we specify this equation when one considers the Hermite polynomials. This is important because the most popular normality test in the econometric literature, namely the Jarque and Bera (1980) test, is based on moment conditions on the third and fourth Hermite polynomials. Finally, we relate the Stein equation to the first moment conditions derived by Hansen and Scheinkman (1995) in the case of a continuous time process.

2.1 The Stein equation

Stein (1972) shows that a random variable X has a standard normal distribution $\mathcal{N}(0,1)$ if and only if, for any differentiable function f such that $E|f'(Z)| < +\infty$ where Z is $\mathcal{N}(0,1)$,[3] we have

$$E[f'(X) - Xf(X)] = 0. \tag{2.1}$$

It is straightforward to show that (2.1) holds under normality. Hence, the main result of Stein (1972) is that (2.1) characterizes the normal distribution. The Stein equation (2.1) has several implications, like the recursive moment equation $E[X^{i+1}] = iE[X^{i-1}]$, (with $f(X) = X^i$). It is worth noting that Amemiya (1977) used this equality to show the consistency of the maximum likelihood estimator for nonlinear simultaneous equation models while Davidson and MacKinnon (1984) used it when they developed their specification tests based on double-length artificial linear regressions.

The Stein equation (2.1) is the basic test function we consider for testing normality. This may be applied to monomials, polynomials and more general functions. An important property

[3]Observe that in the normal case, we have $E|Xf(X)| < +\infty$ when $E|f'(X)| < +\infty$.

of the Stein equation is by construction, the expectation of the considered function is zero. Therefore, one does not compute the analytic formula of the moment as one would when he uses, for instance, marginal moments. In other words, if one considers an integrable function g and wants to check that the empirical counterpart of E[g(X)] is close to the theoretical formula, then by using the Stein equation, one will get another function (namely $XG(X)$ where $G(\cdot)$ is any primitive function of $g(\cdot)$) whose population mean equals $E[g(X)]$. Therefore, one can test normality by comparing the empirical counterparts of $E[g(X)]$ and $E[XG(X)]$.[4]

There are some functions of interest for which the Stein equation becomes simple. This is the case for the Hermite polynomials we consider below.

2.2 Hermite polynomials

The normalized Hermite polynomial H_i associated with the $\mathcal{N}(0,1)$ distribution is defined by

$$H_i(x) = \exp(\frac{x^2}{2})\frac{(-1)^i}{\sqrt{i!}}\frac{d^i \exp(-x^2/2)}{dx^i}. \tag{2.2}$$

From (2.2), it is easy to show that the Hermite polynomials are given by the recursive formula

$$\forall i > 1, H_i(x) = \frac{1}{\sqrt{i}}\{xH_{i-1}(x) - \sqrt{i-1}H_{i-2}(x)\},\ H_0(x) = 1,\ H_1(x) = x. \tag{2.3}$$

By applying (2.3), we have

$$H_2(x) = \frac{1}{\sqrt{2}}(x^2-1);\ H_3(x) = \frac{1}{\sqrt{6}}(x^3-3x);\ H_4(x) = \frac{1}{\sqrt{24}}(x^4-6x^2+3). \tag{2.4}$$

When a random variable X follows a normal distribution $\mathcal{N}(0,1)$, the transformed random variables $H_i(X)$, $i = 0, 1, ...$, have some interesting properties. In particular, they are orthonormal, that is:

$$E[H_i(X)H_j(X)] = \delta_{ij}, \tag{2.5}$$

where δ_{ij} is the Kronecker symbol. By applying (2.5) to $j = 0$ and $i \neq 0$, one gets

$$\forall i > 0,\ E[H_i(X)] = 0, \tag{2.6}$$

that is, the Hermite polynomials $H_i(X)$ are centered for $i > 0$.

In order to characterize the relationships between the Stein equation (2.1) and the Hermite polynomials, note that (2.2) implies the following restrictions are fulfilled by the derivatives of the Hermite polynomials:

$$H_i^{'}(x) = \sqrt{i}H_{i-1}(x) \quad \text{and} \quad H_i^{''}(x) - xH_i^{'}(x) + iH_i(x) = 0. \tag{2.7}$$

[4] Another solution is to define, when it is possible, the function $h(X) \equiv g(X)/X$. In this case, if one can use the function h in the Stein equation, then one gets $E[g(X)] = E[Xh(X)] = E[h'(X)]$.

Let us now apply the Stein equation (2.1) to the function $H_i'(x)/\sqrt{i}$. This function is clearly differentiable and integrable. Therefore, we have

$$\frac{1}{\sqrt{i}}E[H_i''(X) - XH_i'(X)] = 0,$$

which implies (2.6) by using the second result in (2.7). As a consequence, the Stein equation (2.1) implies (2.6). It turns out that the converse also holds:

Proposition 2.1 *Let X be a random variable such that $\forall i > 0$, $E[H_i(X)] = 0$. Then, the equation (2.1) holds for any differentiable function f such that $E[|f'(Z)|] < +\infty$ where Z is assumed to be $\mathcal{N}(0,1)$. Consequently, a random variable X is $\mathcal{N}(0,1)$ if and only if (2.6) holds.*

This result is established by Gallant (1980, Theorem 3, page 192) for any distribution that admits some polynomials as a basis of the space of the square-integrable functions, which is the case for the normal distribution. We therefore dropped it from this paper; see however the previous version, Bontemps and Meddahi (2002), for a proof. This proposition means that for statistical inference purposes, in particular testing, one could use Hermite polynomials only.

2.3 Continuous time case

Consider a univariate diffusion process X_t assumed to be the stationary solution of

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{2.8}$$

where W_t is a standard Brownian process. Then, Hansen and Scheinkman (1995) provide two sets of moment conditions related to the marginal and conditional distributions of X_t respectively. For the marginal distribution, Hansen and Scheinkman (1995) show that

$$E[\mathcal{A}g(X_t)] = 0, \tag{2.9}$$

where g is assumed to be twice differentiable and square-integrable with respect to the marginal distribution of X_t and $\mathcal{A}$ is the infinitesimal generator associated with the diffusion (2.8), that is:

$$\mathcal{A}g(x) = \mu(x)g'(x) + \frac{\sigma^2(x)}{2}g''(x). \tag{2.10}$$

A well-known continuous time process for which the marginal distribution is $\mathcal{N}(0,1)$ is the standardized Ornstein-Uhlenbeck process defined by

$$dX_t = -kX_t dt + \sqrt{2k}dW_t, \quad k > 0, \quad X_0 \sim \mathcal{N}(0,1). \tag{2.11}$$

For this process, Hansen and Scheinkman (1995) moment condition (2.9) becomes:

$$E[-kX_t g'(X_t) + kg''(X_t)] = 0. \tag{2.12}$$

Thus, by considering the function f defined by $f \equiv g'$, we obtain the Stein equation (2.1) (since $k \neq 0$). Thus, the Hansen and Scheinkman (1995) moment condition (2.9) coincides with the Stein equation (2.1).

The continuous time setting provides examples of processes where the marginal distribution is normal while the conditional distribution is not. A first example may be constructed as follows. For a given specification of $\mu(x)$ and $\sigma(x)$, the marginal density function of the process X_t is, up to a scale,[5]

$$\sigma(x)^{-2} \exp\left(\int_z^x \frac{2\mu(u)}{\sigma(u)^2} du\right).$$

This density function suggests that two different specifications of $\mu(x)$ and $\sigma(x)$ may give the same marginal distribution. It turns out that this is the case.[6] As a consequence, it is possible to get a scalar diffusion such that the marginal distribution is $\mathcal{N}(0,1)$ while the conditional distribution is not normal, that is a non Ornstein-Uhlenbeck process.

A second example may be obtained by subordination. More precisely, assume that we observe a sample $x_1, x_2, ..., x_T$ of a process X_t with $X_t = Y_{S_t}$, where Y_t is a stationary scalar diffusion and S_t, $t = 1, ..., T$, is a positive and increasing process with $S_1 = 1$. Under the assumption that the processes $\{Y_\tau, \tau \in \mathbb{R}^+\}$ and $\{S_t, t \in \mathbf{N}^*\}$ are independent, the marginal distribution of the processes X_t and Y_t coincide. Therefore, if the process Y_t is a standardized Ornstein-Uhlenbeck process, the marginal distribution of X_t is $\mathcal{N}(0,1)$ while its conditional distribution is (in general) not normal.

3 Test statistics

In this section, we provide the test statistics for testing normality. All of them are based on the Stein equation (2.1). We study in detail the cross-sectional and the time series cases. We assume that we observe a sample of the random variable of interest, i.e., we do not take into account the potential problem of parameter uncertainty (studied in the next section).

3.1 The general case

Consider a sample $x_1, ..., x_T$, of the variable of interest denoted by X. The observations may be independent or dependent. We assume that the marginal distribution of X is $\mathcal{N}(0,1)$. Let f_1, ..., f_p, be p differentiable functions such that f_i' is integrable. For a real x, define the vector $g(x) \in \mathbb{R}^p$, whose components are $(f_i'(x) - x f_i(x))$ for $i = 1, ..., p$. Thus, by the Stein equation (2.1), we have $E[g(X)] = 0$. Throughout the paper, we assume that any component of the

[5]For a given z, the scale parameter is chosen so that the density integral equals one.

[6]See Aït-Sahalia, Hansen and Scheinkman (2001) for a review of all the properties of diffusion processes we consider in this paper.

vector $g(X)$ is square-integrable and that the matrix Σ defined by

$$\Sigma \equiv \lim_{T \to +\infty} Var[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(x_t)] = \sum_{h=-\infty}^{+\infty} E[g(x_t) g(x_{t-h})^\top], \tag{3.1}$$

is finite and positive definite, then we have (see Hansen, 1982)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(x_t) \longrightarrow \mathcal{N}(0, \Sigma) \tag{3.2}$$

while

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(x_t) \right)^\top \Sigma^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(x_t) \right) \sim \chi^2(p). \tag{3.3}$$

For the feasibility of the test procedure, one needs the matrix Σ or at least a consistent estimator. It is clear that if one does not specify the dependence between the observations x_1, x_2,..., x_T, then one needs to estimate Σ.

3.2 The cross-sectional case

Consider the cross-sectional case and assume that the observations are independent and identically distributed (i.i.d.). In this case, we have

$$\Sigma = Var[g(X)] = E[g(X) g(X)^\top].$$

Observe that by applying the Stein equation (2.1) to $\phi(X) = X f(X) f(X)^\top$, one can show that

$$E[g(X) g(X)^\top] = E[f(X) f(X)^\top + f'(X) f'(X)^\top]. \tag{3.4}$$

Two cases may arise. In the first case, one can explicitly compute the matrix Σ and, hence, one can use the test statistic (3.3). This is the case for the Hermite polynomials that we consider below. In the second case, computing Σ explicitly is not possible (or difficult), then one can use any consistent estimator of Σ and denoted by $\hat{\Sigma}_T$, like

$$\hat{\Sigma}_{1,T} = \frac{1}{T} \sum_{t=1}^{T} g(x_t) g(x_t)^\top \text{ or } \hat{\Sigma}_{2,T} = \frac{1}{T} \sum_{t=1}^{T} \left(f(x_t) f(x_t)^\top + f'(x_t) f'(x_t)^\top \right). \tag{3.5}$$

In this case, one can use the following test statistic

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(x_t) \right)^\top \hat{\Sigma}_T^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(x_t) \right) \sim \chi^2(p). \tag{3.6}$$

Assume now that we consider the Hermite polynomials. We have shown in the previous section that when one applies the Stein equation (2.1) to the function $f_i(x) = H'_{i+1}(x)/\sqrt{i}$, one gets $E[H_i(X)] = 0$. But the unconditional variance of $H_i(x)$ is one. Hence, for $i \geq 0$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_i(x_t) \longrightarrow \mathcal{N}(0,1) \quad \text{and} \quad \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_i(x_t)\right)^2 \sim \chi^2(1). \tag{3.7}$$

Moreover, the Hermite polynomials are orthogonal. Hence, the test statistic based on different Hermite polynomials are asymptotically independent. In other words, when one uses a test statistic based on several Hermite polynomials, the corresponding matrix Σ derived previously is diagonal. The diagonal matrix Σ is indeed the identity since the variance of each Hermite polynomial equals one. For instance, if we consider the Hermite polynomials $H_3, H_4, ..., H_p$, then the test statistic is

$$\sum_{i=3}^{p}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_i(x_t)\right)^2 \sim \chi^2(p-2). \tag{3.8}$$

It is worth noting that this result is more general than the one obtained by Kiefer and Salmon (1982). They established (3.8) when the variables x_t are estimated residuals in a linear model when one uses the maximum likelihood method. We will discuss Kiefer and Salmon's (1982) results in more detail in the next section where we consider the parameter uncertainty problem.

3.3 The time series case

Assume now that the observations are correlated and represent a sample of a process. Then, without additional assumptions on the dependence, one can not explicitly compute the matrix Σ and has to estimate it. A traditional solution is to estimate this matrix by using a Heteroskedastic-Autocorrelation-Consistent (HAC) method like Newey and West (1987) or Andrews (1991). This is one of the motivations of using a GMM approach for testing normality. This has been used by Smith and Richardson (1991), and was more recently and independently of our work, highlighted by Bai and Ng (2002).

In contrast to the cross-sectional case, one can not show that test statistics based on two different Hermite polynomials are asymptotically independent. More precisely, consider a component (i, j), with $i \neq j$, of the matrix Σ. In this case, $E[H_i(x_t)H_j(x_t)]$ is zero by the orthogonality of the Hermite polynomials (2.5). However, without additional restrictions, $E[H_i(x_t)H_j(x_{t-h})]$ is in general nonzero for $h \neq 0$ and the matrix Σ will be non-diagonal. This is probably the case for the following scalar diffusion whose marginal distribution is $\mathcal{N}(0,1)$:

$$dx_t = \frac{1}{2}x_t(1-x_t^2)dt + \sqrt{1+x_t^2}dW_t.$$

In contrast, the asymptotic independence of the tests may hold if one makes additional assumptions on the dependence of the process x_t. An important example is when one assumes that the process x_t is a normal autoregressive process of order one, AR(1), that is

$$x_t = \gamma x_{t-1} + \sqrt{1-\gamma^2}\, \varepsilon_t, \quad \varepsilon_t \text{ is i.i.d. and } \sim \mathcal{N}(0,1), \quad \text{and} \quad |\gamma| < 1. \tag{3.9}$$

In this case, any Hermite polynomial $H_i(x_t)$ is an AR(1) process whose autoregressive coefficient equals γ^i, that is

$$E[H_i(x_{t+1})|x_\tau, \tau \leq t] = \gamma^i H_i(x_t). \tag{3.10}$$

In this case, it is easy to show that

$$\Sigma_{ij} = \sum_{h=-\infty}^{+\infty} E[H_i(x_t) H_j(x_{t-h})] = \frac{1+\gamma^i}{1-\gamma^i}\delta_{ij}. \tag{3.11}$$

As a consequence, the matrix Σ is diagonal and, hence, the test statistics based on different Hermite polynomials are asymptotically independent.[7] Besides, when one tests normality and ignores the dependence of the Hermite polynomials, one gets a wrong distribution for the test statistic. For instance, assume that one considers a test based on a particular Hermite polynomial H_i. Then, the test statistic becomes

$$\frac{1-\gamma^i}{1+\gamma^i} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} H_i(x_t) \right)^2 \sim \chi^2(1). \tag{3.12}$$

Thus, by ignoring the dependence of the Hermite polynomial $H_i(x_t)$, one overrejects the normality when $\gamma \geq 0$ or i is even and underrejects otherwise. Monte Carlo simulations in the sixth section will assess this. This is important in practice since many economic time series are positively autocorrelated.

Allowing for potential serial correlation of unknown form in testing normality is particularly important when one uses a nonlinear transform to get the variable of interest. For instance, in the density forecast analysis (e.g., Diebold, Gunther and Tay, 1998), the distribution of interest is not a Gaussian one, and therefore one uses a transform (through the cumulative distribution functions of the variable of interest and the Gaussian variable) to get a normal distribution.[8] For this example, even if the marginal distribution is Gaussian, there is no reason to assume that this is also the case for the conditional distribution.

3.4 Skewness and excess kurtosis

A traditional approach for testing normality is to study the skewness and excess kurtosis of the variable of interest (Mardia, 1970; Bowman and Shenton, 1975; Jarque and Bera, 1980).

[7]This result also holds when one assumes that the process $\{x_t\}$ is Gaussian, which implies that (x_t, x_{t-h}) is Gaussian, $\forall\ t,\ h$. Bai and Ng (2002) used this assumption when they showed the asymptotic independence of the skewness and excess kurtosis tests.

[8]Let Y be a continuous random variable with cumulative distribution function $F(\cdot)$; then $\Phi^{-1}(F(X))$ is a $\mathcal{N}(0,1)$, where $\Phi(\cdot)$ is the cumulative distribution function of $\mathcal{N}(0,1)$ random variable.

More precisely, when a random variable X is distributed as $\mathcal{N}(0,\sigma^2)$, we have

$$E[X^3]=0 \text{ and } E[X^4-3\sigma^4]=0, \tag{3.13}$$

where the first condition deals with skewness while the second one deals with excess kurtosis. For simplicity, assume that we observe an i.i.d. sample $x_1, x_2, ..., x_T$. Then, when the parameter σ^2 is known, the test statistic implied by moment condition in (3.13) is

$$\sqrt{T}\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} x_t^3 \\ \frac{1}{T}\sum_{t=1}^{T} x_t^4 - 3\sigma^4 \end{pmatrix} \overset{n\to+\infty}{\longrightarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 15\sigma^6 & 0 \\ 0 & 96\sigma^8 \end{pmatrix}\right). \tag{3.14}$$

Thus, skewness and excess kurtosis test statistics are asymptotically independent. Moreover, we have the following test statistic

$$\frac{1}{15}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t/\sigma)^3\right)^2 + \frac{1}{96}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(x_t/\sigma)^4-3]\right)^2 \overset{n\to+\infty}{\longrightarrow} \chi^2(2). \tag{3.15}$$

For future reference, observe that this test statistic is different from one that Jarque and Bera (1980) had given by

$$\frac{1}{6}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t/\hat{\sigma})^3\right)^2 + \frac{1}{24}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(x_t/\hat{\sigma})^4-3]\right)^2 \overset{n\to+\infty}{\longrightarrow} \chi^2(2), \tag{3.16}$$

where $\hat{\sigma}$ is the MLE of σ in a regression model. This difference is due to parameter uncertainty that we consider in the following section.

4 Parameter uncertainty

In most empirical examples, the normality assumption is made for an unobservable random variable. This is the case for a regression, linear or nonlinear, homoskedastic or heteroskedastic, where the normality assumption is in general made on the (standardized) residuals. This is also the case for nonlinear time series as volatility models (e.g., GARCH or stochastic volatility models). Thus, one must first estimate the parameters of the model and then get fitted residuals. Then, one tests the normality assumption of the residuals by using the fitted residuals. In other empirical examples, the normality assumption is made on observable variables but the parameter of the normal distribution, i.e., the mean and the variance, are unknown. Therefore, one must also estimate these parameters in order to test normality.

It is well-known that the asymptotic distribution of a test statistic that depends on an unknown parameter, denoted by θ^0, may be different from the asymptotic distribution of the same test statistic applied by using a consistent estimator of θ^0, denoted by $\hat{\theta}_T$. The main reason is one has to take into account the uncertainty of $\hat{\theta}_T$ in the testing procedure. This is known as the parameter uncertainty problem.

4.1 The traditional approach

The GMM approach is well-suited for this problem, which is the first reason we are using it in the paper to test normality. Newey (1985) and Tauchen (1985) provided a general theory for taking into account the parameter uncertainty in testing procedures. Their approach is the following. Assume that one has to test the following moment condition

$$E[g(z_t, \theta^0)] = 0, \tag{4.1}$$

where z_t is a random variable, potentially multivariate, and θ^0 is an unknown (vectorial) parameter. Under the null hypothesis, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \theta^0) \ \rightarrow \ \mathcal{N}(0, \Sigma_g) \text{ where } \Sigma_g = \lim_{T \rightarrow +\infty} Var[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \theta^0)]. \tag{4.2}$$

Assume that one has a square-root T consistent estimator of θ^0, denoted by $\hat{\theta}_T$, i.e.

$$\sqrt{T}(\hat{\theta}_T - \theta^0) \quad \rightarrow \quad \mathcal{N}(0, V_\theta). \tag{4.3}$$

Then, a natural approach to test (4.1) is by using $T^{-1/2} \sum_{t=1}^{T} g(z_t, \hat{\theta}_T)$. Therefore, one needs the asymptotic distribution of this test statistic. It is easily obtained by using a Taylor approximation around the unknown parameter θ^0. More precisely, we have:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \hat{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \theta^0) + \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(z_t, \theta^0)}{\partial \theta^\top} \right] \sqrt{T}(\hat{\theta}_T - \theta^0) + o_p(1). \tag{4.4}$$

Define the matrix P_g by

$$P_g = \lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g(z_t, \theta^0)}{\partial \theta^\top}. \tag{4.5}$$

Then, we can rewrite (4.4) in the following form:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \hat{\theta}_T) = [I_p \ P_g] \ \left[\begin{array}{c} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(z_t, \theta^0) \\ \sqrt{T}(\hat{\theta}_T - \theta^0) \end{array} \right] + o_p(1), \tag{4.6}$$

where I_p is the $p \times p$ identity matrix and p the dimension of g. From (4.6), it is clear that the asymptotic distribution of the test statistic in the left hand side of (4.6) depends on the asymptotic distributions of two random variables, $T^{-1/2} \sum_{t=1}^{T} g(z_t, \theta^0)$ and $\sqrt{T}(\hat{\theta}_T - \theta^0)$ which are given respectively in (4.2) and (4.3), and their asymptotic covariance.[9] As a consequence, the parameter uncertainty generally changes the asymptotic distribution of the test statistic when one uses an estimator instead of the unknown parameter θ^0. This is why the Jarque

[9]We assume that the asymptotic distribution of the right hand side of (4.6) is normal.

and Bera (1980) test (3.16) differs from (3.15). Bai and Ng (2002) also adopted the same approach by including the third and fourth moment conditions in their estimation procedure by the GMM.

In general, the matrices that appear in the asymptotic distributions (4.2) and (4.3) are easily estimated. However, it is difficult to estimate the asymptotic covariance matrix between $T^{-1/2}\sum_{t=1}^{T} g(z_t, \theta^0)$ and $\sqrt{T}(\hat{\theta}_T - \theta^0)$. This is the case when one uses advanced estimation methods, especially simulation techniques[10] as in the case of stochastic volatility and latent factor models. Several solutions have been proposed in the literature to solve this problem. Bai (2002) as well as Duan (2003) proposed transformations of their test statistics that are robust against parameter uncertainty; Thompson (2002) proposed upper bound critical values for his tests; Hong and Li (2002) split their sample and used the first part of the sample to do the estimation and made the test on the second part of the sample whose sample size, T_2, is assumed to be small with respect to the total sample size T (i.e., $T_2/T \to 0$ when $T \to +\infty$); Corradi and Swanson (2002) used the bootstrap to take into account parameter uncertainty (and serial correlation); finally, Dufour, Farhat, Gardiol and Khalaf (1998) developed Monte Carlo exact tests in the linear homoskedastic regression model in finite samples.

4.2 Our approach

An alternative method that we adopt in this paper is to consider moment conditions such that the matrix P_g equals zero, i.e.,

$$P_g = 0. \tag{4.7}$$

In this case, the asymptotic distribution of $T^{-1/2}\sum_{t=1}^{T} g(z_t, \theta^0)$ and $T^{-1/2}\sum_{t=1}^{T} g(z_t, \hat{\theta}_T)$ coincide. Hence, the test statistic is robust against the parameter uncertainty.

In the sequel, we need to be more specific about the examples we will consider in order to characterize the moment conditions that are robust against the parameter uncertainty. We consider three examples:

Example 1: regression with exogenous variables. Let $z_t = (y_t, x_t^\top)^\top$ be a vector where y_t is an endogenous variable, x_t is a (vectorial) exogenous variable. We assume that there exists a unique parameter $\theta^0 = (\alpha^{0\top}, \beta^{0\top})^\top$ such that

$$y_t = m(x_t, \beta^0) + \sigma(x_t, \beta^0, \alpha^0)\ u_t, \quad \text{and } u_t \sim \mathcal{N}(0,1), \tag{4.8}$$

where α^0 and β^0 are real vectors and $m(x, \beta)$ and $\sigma(x, \beta, \alpha)$ are two real functions. A special example is the cross-sectional case where the random variable u_t is i.i.d. by assumption. Another example is the time series case where the variable u_t may be serially correlated. However, it is assumed to be independent from x_t.

[10]See Gouriéroux and Monfort (1996) for a review.

The model adopted by Jarque and Bera (1980) is the special case where

$$m(x_t, \beta) = x_t^\top \beta \quad \text{and} \quad \sigma(x_t, \beta, \alpha) = \alpha, \tag{4.9}$$

i.e., they considered a linear homoskedastic regression model with potentially correlated residuals. Kiefer and Salmon (1983) also adopted a special case of (4.8) by assuming

$$\sigma(x_t, \beta, \alpha) = \alpha \quad \text{and} \quad u_t \text{ is i.i.d.}, \tag{4.10}$$

i.e., a nonlinear regression model with homoskedastic and i.i.d. errors.

Example 2: time series regression. This example is similar to the first one, but we now assume that the variables x_t are lagged values of y_t and u_t is i.i.d., i.e.,[11]

$$E[y_t \mid y_\tau, \tau \leq t-1] = m_t(\beta^0), \; Var[y_t \mid y_\tau, \tau \leq t-1] = \sigma_t^2(\beta^0, \alpha^0), \tag{4.11}$$

$$u_t \equiv \frac{y_t - m_t(\beta^0)}{\sigma_t(\beta^0, \alpha^0)} \quad \text{and} \quad u_t \text{ is i.i.d. and } \sim \mathcal{N}(0,1). \tag{4.12}$$

Special examples of this case are ARMA models with GARCH errors (Bollerslev, 1986).

Example 3: marginal distribution of a process. In this case, we assume that we observe a sample $y_1, ..., y_T$, of a process whose marginal distribution is assumed to be $\mathcal{N}(m^0, \sigma^{02})$ where m^0 and σ^0 are unknown parameters. Hence, the standardized process is $\mathcal{N}(0,1)$, that is

$$u_t \equiv \frac{y_t - m^0}{\sigma^0} \text{ and } \; u_t \; \sim \; \mathcal{N}(0,1). \tag{4.13}$$

Observe that in all of these examples, the normal variable of interest u_t may be written as

$$u_t(\theta) = \frac{y_t - m_t(\theta)}{\sigma_t(\theta)}, \tag{4.14}$$

where the normality assumption holds for $u_t(\theta^0)$ and denoted by u_t. We can now characterize the test functions g that are robust against parameter uncertainty in Examples 1, 2 or 3.

Proposition 4.1 *Consider u_t as defined in Examples 1, 2 or 3. Let $\hat{\theta}_T$ be a square-root T consistent estimator of θ^0 such that (4.3) applies and denote by $\hat{u}_t$ the corresponding estimated residuals. Define the function $\tilde{g}(.)$ by $\tilde{g}(u_t(\theta)) = g(z_t, \theta)$. Then, a sufficient condition such that the asymptotic distribution of the test statistics $T^{-1/2} \sum_{t=1}^T g(z_t, \theta^0)$ and $T^{-1/2} \sum_{t=1}^T g(z_t, \hat{\theta}_T)$ coincide is*

$$E[\tilde{g}'(u_t)] = 0 \quad \text{and} \quad E[u_t \tilde{g}'(u_t)] = 0. \tag{4.15}$$

[11] Observe that we adopt a different notation than used in the first example to incorporate non Markovian models like MA and GARCH.

This proposition means that a sufficient condition ensuring the robustness of our test statistic against the parameter uncertainty is the orthogonality of $\tilde{g}'$ with H_0 and H_1,[12] i.e.,

$$E[H_0(u_t)\tilde{g}'(u_t)] = 0 \quad \text{and} \quad E[H_1(u_t)\tilde{g}'(u_t)] = 0. \tag{4.16}$$

It is worth noting that we do not assume that the considered test statistic comes from the Stein equation (2.1). Indeed, this result encompasses the results of White and MacDonald (1980). Besides, this proposition holds in both cross-sectional and time series cases. Finally, while (4.15) is a sufficient condition, it is generically necessary. It may not be necessary for some estimators with very particular asymptotic variances.

Before further characterizing (4.16) when one considers the Stein equation (2.1), let us apply this proposition when one considers tests based on excess skewness and kurtosis as did Jarque and Bera (1980) and Bai and Ng (2002). More precisely, assume that one considers the moment conditions

$$E[\tilde{g}_1(u_t)] = 0 \text{ and/or } E[\tilde{g}_2(u_t)] = 0, \text{ where } \tilde{g}_1(u_t) = u_t^3 \text{ and } \tilde{g}_2(u_t) = u_t^4 - 3. \tag{4.17}$$

It is clear that the function $\tilde{g}_1$ violates the first condition in (4.15) while $\tilde{g}_2$ violates the second one. Thus, in this case, one must correct the asymptotic distribution of the test statistic by taking into account the parameter uncertainty as did Jarque and Bera (1980).

When one considers test statistics based on the Stein equation (2.1), that is when one assumes that

$$\tilde{g}(x) = f'(x) - xf(x),$$

the condition (4.15) may be characterized through the function $f(x)$:

Proposition 4.2 *Let $f(x)$ be a differentiable function and define $\tilde{g}(x)$ by $\tilde{g}(x) \equiv f'(x) - xf(x)$. Then, the condition (4.16) holds if and only if*

$$E[f(u_t)] = 0 \quad \text{and} \quad E[f'(u_t)] = 0. \tag{4.18}$$

This proposition may be easily applied in practice. One has to take any integrable function denoted by $s(x)$ such that $E[|s(Z)|] < +\infty$ where Z is assumed to be $\mathcal{N}(0,1)$. Then, define the function $\bar{s}(x)$ by $\bar{s}(x) = s(x) - E[s(Z)]$ and the function $\bar{f}(x)$ as the primitive of $\bar{s}(x)$ which is centered, that is $E[\bar{f}(Z)] = 0$. Then, by construction, the condition (4.18) holds for $\bar{f}(x)$.

When on uses the conditions based on the Hermite polynomials (2.6), the conditions (4.15) and (4.18) hold for any (linear combination of) Hermite polynomial $H_i(x)$ with $i \geq 3$. This is the main result of our paper:

[12] Observe that we explicitly use the form (4.14) of u_t. If one considers a more general framework like $u_t = f(z_t, \theta^0)$ as in Davidson and MacKinnon (1984), and want to test $E[\tilde{g}(u_t)] = E[\tilde{g}(f(z_t, \theta^0))] = 0$, the sufficient condition (4.15) becomes $E[\tilde{g}'(u_t)\dfrac{\partial f}{\partial \theta^\top}(z_t, \theta^0)] = 0$.

Proposition 4.3 *Consider u_t as defined in Example 1, 2 or 3. Let $\hat{\theta}_T$ be a square-root T consistent estimator of θ^0 such that (4.3) applies. Let g be a vectorial function such that any component is a linear combination of Hermite polynomials $H_i(x)$ with $i \geq 3$. Then, the asymptotic distribution of the test statistics $T^{-1/2}\sum_{t=1}^{T} g(z_t, \theta^0)$ and $T^{-1/2}\sum_{t=1}^{T} g(z_t, \hat{\theta}_T)$ coincide.*

This result was already stated in Kiefer and Salmon (1983). However these authors assumed that the model is a nonlinear regression with homoskedastic and i.i.d. errors, that is under (4.8) and (4.10). Moreover, they found this result when $\hat{\theta}_T$ is the maximum likelihood estimator. In other words, both assumptions are relaxed in the previous proposition. This is very important in many empirical examples where computing the maximum likelihood estimator is difficult or unfeasible.

We now characterize the relationship of the Jarque and Bera (1980) test with the previous proposition. The test statistic they proposed is, for example, one under (4.9). More precisely, let $\hat{u}_t$ defined by $\hat{u}_t = (y_t - x_t^\top \hat{\beta})/\hat{\alpha}$, where $\hat{\alpha}$ and $\hat{\beta}$ are the MLE of α and β. Then, Jarque and Bera (1980) found that

$$JB \equiv \frac{1}{6}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{u}_t^3\right)^2 + \frac{1}{24}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^4 - 3]\right)^2 \overset{n\to+\infty}{\longrightarrow} \chi^2(2). \tag{4.19}$$

In Jarque and Bera (1980), the constant is in the regressors and $\hat{\alpha}^2$ is given by $\hat{\alpha}^2 = T^{-1}\sum_{t=1}^{T}(y_t - x_t^\top\hat{\beta})^2$. Therefore, $\sum_{t=1}^{T}\hat{u}_t = 0$ and $T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 = 1$, and the JB test-statistic (4.19) becomes

$$\begin{aligned} JB &= \frac{1}{6}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^3 - 3\hat{u}_t]\right)^2 + \frac{1}{24}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^4 - 6\hat{u}_t^2 + 3]\right)^2, \\ &= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_3(\hat{u}_t)\right)^2 + \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_4(\hat{u}_t)\right)^2. \end{aligned}$$

In other words, the Jarque and Bera (1980) test coincides with the joint test based on third and fourth Hermite polynomials. However, the setting they considered is less general than ours. In addition, their estimation method is the ML while in our case we only need a square-root T consistent estimator.

In summary, when one wants to test normality, $\mathcal{N}(0,1)$, through skewness and excess kurtosis, one has two methods that are robust against parameter uncertainty.[13] One can either use the third and fourth Hermite polynomials on the fitted residuals whatever the estimation method, or, use the Jarque and Bera (1980) test on the standardized residuals, i.e., the fitted residuals minus their empirical mean divided by their standard deviation.[14] Note that these two

[13]Tests based on OPG regression is a third robust method; see the following subsection.

[14]Of course, for time series, the second method is not valid while one has to use a HAC method for estimating the variance-covariance matrix in the first method.

methods are sufficient only. For instance, Fiorentini, Sentana and Calzolari (2003b) established that for some symmetric and heteroskedastic models (like GARCH) estimated by the ML method, the Jarque and Bera (1980) test statistic is still valid.

4.3 The OPG regression approach

Another approach to handle the parameter uncertainty problem is the use of outer-product-of-the-gradient (OPG) regressions. Indeed, Davidson and MacKinnon (1993) considered the OPG regression approach for testing normality through skewness and excess kurtosis.

More precisely, as in Davidson and MacKinnon (1993), assume that one is interested in testing that y_t is $\mathcal{N}(\beta, \sigma^2)$ where σ^2 is a known parameter and define x_t by $x_t = y_t - \beta$. The assumption that the mean and variance of x_t are zero and σ^2 could be tested by the following OPG regression

$$1 = s_1 x_t + s_2(x_t^2 - \sigma^2) + \text{residual}. \tag{4.20}$$

Assume now that one is interested in testing skewness. Then, Davidson and MacKinnon (1993) propose to add in (4.20) the regressor x_t^3, i.e.,

$$1 = s_1 x_t + s_2(x_t^2 - \sigma^2) + a x_t^3 + \text{residual},$$

and to test that the coefficient a is zero. The test statistic being the t statistic of the estimator of a. Given that x_t^3 is orthogonal with $x_t^2 - \sigma^2$ and not with x_t, the numerator of the t test is the mean of x_t^3 minus the mean of its projection on x_t, i.e.,

$$\frac{1}{T}\sum_{t=1}^{T} x_t^3 - \left(\frac{\sum_{t=1}^{T} x_t^4}{\sum_{t=1}^{T} x_t^2}\right)\frac{1}{T}\sum_{t=1}^{T} x_t \quad \approx \quad \frac{1}{T}\sum_{t=1}^{T} x_t^3 - 3\sigma^2\frac{1}{T}\sum_{t=1}^{T} x_t.$$

where the last approximation holds under normality and when T is large. Davidson and MacKinnon (1993) show the variance of $x_t^3 - 3\sigma^2 x_t$ is $6\sigma^6$. Thus, the t statistic is close to

$$\frac{\frac{1}{T}\sum_{t=1}^{T} x_t^3 - 3\sigma^2\frac{1}{T}\sum_{t=1}^{T} x_t}{\sqrt{6\sigma^6}} = \frac{1}{T}\sum_{t=1}^{T} H_3(x_t/\sigma). \tag{4.21}$$

Given that the variance of $H_3(x_t/\sigma)$ is one, the test statistic (4.21) is also the t-statistic of the coefficient $\tilde{a}$ in the two OPG regressions

$$\begin{aligned} 1 &= s_1 x_t + s_2(x_t^2 - \sigma^2) + \tilde{a} H_3(x_t/\sigma) \ + \ \text{residual}, \quad \text{and} \\ 1 &= \tilde{s_1} H_1(x_t/\sigma) + \tilde{s_2} H_2(x_t/\sigma) + \tilde{a} H_3(x_t/\sigma) \ + \ \text{residual}. \end{aligned}$$

Therefore, testing that the empirical mean of the third Hermite polynomial is numerically the same as testing that the coefficient $\tilde{a}$ in the previous regressions is zero if one takes into account the orthonormality of the Hermite polynomials (in particular the unit variance of the Hermite polynomials). When one ignores this orthonormality, one gets an asymptotic equivalence.

Similarly, one easily obtains the same results when one tests excess kurtosis by using the fourth Hermite polynomial. In addition, the orthogonality between the third and fourth Hermite polynomials implies that the same result holds when one tests jointly the skewness and excess kurtosis through the third and fourth Hermite polynomials. More generally, the orthonormality of the Hermite polynomials means that the result holds for any and higher order Hermite polynomials; for more details, see the previous version, Bontemps and Meddahi (2002).

5 A Monte Carlo study

This section provides some Monte Carlo experiments to study the finite sample properties of the tests we proposed. We also compare our tests with the more popular ones, that is the Kolmogorov-Smirnov, Jarque-Bera and OPG-type tests. The first two tests are respectively denoted by KS and JB in the tables. Note that when we consider the parameter uncertainty problem, we also provide the Lilliefors modified Kolmogorov-Smirnov test and denote these by M-KS in the tables.

All the test functions we consider are based on Hermite polynomials given their generality (Proposition 2.1) and their robustness against parameter uncertainty. We consider test functions based on individual Hermite polynomials H_i for $i = 3, ..., 10$. We also consider joint tests based on $(H_3, H_4, ..., H_i)$, for $i = 4, ..., 10$. These tests are denoted by H_{3-i} for $i = 4, ..., 10$ in the tables.

In all the simulation experiments, we consider four sample sizes: 100, 250, 500 and 1000. All the results are based on 50000 replications. We report in the tables the empirical probability of rejecting the null hypothesis when one considers tests at 5% significance level. Tests based on 10%, 2.5% and 1% produce similar results to that based on 5% and are omitted to save space.[15]

Cross-sectional case. We start by simulating an i.i.d. sample from a $\mathcal{N}(0,1)$. We assume that we know the mean and the variance. Obviously, this is unrealistic in practice. However, it is a good benchmark for the realistic cases where the parameters are unknown and have to be estimated. We report the results in Panel A of Table 1. Consider the tests based on individual Hermite polynomials. Their finite sample properties are clearly good. In particular, they do not reject the null more than the nominal level, even with the smaller sample size. However, while the tests based on higher order polynomials H_i for $i \geq 6$ underreject the null for the four sample sizes, this is not problematic given that we are considering the level of the tests. The tests based on several Hermite polynomials also have very good finite sample properties for

[15]They are available upon request from the authors.

the four sample sizes. Indeed, we do not observe the underrejection when we use high order Hermite polynomials.

Consider now the popular tests, i.e., KS and JB. The KS test works very well whatever the sample size. Interestingly, the properties of tests based on the Hermite polynomials are very close to the KS test and occasionally better for the sample size 100 when one considers a test based on H_4. However the JB test, denoted Naive-JB, does not work well and overreject the null. The main reason is, when the empirical mean of the sample is not zero, the asymptotic distribution of the Jarque and Bera (1980) test is not a $\chi^2(2)$ (which justifies the terminology Naive-JB).

In Panel B of Table 1, we present the results of the same tests[16] on the same samples when one does not know the mean and the variance and estimates them. Thus, the test statistics are based on the standardized residuals. By comparing Panel B with Panel A, it is clear that tests based on Hermite polynomials underreject a little bit the null assumption which again is not problematic. However the difference between knowing or not knowing the mean and the variance, decreases with the sample size and almost vanishes when the sample size is 1000. This confirms the robustness of these tests against parameter uncertainty. This is in contrast with the Kolmogorov-Smirnov test that almost never rejects the null. However, the Lilliefors modified test works well whatever the sample size. This is also the case for the JB test, since by construction, the empirical mean of the standardized residuals is zero. Indeed, the JB test coincides with the joint test based on H_3 and H_4, that is H_{3-4}.[17]

We also provide in Panel B results based on OPG regression as discussed in the previous section. For simplicity, we present three tests that correspond to the hypothesis $E[H_3(x)] = 0$, $E[H_4(x)] = 0$, and $E[H_3(x)] = E[H_4(x)] = 0$, which are respectively denoted OPG_3, OPG_4, and OPG_{3-4}. The main conclusion from the results in Table 1 is that tests based on OPG regression present some important distortions which remain with large sample sizes like 1000 observations. For instance, the probability of rejection of OPG_{3-4} are 31.5% and 10.9% with 100 and 1000 observations respectively. These distortions are in line with ones reported by Davidson and MacKinnon (1992) when they considered information matrix-based tests. Due to these distortions, we will not consider OPG tests in the rest of the paper.[18]

We now study the power of the considered tests against some interesting alternative assumptions for the cross-sectional case. In particular, we consider Student, chi-square and exponential alternatives. We start by simulating i.i.d. random variables from Student distributions with five different degrees of freedom: a) $T(60)$; b) $T(30)$; c) $T(20)$; d) $T(10)$;

[16] We do not consider H_1 and H_2 since these moments are used to estimate the mean and the variance.

[17] There is a small difference in Panel B between the JB and H_{3-4} tests since in JB test, the variance is estimated by $T^{-1}\sum_{t=1}^{T}(x_t - \bar{X})^2$ while in the Hermite case it is estimated by $(T-1)^{-1}\sum_{t=1}^{T}(x_t - \bar{X})^2$ where T is the sample size and $\bar{X}$ the empirical mean.

[18] We also do not consider tests based on double-length artificial linear regressions (Davidson and MacKinnon, 1984) which have better finite sample properties than tests based on OPG regression; see MacKinnon and Magee (1990).

and e) $T(6)$. Recall that for a random variable that follows a $T(\nu)$ distribution, the moments of order higher than ν-1 are not defined. Hence, the moments of H_i are not defined if $i > \nu - 1$. Moreover, the asymptotic distribution of the corresponding test statistics are not chi-square if $2i > \nu - 1$ since the variance of the Hermite polynomial H_i is not defined. The results are presented in Table 2. It is clear that the power of the tests is low when the degree of freedom ν is high. This is not surprising since a $T(\nu)$ distribution tends toward a normal one when $\nu \to +\infty$. However, when the degree of freedom decreases, the power of the tests increases and becomes very good when the degree of freedom is smaller than 10, which is the relevant case in the volatility literature (see the first example in the empirical section). A surprising result is that the fourth Hermite polynomial captures much more the non-normality than the higher polynomials. In contrast, tests based on odd polynomials do not work well. This is not surprising given that the mean in population of any odd Hermite polynomial is zero (when it is well defined) for any symmetric distribution and, hence, for a Student one.

In order to understand the behavior of the power of test statistics against Student distributions, we characterize in the appendix the behavior of those based on the third and fourth Hermite polynomials. In particular, we show that if one observes an i.i.d. sample $y_1, ..., y_T$, of a random variable Y that follows a $T(\nu)$ where $\nu > 8$, then

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_3(x_t)\right)^2 \overset{T\to+\infty}{\longrightarrow} A(\nu)\chi^2(1) \quad \text{and} \quad \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_4(x_t)\right)^2 \overset{T\to+\infty}{\longrightarrow} +\infty, \tag{5.1}$$

where $x_t = y_t\sqrt{(\nu-2)\nu^{-1}}$ and $A(\nu) = (\nu^2 - \nu + 10)/(\nu - 6)(\nu - 4)$. The first result in (5.1) implies that when one uses the third Hermite polynomial for testing normality while the random variable is a Student $T(\nu)$ at, say, 5% level, one accepts normality with a probability that equals $P(A(\nu)\chi^2(1) > 3.84)$ if one assumes that the limit distribution of $(T^{-1/2}\sum_{t=1}^{T} H_3(x_t))^2$ is $\chi^2(1)$.

In Table 3, we provided for all values of ν we considered in the Monte Carlo experiment, the value of $A(\nu)$ and the probability $P(A(\nu)\chi^2(1) > 3.84)$. These results are compatible with the Monte Carlo ones; in particular, the theoretical probabilities of rejection are very close to the Monte Carlo ones for the sample size $T = 1000$. Given that a test based on the third Hermite polynomial is not powerful, this is also the case for any joint test that uses this polynomial. In contrast, the second result in (5.1) explains why a test based on the fourth Hermite polynomial has a good asymptotic power against Student distribution.

Observe that we do not report results based on Hermite polynomials $H_i(\cdot)$ with $i > 7$ in Table 2 and the subsequent tables of the paper, because they do not provide gain in power with respect to tests based on Hermite polynomials with lower order like the test denoted by H_{3-6} in the tables; these simulations are however reported in Bontemps and Meddahi (2002).

Consider now the power of the tests against a $\chi^2(1)$ and an exponential distribution, exp(1). The results are reported in Table 4. They clearly imply that tests based on the third and fourth Hermite polynomials are very powerful whatever the sample size and that they are similar to the modified Kolmogorov-Smirnov test. However, tests based on individual higher order Hermite polynomials are less powerful for small sample sizes.

Similarly, we study the power of our tests against heteroskedastic and Gaussian errors. We follow Bera and Jarque (1981) by considering the model $x_t \sim \mathcal{N}(0, \sigma_t^2)$ with $\sigma_t^2 = 25 + \alpha z_t$ and $\sqrt{z_t} \sim \mathcal{N}(10, 25)$. We simulate two examples, namely $\alpha = .25$ and $\alpha = 1.25$, which respectively correspond to a weak and strong heterogeneity. The results reported in Table 5 indicate that our tests work well and these results are in line with those reported in Table 3 when we study the power of the tests against Student distributions.

The last i.i.d. example we consider is a GARCH(1,1) model:

$$x_t = \mu + \sqrt{h_t} u_t, \quad h_t = \omega + \alpha(\sqrt{h_{t-1}} u_{t-1})^2 + \beta h_{t-1},$$

where $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$, and $\beta = 0.8$. We take different distributions for u_t: $\mathcal{N}(0,1)$ to study the size of the tests, and standardized $T(20)$, $T(10)$, $T(6)$, $\chi^2(1)$, and $\exp(1)$ to study the power of the tests. The parameters μ, ω, α, and β are estimated with a Gaussian-QMLE method, which is consistent for all models given that we correctly specify the conditional mean and variance of x_t (Bollerslev and Wooldridge, 1992). The results reported in Table 6 confirm those reported in the other tables.

We did not adjust for small sample size distortions of the tests we considered. This obviously makes the power comparison less clear. The main reason for not adjusting the size is that our approach is semiparametric because our testing approach does not specify the complete model. And, we follow the same approach in the simulation even if we completely know the model. Of course, one can use the Bootstrap to do this correction, like the parametric bootstrap when one considers a linear regression model with fixed regressors. It is less clear if one can use it for the GARCH models and more importantly in the case where the variable of interest is serially correlated.

Dependent case. Consider now the dependent case, where the variable of interest is serially correlated. We consider several autoregressive normal processes of order one, AR(1), i.e., we assume that the conditional distribution of the variable of interest denoted by x_t given its past is $\mathcal{N}(\rho x_{t-1}, 1-\rho^2)$. Observe that the marginal distribution of x_t is $\mathcal{N}(0,1)$. We consider four values for ρ: a) $\rho = .1$; b) $\rho = .5$; c) $\rho = .7$ and d) $\rho = 0.9$. We did the same tests as for the independent case by assuming that we do not know the unconditional mean and variance of x_t.

We start by ignoring the dependence of the data, that is we assume that the sample size is i.i.d.; the results are reported in Table 7a. They clearly state that all the tests, including M-KS and JB ones, overreject the null when the sample size is higher than 250. This distortion is problematic. Therefore, we take into account the dependence of the data.

Next, we assume that we know the autoregressive structure. This is not always a realistic assumption. We do it however in order to get a benchmark. We consider two cases; in the first one, we assume that we know the autoregressive parameter while we estimate it by OLS in the second case. Given that the autoregressive feature of the data is known, we assume that the weighting matrix that appears in the test statistic is diagonal and that the diagonal coefficients are given by (3.11). The results are provided in Table 7b and Table 7c. These results are clearly good and similar to the ones provided in Panel B of Table 1 for the independent case. We observe again an underrejection of normality, in particular when the autoregressive coefficient increases.

We then test normality by ignoring the autoregressive feature of the data but by taking into account their dependence. Therefore, we do not assume that the weighting matrix Σ is diagonal. Instead, we estimate it by a HAC method. The HAC method is developed by using the quadratic kernel with an automatic lag selection procedure of Andrews (1991). The results are reported in Table 7d. From this table, it is clear that univariate tests work well. However, joint tests overreject the normality assumption, especially for small sample sizes and for tests that are based on three or more Hermite polynomials. The overrejection is relatively small for the test based on the third and fourth Hermite polynomials.

We now study the power of these tests against an autoregressive model of order one where the innovation is a Student one. Again, with the same autoregressive parameters as previously used, we consider almost the same degree of freedom as in the i.i.d. case, i.e., 30, 20, 10, and 5. We consider the T(5) example for comparison purposes with Bai and Ng (2002). Observe that the marginal distribution of the processes are (probably) not Student. However, their tails are clearly fatter than for a normal distribution. The results are reported in Tables 8a, 8b, 8c, and 8d. The main results of the tables can be summarized as follows. The tests based on univariate polynomials and different from the third one work well; however, their power decreases when both the degree of freedom of the Student distribution and the autocorrelation parameter are high. The univariate, bivariate and trivariate tests based on the third Hermite polynomial (denoted in the tables by H_3, H_{3-4} and H_{3-5}) are not powerful, especially when the autocorrelation is high. The main reason is symmetry. The second reason, given by Bai and Ng (2002), is that when the autocorrelation parameter is high, the Central Limit Theorem suggests that process of interest is close to a normal one. Note however that our results for H_{3-4} are different from ones of Bai and Ng (2002) when they test normality (for $\nu = 5$), which are more powerful then ours. The reason is not clear to us. One potential reason is the difference in the estimation method: in order to estimate the mean and variance parameters, we use the first two moments while Bai and Ng (2002) used the first four moments.

6 Empirical examples

This section provides two empirical examples: the first one concerns GARCH models while the second one deal with high frequency realized volatility.

6.1 First example: GARCH model

A very popular model in the volatility literature is GARCH(1,1) in Bollerslev (1986). More precisely, Bollerslev (1986) generalizes the ARCH models of Engle (1982) by assuming that

$$y_t = \sqrt{h_t} u_t \quad \text{with} \quad h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}, \text{ where } \omega \geq 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1, \tag{6.1}$$

and the process u_t is assumed to be i.i.d. and $\mathcal{N}(0,1)$. An important characteristic of GARCH models is that the kurtosis of y_t is higher than for a normal variable. It turns out that financial returns are also leptokurtic, and hence, GARCH models describe financial data well.[19]

However, some empirical studies found that the implied kurtosis of a GARCH(1,1) is lower than empirical ones. These studies lead Bollerslev (1987) to assume that the standardized process u_t may follow a Student distribution. Under this assumption, GARCH(1,1) fit financial returns very well. Indeed, by using a Bayesian likelihood method, Kim, Shephard and Chib (1998) proved that a Student GARCH(1,1) outperforms in terms of likelihood another very popular volatility model, namely the log-normal stochastic volatility model of Taylor (1986) popularized by Harvey, Ruiz and Shephard (1994) and Jacquier, Polson and Rossi (1994).

The first example we consider in our empirical study is testing normality of the standardized residual u_t. We consider the same data as Harvey, Ruiz and Shephard (1994) and Kim, Shephard and Chib (1998),[20] i.e., observations of weekday close exchange rates from 1/10/81 to 28/6/85. The exchange rates are the U.K. Pound, French Franc, Swiss Franc and Japanese Yen, all versus the U.S. Dollar. We estimate the model by a Gaussian QML method. The method is consistent as soon as the variance h_t is well specified (Bollerslev and Wooldridge, 1992). We get the fitted residuals $\hat{u}_t$ and test their normality. The results are provided in Table 9. It is clear that normality of the residuals is strongly rejected by all the tests, in particular those related to the tails (even polynomials). The difference between JB and H_{3-4} tests is relatively small; which is in line with the results of Fiorentini, Sentana and Calzolari (2003b) who found that the test based on the fourth moment for GARCH models is still valid even if the parameters are estimated.[21] Observe that the magnitude of normality rejection is in the following increasing order: FF-US$, UK-US$, SF-US$, and Yen-US$. Interestingly, this order is the same as one implied by the Student GARCH models estimated by Kim, Shephard and Chib (1998), since these authors reported in their Table 13 the following degree of freedom: 12.82, 9.71, 7.57 and 6.86.

[19]The second characteristic that GARCH models share with financial returns is the clustering effect. For a survey on GARCH models, see for instance Bollerslev, Engle and Nelson (1994).

[20]We are grateful to Neil Shephard for providing us with the data.

[21]Recall that exchange rates returns are symmetric.

6.2 Second example: realized volatility

Several recent studies highlight the advantage of using high-frequency data to measure volatility of financial returns. These include Andersen and Bollerslev (1998), ABDL (2001) and Barndorff-Nielsen and Shephard (2001). For a survey of this literature, Andersen, Bollerslev and Diebold (2002) and Barndorff-Nielsen and Shephard (2002) should be consulted. Typically, when one is interested in volatility over, say, a day, then these papers propose to study the estimation of this volatility by the sum of the intra-daily squared returns, such as returns over five or thirty minutes. This measure of volatility is called the realized volatility.

More precisely, consider S_t a continuous time process representing the price of an asset or the exchange rate between two currencies. Assume that it is characterized by the following stochastic differential equation:

$$d\log(S_t) = m_t dt + \sigma_t dW_t \quad \text{with} \quad d\sigma_t^2 = \tilde{m}_t dt + \tilde{\sigma}_t d\tilde{W}_t, \tag{6.2}$$

where W_t and $\tilde{W}_t$ are standard Brownian processes, potentially correlated. Assume that the time t is measured in units of one day. Consider a real h such that $1/h$ is a positive integer. Then, integrated and realized volatility denoted respectively by IV_t and $RV_t(h)$ are defined by

$$IV_t \equiv \int_{t-1}^{t} \sigma_u^2 du \quad \text{and} \quad RV_t(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2}, \tag{6.3}$$

where $r_{t-1+ih}^{(h)}$ is the return over the period $[t-1+(i-1)h; t-1+ih]$, given by $r_{t-1+ih}^{(h)} \equiv \log(S_{t-1+ih}) - \log(S_{t-1+(i-1)h})$. It turns out that when h goes to zero, realized volatility $RV_t(h)$ converges (in probability) to integrated volatility IV_t.

An assumption made in ABDL (2003) is conditional normality of the log of realized volatility. Hence, log-realized volatility are also unconditionally normal. This is our second example. We consider the same data as in ABDL (2003),[22] i.e., returns of three exchange rates, DM-US\$, Yen-US\$ and Yen-DM, from December 1, 1986 through June 30, 1999. The realized volatilities are based on observations at five and thirty minutes. Therefore, we have six series.

In Table 10, we provide the results of testing normality of log-realized volatility with unknown mean and variance. The weighting matrix is estimated by a HAC procedure of Andrews (1991). It is clear that unconditional normality of log-realized volatility is rejected, particularly for realized volatility based on five-minute returns.[23] Observe that the rejection is due to the asymmetry of the distribution. Thomakos and Wang (2003) also studied the log-normality of the realized volatility process by using among other tests, the KS and JB. They concluded that realized volatility is log-normal. However, these authors used a Monte Carlo study to correct the size of their tests, which is therefore a model based correction.

[22]We are grateful to Ramazan Gençay for providing us the OLSEN data and to Torben Andersen and Paul Labys for providing us their data.

[23]In their study, ABDL (2003) used thirty-minute realized volatilities.

We conclude this empirical section by one remark. In our tests, we assumed that the weighting matrix is well defined. This is not necessarily the case. In particular, ABDL (2003) reported results that clearly indicate a presence of long memory in log-realized volatility. In this case, the weighting matrix is not well defined and our test procedures are not valid. However, this is also the case for the procedures of ABDL (2003) which are based on the skewness, kurtosis, and nonparametric estimation of density of log-realized volatilities. Actually, some theoretical results are derived in Beran and Ghosh (1991). Following Taqqu (1979), they considered an expansion approach of the test statistic of interest onto the Hermite polynomials and show that its speed of convergence and the asymptotic distribution depends on the long memory parameter. For instance, the speed of convergence of the third Hermite polynomial may differ from one of the fourth Hermite polynomial. This is why Thomakos and Wang (2003) proved in their Monte Carlo study that the correction depends on the sample size and indeed on the long memory parameter. It is worth noting that Beran and Ghosh (1991) assumed that the process is Gaussian which means that both unconditional and conditional distributions are Gaussian, which is not our case. In addition, they did not take into account the parameter uncertainty problem. Testing normality under long memory in our framework is more difficult and is left for future research.

7 Conclusion

In this paper, we have considered testing marginal normal distributional assumptions for both cross-section and time series data. We used the GMM approach to test moment conditions given by Stein (1972) equations and the first class of moment conditions derived by Hansen and Scheinkman (1995) when the process of interest is a scalar diffusion. The main advantage of our approach is that tests based on Hermite polynomials are robust against parameter uncertainty. In addition, the GMM setting is well suited to take into account serial correlation by using a HAC procedure. We provided simulation results that clearly show the usefulness of our approach. We also applied our approach to test for normality in three volatility models.

Three main extensions have to be considered. The first one is to extend our approach to the multivariate case. The second is to consider other distributions, in particular Pearson ones. These two extensions are under consideration by using the Hansen and Scheinkman (1995) moment conditions which are valid in both multivariate normal and non normal cases. A third important extension will be testing normality under long memory.

Acknowledgments. The authors would like to thank Manuel Arellano, Bryan Campbell, Marine Carrasco, Xiaohong Chen, Russell Davidson, Frank X. Diebold, Jean-Marie Dufour, Jean-Pierre Florens, Ron Gallant, René Garcia, Ramazan Gençay, Silvia Gonçalves, Lynda Khalaf, James MacKinnon, Benoit Perron, Enrique Sentana, the editor, the associate editor and two anonymuous referees for helpful comments on an earlier draft and participants at CEMFI, Montréal, Duke, Princeton, Queen's University, Rochester, Toulouse, University of Iowa and the University of Pittsburgh seminars, the Canadian Economic Association Meeting, Montréal, June, 2001, the European Econometric Society Meeting, Lausanne, August, 2001, the American Statistical Association Meeting, New-York, August 2002, the NSF-NBER Time Series Conference, Philadelphia, September 2002, the CIRANO-CIREQ Conference on Financial Econometrics, Montréal, May 2003. The authors are responsible for any remaining errors. The second author acknowledges financial support from FCAR, IFM2, MITACS, NSERC, SSHRC, and the Econometrics Chair of Jean-Marie Dufour (Canada Research Chair Program).

Appendix

Proof of Proposition 4.1. Consider the first example and observe that

$$g(z_t,\theta)=\tilde{g}(u_t(\theta))=\tilde{g}\left(\frac{y_t-m(x_t,\theta)}{\sigma(x_t,\theta)}\right).$$

Then, we have:

$$\frac{\partial g}{\partial\theta^\top}(z_t,\theta)=\tilde{g}'\left(\frac{y_t-m(x_t,\theta)}{\sigma(x_t,\theta)}\right)\frac{\partial m}{\partial\theta^\top}(x_t,\theta)-\tilde{g}'\left(\frac{y_t-m(x_t,\theta)}{\sigma(x_t,\theta)}\right)\frac{(y_t-m(x_t,\theta))}{\sigma^2(x_t,\theta)}\frac{\partial\sigma}{\partial\theta^\top}(x_t,\theta)$$

$$=\tilde{g}'(u_t(\theta))\frac{\partial m}{\partial\theta^\top}(x_t,\theta)-\tilde{g}'(u_t(\theta))u_t(\theta)\frac{1}{\sigma(x_t,\theta)}\frac{\partial\sigma}{\partial\theta^\top}(x_t,\theta).$$

Hence, we have:

$$E\left[\frac{\partial g}{\partial\theta^\top}(z_t,\theta^0)\right]=E\left[\tilde{g}'(u_t(\theta^0))\right]E\left[\frac{\partial m}{\partial\theta^\top}(x_t,\theta^0)\right]-E\left[\tilde{g}'(u_t(\theta^0))u_t(\theta^0)\right]E\left[\frac{1}{\sigma(x_t,\theta)}\frac{\partial\sigma}{\partial\theta^\top}(x_t,\theta^0)\right]$$

since x_t is an exogenous variable. Hence, under (4.15), we have $E\left[\frac{\partial g}{\partial\theta}(z_t,\theta^0)\right]=0$, i.e., (4.7) holds. This achieves the proof for the first example. The same proof holds for the second example since u_t is independent of y_τ, $\tau\le t-1$.

Consider now the third example. We still have $\theta=(m,\sigma)$. Hence: $\frac{\partial g}{\partial\theta^\top}(z_t,\theta)=\tilde{g}'(u_t(\theta))\begin{bmatrix}1\\0\end{bmatrix}-\tilde{g}'(u_t(\theta))u_t(\theta)\frac{1}{\sigma}\begin{bmatrix}0\\1\end{bmatrix}$, and $E\left[\frac{\partial g}{\partial\theta}(z_t,\theta^0)\right]=0$ under (4.15). This achieves the proof for the third example.■

Proof of Proposition 4.2. Since $\tilde{g}(x)=f'-xf(x)$, we have $\tilde{g}'(x)=f''(x)-xf'(x)-f(x)$ and $x\tilde{g}'(x)=(xf'(x))'-x(xf'(x))+f'(x)-xf(x)-2f'(x)$. Applying the Stein equation (2.1) to $f'(x)$ and $xf'(x)$ proves the proposition.■

Proof of Proposition 4.3. The orthogonality property of the Hermite polynomials, i.e. (2.5), and the first result in (2.7) prove the proposition.■

Lemma. *Let $y_1, y_2, ..., y_T$, an i.i.d. sample of a random variable Y that follows a $T(\nu)$ where ν is assumed to be higher than eight ($\nu > 8$), and define the random variable X by $X = Y\sqrt{(\nu-2)\nu^{-1}}$. Then:*

$$\sqrt{T}\left(\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T} H_3(x_t) \\ \frac{1}{T}\sum_{t=1}^{T} H_4(x_t) \end{array} - \begin{array}{c} 0 \\ \sqrt{\frac{3}{2}\frac{1}{\nu-4}} \end{array}\right) \overset{T\to+\infty}{\longrightarrow} \left(\begin{array}{c} 0 \\ 0 \end{array}, \left(\begin{array}{cc} A(\nu) & 0 \\ 0 & B(\nu) \end{array}\right)\right), \tag{A.1}$$

where

$$A(\nu) = \frac{\nu^2-\nu+10}{(\nu-6)(\nu-4)} \quad \text{and} \quad B(\nu) = \frac{24\nu^3+1321\nu^2+708\nu-1572}{(\nu-8)(\nu-6)(\nu-4)}.$$

As a consequence:

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_3(x_t)\right)^2 \overset{T\to+\infty}{\longrightarrow} A(\nu)\chi^2(1), \quad \text{and} \tag{A.2}$$

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_4(x_t)\right)^2 \overset{T\to+\infty}{\longrightarrow} +\infty. \tag{A.3}$$

In addition, when T is large, we have the following approximation result:

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_4(x_t)\right)^2 \sim \frac{1}{T}B(\nu)\chi^2(1, T^2C(\nu)) \quad \text{where} \quad C(\nu) = \frac{3}{2}\frac{1}{(\nu-4)B(\nu)}. \tag{A.4}$$

Proof: Given that $Y \sim \mathcal{T}(\nu)$, EY^p is well defined when $p < \nu$. In this case, we have $EY^p = \nu^{p/2}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{\nu-p}{2}\right)/\left(\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)\right)$ if p is even and $EY^p = 0$ otherwise.
Thus, for $\nu > 8$, we have $E[H_3(X)] = Cov(H_3(X), H_4(X)) = 0$, $E[H_4(X)] = \sqrt{\frac{3}{2}\frac{1}{\nu-4}}$.
In addition, we have: $Var(H_3(X)) = EH_3^2(X) = 6^{-1}(EX^6 - 6EX^4 + 9EX^2) = A(\nu)$. The same computations lead to show that $Var(H_4(X)) = B(\nu)$. This achieves the proof of (A.1). The results (A.2), (A.3) and (A.4) are implied by (A.1).■

Table 1: Size of the tests.

Panel A: Mean and variance are known.

T	100	250	500	1000
H_1	5.1	5.1	5.2	5.0
H_2	5.1	4.8	5.0	4.9
H_3	5.5	5.4	5.6	5.2
H_4	4.8	4.7	4.9	5.0
H_5	3.6	4.3	4.8	5.0
H_6	2.0	2.7	3.2	3.6
H_7	1.3	1.6	2.0	2.4
H_8	1.4	1.2	1.3	1.4
H_9	1.9	1.7	1.4	1.3
H_{10}	1.2	1.6	1.8	1.6
H_{3-4}	5.8	5.6	5.7	5.3
H_{3-5}	6.0	6.2	6.3	6.0
H_{3-6}	5.4	5.8	6.1	6.1
H_{3-7}	5.0	5.2	5.5	5.6
H_{3-8}	5.0	5.0	5.1	5.1
H_{3-9}	4.8	4.9	5.0	4.9
H_{3-10}	4.5	4.6	4.8	4.7
KS	4.5	4.7	4.9	4.9
Naive-JB	15.0	16.3	17.4	17.4

Panel B: Mean and variance are estimated.

T	100	250	500	1000
H_3	4.3	4.8	4.9	5.0
H_4	3.1	3.8	4.3	4.6
H_5	2.4	3.6	4.3	4.8
H_6	1.2	2.1	2.9	3.4
H_7	0.7	1.2	1.7	2.2
H_8	0.8	0.9	1.0	1.3
H_9	1.2	1.4	1.1	1.1
H_{10}	0.6	1.3	1.7	1.5
H_{3-4}	4.1	4.5	4.6	4.8
H_{3-5}	4.2	5.1	5.3	5.4
H_{3-6}	3.6	4.7	5.3	5.5
H_{3-7}	3.3	4.2	4.7	5.1
H_{3-8}	3.2	4.0	4.3	4.6
H_{3-9}	3.1	3.9	4.2	4.4
H_{3-10}	2.8	3.7	4.1	4.3
KS	0.0	0.0	0.0	0.0
JB	4.3	4.7	4.7	4.8
M-KS	5.2	5.4	5.7	6.0
OPG_3	24.9	16.2	11.9	9.2
OPG_4	33.6	21.4	15.6	11.6
OPG_{3-4}	31.5	20.1	14.5	10.9

Note: The data are i.i.d. from a $\mathcal{N}(0,1)$ distribution. We tested the normality assumption. Either the mean and variance are not estimated (Panel A), or they are estimated (Panel B). The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. H_{i-j} corresponds to the joint test based on H_k, $i \leq k \leq j$. KS and JB are the Kolmogorov-Smirnov and Jarque-Bera tests; the Naive-JB statistic equals the JB statistic, but is not valid when the residuals are not standardized. M-KS is the modified Kolmogorov-Smirnov test. OPG_3, OPG_4 and OPG_{3-4} are the results of a normality test based on a OPG regression and testing respectively the third, fourth moments and both.

Table 2: Power of the tests against Student distributions.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}	KS	M-KS	JB
$\nu = 60$	100	5.7	4.9	3.7	2.1	6.3	6.4	5.7	0.0	0.0	6.6
	250	6.5	7.3	6.1	3.9	8.4	9.1	8.6	0.0	0.0	8.5
	500	6.8	9.5	7.8	5.5	10.0	11.3	10.8	0.0	0.0	10.2
	1000	6.8	13.0	9.5	7.4	12.6	14.0	13.9	0.0	0.0	12.6
$\nu = 30$	100	7.4	7.8	5.6	3.3	9.3	9.4	8.6	0.0	0.0	9.7
	250	8.6	13.3	9.8	6.6	14.0	14.9	14.1	0.0	0.0	14.3
	500	9.0	19.4	12.9	9.9	19.1	20.5	19.7	0.0	0.0	19.3
	1000	9.4	30.3	16.4	13.7	28.0	29.4	28.7	0.0	0.0	28.1
$\nu = 20$	100	9.6	11.2	7.7	4.9	12.9	12.8	11.9	0.0	0.0	13.4
	250	11.7	21.1	14.2	10.3	21.5	22.3	21.3	0.1	0.0	21.8
	500	12.3	34.2	19.8	15.8	32.6	33.5	32.3	0.1	0.0	32.8
	1000	13.0	54.0	25.3	22.7	50.0	50.0	48.8	0.1	0.0	50.1
$\nu = 10$	100	18.1	26.8	17.1	12.4	28.1	27.5	26.5	0.1	0.0	28.8
	250	23.1	52.3	31.6	26.2	51.2	50.6	49.9	0.2	0.0	51.6
	500	25.6	77.5	43.7	40.2	74.5	73.5	72.9	0.5	0.0	74.7
	1000	27.8	95.6	54.5	56.6	94.1	93.2	93.1	2.1	0.0	94.1
$\nu = 6$	100	31.4	50.6	33.2	27.1	51.2	50.1	49.5	1.1	0.0	51.9
	250	40.8	84.5	56.4	52.2	83.0	81.8	82.0	3.4	0.0	83.2
	500	46.5	98.0	71.4	72.8	97.4	97.0	97.1	11.5	0.0	97.4
	1000	51.8	100.0	82.2	89.4	100.0	100.0	100.0	41.7	0.0	100.0

Note: The data are i.i.d. from a $T(\nu)$ distribution. We test the normality assumption. Thus, we estimated the mean and variance. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. H_{i-j} is the joint test based on H_k, $i \le k \le j$. KS, M-KS and JB are the Kolmogorov-Smirnov, modified KS and Jarque-Bera tests.

Table 3: Probability of rejection for Student distributions.

ν	$A(\nu)$	$P(A(\nu)\chi^2(1) > 3.84)$
60	1.17	.071
30	1.41	.099
20	1.74	.137
10	4.16	.337
6	–	1

Table 4: Power of the tests against asymmetric distributions.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}	KS	M-KS	JB
$\chi^2(1)$	100	100.0	98.9	84.2	82.1	100.0	100.0	100.0	100.0	100.0	100.0
	250	100.0	100.0	97.8	98.6	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
exp(1)	100	100.0	89.6	70.0	61.6	100.0	100.0	100.0	91.9	100.0	100.0
	250	100.0	99.8	87.5	90.9	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	97.0	99.1	100.0	100.0	100.0	100.0	100.0	100.0
	1000	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: The data are i.i.d. from a $\chi^2(1)$ and exp(1) distributions. We tested the normality assumption. Thus, we estimated the mean and variance. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. H_{i-j} is the joint test based on H_k, $i \le k \le j$. KS, M-KS and JB are the Kolmogorov-Smirnov, modified KS and Jarque-Bera tests.

Table 5: Power of the tests against heteroskedastic and Gaussian errors

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}	KS	M-KS	JB
$\alpha = 0.25$	100	49.8	89.7	55.5	56.4	88.8	87.1	90.3	15.5	79.2	89.3
	250	56.5	99.9	77.6	79.1	99.8	99.8	99.9	69.2	99.3	99.8
	500	60.6	100.0	86.2	91.8	100.0	100.0	100.0	98.9	100.0	100.0
	1000	62.7	100.0	90.5	98.6	100.0	100.0	100.0	100.0	100.0	100.0
$\alpha = 1.25$	100	55.0	96.0	62.8	68.3	95.3	94.2	97.1	47.5	96.7	95.5
	250	61.7	100.0	82.2	84.8	100.0	100.0	100.0	98.3	100.0	100.0
	500	64.6	100.0	89.3	94.9	100.0	100.0	100.0	100.0	100.0	100.0
	1000	66.9	100.0	92.4	99.3	100.0	100.0	100.0	100.0	100.0	100.0

Note: We considered heteroscedasticity : $x_t \sim \mathcal{N}(0, \sigma_t^2)$ with $\sigma_t^2 = 25 + \alpha z_t$ and $\sqrt{z_t} \sim \mathcal{N}(10, 25)$. Two cases are treated : weak heterogeneity ($\alpha = 0.25$) and strong heterogeneity ($\alpha = 1.25$). These simulations are the same as in Bera and Jarque (1981). We tested the normality assumption. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j}, KS, M-KS and JB are defined in Table 3.

Table 6: Size and power of the tests with GARCH(1,1) errors.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}	KS	M-KS	JB
$u_t \sim \mathcal{N}(0,1)$	100	4.5	3.0	2.5	1.3	4.0	4.2	3.7	0.0	7.0	3.7
	250	4.5	3.5	3.3	2.1	4.0	4.6	4.3	0.0	7.7	4.3
	500	4.7	4.1	4.1	2.6	4.3	5.0	4.7	0.0	7.8	4.9
	1000	4.9	4.5	4.5	3.2	4.5	5.0	5.1	0.0	7.8	5.1
$u_t \sim \mathcal{T}(20)$	100	8.6	9.7	6.9	4.4	11.1	11.4	10.5	0.1	8.2	10.9
	250	10.4	17.6	12.2	8.5	18.4	19.2	18.1	0.1	11.1	19.0
	500	11.3	29.6	17.6	13.7	28.6	29.6	28.5	0.1	13.9	29.7
	1000	12.6	51.0	23.8	21.1	47.1	47.0	46.1	0.2	19.4	48.1
$u_t \sim \mathcal{T}(10)$	100	15.4	22.4	14.6	10.5	23.9	23.7	22.9	0.3	13.5	23.8
	250	20.0	45.8	27.3	22.1	44.8	44.4	43.5	0.5	22.5	45.9
	500	23.7	73.7	40.1	36.7	70.7	69.6	68.9	1.2	36.3	71.9
	1000	26.4	94.9	52.6	53.9	93.1	92.1	92.0	4.0	60.0	93.4
$u_t \sim \mathcal{T}(6)$	100	26.9	43.5	28.0	22.9	44.3	43.4	43.1	1.7	25.6	43.9
	250	39.4	82.1	54.2	50.8	80.7	79.5	79.9	4.6	49.0	80.9
	500	47.5	97.9	72.2	73.8	97.3	96.8	97.0	11.8	72.7	97.2
	1000	53.5	100.0	83.2	90.9	100.0	100.0	100.0	33.2	93.0	100.0
$u_t \sim \chi^2(1)$	100	100.0	98.4	80.7	77.6	100.0	100.0	100.0	99.9	100.0	100.0
	250	100.0	100.0	96.6	97.6	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$u_t \sim \exp(1)$	100	100.0	86.0	67.0	56.1	100.0	100.0	100.0	81.2	99.9	100.0
	250	100.0	99.7	84.3	87.7	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	95.7	98.7	100.0	100.0	100.0	100.0	100.0	100.0
	1000	100.0	100.0	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: The data follow a GARCH(1,1) process : $x_t = \mu + \sqrt{h_t} u_t$ with $h_t = \omega + \alpha(\sqrt{h_{t-1}} u_{t-1})^2 + \beta h_{t-1}$ and $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$, $\beta = 0.8$. u_t follows a Normal, Student, exponential or chi-square distribution up to some affine transformation which guarantees that $Eu_t = 0$ and $Vu_t = 1$. μ, ω, α and β are estimated with a QMLE method. We tested the normality assumption. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j}, KS, M-KS and JB are defined in Table 3.

Table 7a: Size of the tests under serial correlation which is ignored.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}	KS	M-KS	JB
$\rho = 0.1$	100	4.2	3.0	2.3	1.2	3.9	4.1	3.6	0.0	0.0	4.1
	250	4.7	3.9	3.5	2.1	4.6	5.2	4.7	0.0	0.0	4.7
	500	4.8	4.2	4.2	2.8	4.5	5.3	5.2	0.0	0.0	4.6
	1000	5.0	4.5	4.8	3.5	4.7	5.4	5.5	0.0	0.0	4.8
$\rho = 0.5$	100	5.9	2.7	2.0	0.9	4.7	4.5	3.9	0.0	0.0	4.9
	250	7.3	4.2	3.4	1.9	6.1	6.2	5.5	0.0	0.0	6.2
	500	7.7	5.0	4.3	2.6	6.9	7.0	6.5	0.0	0.0	7.0
	1000	8.1	5.7	4.9	3.3	7.5	7.5	7.1	0.0	0.0	7.5
$\rho = 0.7$	100	9.7	3.0	2.3	1.2	6.6	6.9	5.8	0.1	0.0	6.9
	250	13.5	5.9	3.7	2.0	10.7	10.4	9.1	0.1	0.0	10.9
	500	15.0	8.5	5.3	3.0	14.0	13.1	11.6	0.2	0.0	14.1
	1000	16.1	9.9	6.5	3.9	15.9	14.9	13.5	0.2	0.0	16.0
$\rho = 0.9$	100	19.2	7.9	11.3	6.7	15.2	19.8	21.2	3.1	0.0	16.2
	250	30.1	19.6	13.7	9.6	34.0	33.5	34.2	5.1	0.0	34.6
	500	35.9	26.2	16.0	11.1	44.2	42.1	41.7	6.1	0.0	44.6
	1000	39.1	30.3	17.9	12.7	49.8	48.2	46.9	6.3	0.0	50.0

Note: The data follow an AR(1) process: $x_t \mid x_{t-1} \sim \mathcal{N}(\rho x_{t-1}, 1-\rho^2)$. We tested the normality assumption. We did not take into account the serial correlation in the tests. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j}, KS, M-KS and JB are defined in Table 5.

Table 7b: Size of the tests under serial correlation.
The serial correlation is known and taken into account; ρ is known.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
$\rho = 0.1$	100	4.1	3.0	2.3	1.2	3.9	4.1	3.5
	250	4.6	3.9	3.5	2.1	4.5	5.2	4.7
	500	4.8	4.2	4.2	2.8	4.5	5.3	5.2
	1000	5.0	4.5	4.8	3.5	4.7	5.4	5.5
$\rho = 0.5$	100	3.5	2.2	1.7	0.9	3.1	3.2	2.8
	250	4.3	3.3	3.1	1.8	4.0	4.5	4.0
	500	4.6	3.6	3.8	2.6	4.2	4.8	4.8
	1000	4.8	4.3	4.5	3.4	4.7	5.3	5.4
$\rho = 0.7$	100	2.6	1.4	1.0	0.5	2.4	2.2	1.8
	250	3.9	2.6	2.1	1.3	3.5	3.6	3.3
	500	4.2	3.5	3.1	2.0	4.0	4.4	4.2
	1000	4.4	3.6	3.9	2.7	4.1	4.7	4.5
$\rho = 0.9$	100	0.9	0.4	0.3	0.1	0.8	0.7	0.6
	250	2.1	1.1	0.7	0.4	1.9	1.7	1.4
	500	3.2	1.9	1.4	0.8	3.0	2.9	2.5
	1000	4.0	2.6	2.2	1.3	3.7	3.8	3.4

Note: The data follow an AR(1) process: $x_t \mid x_{t-1} \sim \mathcal{N}(\rho x_{t-1}, 1-\rho^2)$. We tested the normality assumption. We took into account the serial correlation in the tests. We assume that we knew the AR(1) dynamics and that we knew ρ. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 7c: Size of the tests under serial correlation.
The serial correlation is known and taken into account; ρ is estimated.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
	100	4.1	3.0	2.3	1.2	3.9	4.1	3.6
$\rho = 0.1$	250	4.6	3.9	3.5	2.1	4.6	5.2	4.7
	500	4.8	4.2	4.2	2.8	4.5	5.3	5.2
	1000	5.0	4.5	4.8	3.5	4.7	5.4	5.5
	100	3.4	2.2	1.7	0.9	3.1	3.2	2.8
$\rho = 0.5$	250	4.3	3.3	3.1	1.8	4.0	4.4	4.1
	500	4.6	3.6	3.8	2.6	4.2	4.9	4.7
	1000	4.8	4.3	4.5	3.3	4.7	5.3	5.4
	100	2.4	1.3	1.0	0.5	2.0	2.0	1.7
$\rho = 0.7$	250	3.8	2.4	2.0	1.3	3.4	3.6	3.2
	500	4.1	3.4	3.0	2.0	3.8	4.3	4.1
	1000	4.4	3.6	3.9	2.7	4.0	4.7	4.6
	100	0.2	0.2	0.2	0.1	0.3	0.3	0.2
$\rho = 0.9$	250	1.6	0.9	0.6	0.3	1.4	1.4	1.1
	500	2.9	1.7	1.3	0.8	2.7	2.7	2.3
	1000	3.8	2.6	2.1	1.3	3.5	3.7	3.3

Note: The data follow an AR(1) process: $x_t \mid x_{t-1} \sim \mathcal{N}(\rho x_{t-1}, 1-\rho^2)$. We tested the normality assumption. We took into account the serial correlation in the tests. We assumed that we knew the AR(1) dynamics but not ρ which is estimated by OLS. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 7d: Size of the tests under serial correlation.
The serial correlation is unknown; Σ is estimated by a HAC procedure.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
	100	3.4	5.6	4.0	4.4	4.2	1.9	1.6
$\rho = 0.1$	250	4.2	7.6	4.1	4.2	8.2	6.4	3.6
	500	4.6	7.5	4.2	4.6	8.9	10.2	13.6
	1000	4.8	6.9	4.2	5.0	8.2	10.9	22.2
	100	3.3	3.9	4.2	4.5	2.8	1.4	1.3
$\rho = 0.5$	250	4.2	6.7	4.3	4.4	7.0	4.2	2.3
	500	4.6	7.3	4.3	4.6	8.8	8.0	6.5
	1000	5.0	7.1	4.3	4.7	8.8	10.2	16.0
	100	2.9	2.3	4.8	5.3	1.5	1.1	1.3
$\rho = 0.7$	250	4.0	5.1	4.8	4.8	4.6	2.4	1.8
	500	4.4	7.1	4.7	4.6	7.7	4.8	2.8
	1000	4.8	7.5	4.6	4.7	8.7	7.8	6.6
	100	1.6	0.7	4.5	4.9	0.5	0.5	0.4
$\rho = 0.9$	250	2.6	1.8	5.0	6.4	1.2	0.9	1.3
	500	3.5	3.8	5.2	5.8	2.8	1.5	1.6
	1000	4.0	6.1	5.0	5.2	5.7	2.8	1.9

Note: The data follow an AR(1) process: $x_t \mid x_{t-1} \sim \mathcal{N}(\rho x_{t-1}, 1-\rho^2)$. We tested the normality assumption. We took into account the serial correlation in the tests. We assumed that we do not knew the AR(1) dynamics. We used a HAC method. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 8a: Power of the tests under serial correlation against T(30) innovations.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
	100	5.0	82.2	84.1	82.8	5.8	0.5	76.7
$\rho = 0.1$	250	14.0	86.9	85.3	86.2	9.4	1.1	79.5
	500	20.6	91.8	85.9	89.6	12.9	3.6	79.3
	1000	24.6	94.9	78.5	82.3	7.6	8.9	56.3
	100	1.6	88.1	89.8	89.3	2.5	0.2	84.6
$\rho = 0.5$	250	5.1	90.1	90.7	91.0	4.3	0.3	87.3
	500	9.3	91.8	90.8	92.1	6.6	0.7	87.6
	1000	16.1	85.1	66.1	71.0	8.8	7.3	41.3
	100	0.4	91.2	92.7	92.7	1.2	0.1	86.0
$\rho = 0.7$	250	1.3	92.7	93.7	93.8	1.7	0.2	91.5
	500	3.0	93.5	94.2	94.4	2.9	0.2	92.4
	1000	7.0	63.4	50.8	52.4	8.5	5.9	24.2
	100	0.0	0.7	0.4	1.5	0.8	0.0	0.2
$\rho = 0.9$	250	0.1	96.7	97.3	97.4	0.3	0.0	92.4
	500	0.1	97.2	97.7	97.7	0.4	0.0	96.7
	1000	0.4	4.7	4.9	6.1	9.4	1.6	2.2

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ T(30). We tested the normality assumption. We took into account the serial correlation in the tests. We assumeed that we do not knew the AR(1) dynamics. We used a HAC method. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 8b: Power of the tests under serial correlation against T(20) innovations.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
	100	7.9	79.1	79.8	78.8	8.1	0.7	69.6
$\rho = 0.1$	250	19.6	87.6	81.8	84.8	11.6	2.6	72.5
	500	27.7	94.9	84.0	90.5	15.2	9.0	73.4
	1000	37.3	99.3	87.0	96.2	21.3	23.3	75.8
	100	2.6	85.1	87.0	86.4	3.7	0.3	79.4
$\rho = 0.5$	250	7.8	89.0	88.7	89.6	5.8	0.5	83.7
	500	13.2	92.4	88.8	91.6	8.4	1.9	83.5
	1000	18.3	96.8	89.7	94.7	11.5	6.6	83.7
	100	0.7	88.7	90.5	90.5	1.7	0.1	80.5
$\rho = 0.7$	250	2.3	91.1	92.3	92.4	2.7	0.2	89.0
	500	5.0	92.2	92.6	93.3	4.2	0.3	90.1
	1000	8.2	93.7	92.5	94.2	5.9	0.9	90.0
	100	0.1	1.4	0.7	2.5	1.1	0.0	0.3
$\rho = 0.9$	250	0.1	95.6	96.4	96.5	0.5	0.0	89.0
	500	0.2	96.6	97.2	97.2	0.5	0.1	95.9
	1000	0.5	96.8	97.3	97.4	0.8	0.1	96.4

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ T(20). We tested the normality assumption. We took into account the serial correlation in the tests. We assumeed that we do not knew the AR(1) dynamics. We used a HAC method. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 8c: Power of the tests under serial correlation against T(10) innovations.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
	100	15.3	77.1	70.7	71.3	11.7	2.6	54.0
$\rho = 0.1$	250	31.8	94.2	77.7	86.5	13.0	10.8	60.4
	500	47.4	99.5	85.0	95.7	16.8	28.3	68.8
	1000	62.1	100.0	92.3	99.6	23.8	47.4	80.8
	100	5.8	79.8	78.9	78.9	6.1	0.9	63.5
$\rho = 0.5$	250	15.0	90.2	81.8	87.3	8.2	3.0	69.4
	500	24.0	97.1	84.5	93.5	10.8	10.4	70.8
	1000	37.2	99.8	88.4	98.3	15.2	26.1	74.1
	100	1.6	81.1	83.8	83.3	3.0	0.4	61.2
$\rho = 0.7$	250	5.5	87.8	87.4	88.6	4.8	0.6	80.0
	500	10.2	92.1	88.1	91.9	6.5	1.8	80.8
	1000	16.0	96.9	89.1	95.4	8.7	6.1	80.7
	100	0.1	3.0	2.0	4.9	2.1	0.0	0.5
$\rho = 0.9$	250	0.2	92.8	94.1	94.3	0.8	0.1	73.1
	500	0.6	94.4	95.3	95.5	1.1	0.1	92.9
	1000	1.4	95.2	95.8	96.0	1.6	0.1	94.2

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ T(10). We tested the normality assumption. We took into account the serial correlation in the tests. We assumeed that we do not knew the AR(1) dynamics. We used a HAC method. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 8d: Power of the tests under serial correlation against T(5) innovations.

	T	H_3	H_4	H_5	H_6	H_{3-4}	H_{3-5}	H_{3-6}
	100	22.9	85.8	67.5	71.4	11.3	7.6	44.9
$\rho = 0.1$	250	45.7	99.5	86.7	94.0	9.3	22.3	70.6
	500	61.0	100.0	96.1	99.6	8.3	36.4	89.2
	1000	72.6	100.0	98.9	100.0	8.2	48.8	95.9
	100	11.1	77.8	65.0	68.4	8.8	4.2	41.2
$\rho = 0.5$	250	26.8	96.8	77.5	89.9	8.2	13.3	56.6
	500	45.9	99.9	88.7	98.4	8.8	28.8	73.8
	1000	64.0	100.0	95.8	100.0	10.2	44.5	88.3
	100	4.0	68.0	65.4	64.9	5.8	2.1	34.1
$\rho = 0.7$	250	11.3	88.2	74.7	84.5	6.8	4.5	54.3
	500	21.6	97.3	79.6	94.3	7.4	12.6	59.3
	1000	40.0	99.9	86.0	99.1	8.9	29.0	68.7
	100	0.2	4.2	4.0	6.4	5.0	0.3	1.1
$\rho = 0.9$	250	0.8	83.4	86.2	86.4	1.8	0.3	40.8
	500	2.0	90.7	90.5	91.7	2.4	0.4	81.1
	1000	4.3	93.7	91.2	94.3	3.1	0.7	84.6

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ T(5). We tested the normality assumption. We took into account the serial correlation in the tests. We assumeed that we do not knew the AR(1) dynamics. We used a HAC method. The results are based on 50000 replications. For each sample size, we provided the percentage of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 9: **Testing $\mathcal{N}(0,1)$ of fitted residuals for a GARCH(1,1) model.**

	UK-US$	FF-US$	SF-US$	Yen-US$
H_3	1.9 (0.17)	1.2 (0.26)	51.0 (0.00)	17.5 (0.00)
H_4	42.6 (0.00)	38.7 (0.00)	189 (0.00)	577 (0.00)
H_5	9.5 (0.00)	15.5 (0.00)	590 (0.00)	3713 (0.00)
H_6	46.8 (0.00)	126 (0.00)	2562 (0.00)	28553 (0.00)
H_7	26.9 (0.00)	8.7 (0.00)	10956 (0.00)	181020 (0.00)
H_8	45.9 (0.00)	8.5 (0.00)	35135 (0.00)	945122 (0.00)
H_9	17.2 (0.00)	2.4 (0.12)	88029 (0.00)	4186878 (0.00)
H_{10}	9.1 (0.00)	42.9 (0.00)	177511 (0.00)	15683206 (0.00)
H_{3-4}	44.4 (0.00)	39.9 (0.00)	240 (0.00)	594 (0.00)
H_{3-5}	54.1 (0.00)	55.4 (0.00)	830 (0.00)	4308 (0.00)
H_{3-6}	100 (0.00)	182 (0.00)	3393 (0.00)	32861 (0.00)
H_{3-7}	127 (0.00)	191 (0.00)	14349 (0.00)	213882 (0.00)
H_{3-8}	173 (0.00)	271 (0.00)	49485 (0.00)	1159004 (0.00)
H_{3-9}	191 (0.00)	273 (0.00)	137514 (0.00)	5345882 (0.00)
H_{3-10}	200 (0.00)	316 (0.00)	315026 (0.00)	21029088 (0.00)
KS	1.0	0.8	1.3	1.1
JB	44.6 (0.00)	40.1 (0.00)	240 (0.00)	597 (0.00)

Note: We tested the $\mathcal{N}(0,1)$ assumption of the standardized residuals. The volatility model is a GARCH(1,1) and is estimated by the Gaussian QML method. We reported the test statistics and their corresponding p-values in parentheses. The data are daily exchange rate returns used by Harvey, Ruiz and Shephard (1994) and Kim, Shephard and Chib (1998). H_{i-j} is the joint test based on H_k, $i \leq k \leq j$. KS and JB are the Kolmogorov-Smirnov and Jarque-Bera tests. The critical values of the KS and M-KS (see note of Table 1) are respectively: 1.63 and 1.031 (1%), 1.36 and .886 (5%), 1.22 and .805 (10%).

Table 10: Testing log-normality of realized volatility.

	DM-US$-5	DM-US$-30	Yen-US$-5	Yen-US$-30	Yen-DM-5	Yen-DM-30
H_3	10.6 (0.00)	8.3 (0.00)	10.2 (0.00)	7.0 (0.01)	8.7 (0.00)	3.6 (0.06)
H_4	7.5 (0.01)	5.9 (0.02)	3.6 (0.06)	3.4 (0.06)	5.2 (0.02)	3.4 (0.07)
H_5	0.1 (0.70)	2.9 (0.09)	1.1 (0.29)	2.2 (0.14)	0.7 (0.39)	0.02 (0.97)
H_6	0.0 (1.00)	0.04 (0.85)	0.7 (0.39)	1.0 (0.32)	1.8 (0.18)	1.0 (0.32)
H_7	8.9 (0.00)	0.8 (0.36)	1.0 (0.32)	1.6 (0.20)	0.7 (0.39)	2.7 (0.10)
H_8	0.7 (0.40)	0.5 (0.49)	0.9 (0.34)	1.3 (0.26)	1.0 (0.31)	1.7 (0.20)
H_9	1.5 (0.23)	0.6 (0.45)	0.07 (0.93)	4.0 (0.05)	8.1 (0.00)	0.9 (0.35)
H_{10}	0.6 (0.45)	1.1 (0.30)	2.2 (0.14)	4.0 (0.04)	1.1 (0.30)	0.4 (0.52)
H_{3-4}	16.9 (0.00)	10.5 (0.01)	16.8 (0.00)	9.6 (0.01)	11.0 (0.00)	7.0 (0.03)
H_{3-5}	17.6 (0.00)	10.5 (0.01)	17.4 (0.00)	9.8 (0.02)	17.3 (0.00)	7.7 (0.05)
H_{3-6}	17.7 (0.00)	15.2 (0.00)	19.2 (0.00)	16.0 (0.00)	17.3 (0.00)	7.8 (0.10)
H_{3-7}	24.0 (0.00)	15.5 (0.01)	26.2 (0.00)	24.5 (0.00)	18.6 (0.00)	10.1 (0.07)
H_{3-8}	25.6 (0.00)	15.5 (0.02)	28.2 (0.00)	25.9 (0.00)	18.9 (0.00)	10.2 (0.12)
H_{3-9}	25.9 (0.00)	17.5 (0.01)	28.3 (0.00)	26.4 (0.00)	25.6 (0.00)	12.1 (0.10)
H_{3-10}	27.0 (0.00)	19.9 (0.01)	28.5 (0.00)	27.9 (0.00)	26.2 (0.00)	12.5 (0.13)

Note: We tested the normality assumption of the log of realized volatility. The realized volatility is computed with five-minute and thirty-minute returns. We reported the test statistics and their corresponding p-values in parentheses. We used the same data as ABDL (2003). We used a HAC method of Andrews (1991) to estimate the weighting matrix. H_{i-j} is the joint test based on H_k, $i \leq k \leq j$.

References

Aït-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385-426.

Aït-Sahalia, Y., L.P. Hansen and J. Scheinkman (2001), "Operator Methods for Continuous-Time Markov Models," in Y. Aït-Sahalia and L.P. Hansen (Eds.), *Handbook of Financial Econometrics,* forthcoming.

Amemiya, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955-968.

Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

Andersen, T.G., T. Bollerslev and F.X. Diebold (2002), "Parametric and Nonparametric Measurements of Volatility," in Y. Aït-Sahalia and L.P. Hansen (Eds.), *Handbook of Financial Econometrics,* forthcoming.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.

Andersen, T.G., H.J. Chung and B.E. Sorensen (1999), "Efficient Method of Moment Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61-87.

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 60, 953-966.

Bai, J. (2002), "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, forthcoming.

Bai, J. and S. Ng (2002), "Tests for Skewness, Kurtosis, and Normality for Time Series Data," unpublished manuscript, Johns Hopkins University.

Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian OU based Models and some of their uses in Financial Economics," with discussion, *Journal of the Royal Statistical Society, B*, 63, 167-241.

Barndorff-Nielsen, O.E. and N. Shephard (2002), " Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 5, 457-477.

Bera, A.K. and C.M. Jarque (1981), "Efficient Tests for Normality, Homoscedasticity and serial Independence of Regression Residuals: Monte Carlo Evidence," *Economics Letters*, 7, 313-318.

Bera, A. and S. John (1983), "Tests for Multivariate Normality with Pearson Alternatives," *Commun. Statist. Theor. Meth.,* 12, 103-107.

Beran, J. and S. Ghosh (1991), "Slowly decaying Correlation, Testing Normality, Nuisance Parameters," *Journal of the American Statistical Association*, 86, 785-791.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

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

Bollerslev, T., R. Engle and D.B. Nelson (1994), "ARCH Models," *Handbook of Econometrics*, Vol IV.

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

Bontemps, C. and N. Meddahi (2002), "Testing Normality: A GMM Approach," CIRANO working paper, 2002s-63.

Bowman, K.O. and L.R. Shenton (1975), "Omnibus Test Contours for Departures from Normality based on $\sqrt{b_1}$ and b_2," *Biometrika*, 62, 243-250.

Chen, X., L.P. Hansen and J. Scheinkman (2000), "Principal Components and the Long Run," unpublished manuscript, University of Chicago.

Conley, T., L.P. Hansen, E. Luttmer and J. Scheinkman (1997), "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies,* 10, 525-577.

Corradi, V. and N.R. Swanson (2002), "A Bootstrap Specification Test for Diffusion Processes," working paper, Rutgers University.

Davidson, R. and J.G. MacKinnon (1984), "Model Specification Tests Based on Artificial Linear Regressions," *International Economic Review*, 25, 485-502.

Davidson, R. and J.G. MacKinnon (1992), "A New Form of the Information Matrix Test," *Econometrica*, 60, 145-157.

Davidson, R. and J.G. MacKinnon (1993), "Estimation and Inference in Econometrics," Oxford University Press.

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

Duan, J.C. (2003), "A Specification Test for Time Series Models by a Normality Transformation," working paper, University of Toronto.

Dufour, J-M, A. Farhat, L. Gardiol and L. Khalaf (1998), "Simulation-Based Finite Sample Normality Tests in Linear Regressions," *Econometrics Journal*, 1, C154-C173.

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

Epps, T.W and L.B. Pulley (1983), "A Test for Normality Based on the Empirical Characteristic Function," *Biometrika*, 70, 723-726.

Epps, T.W, K. Singleton and L.B. Pulley (1982), "A Test of Separate Families of Distributions Based on the Empirical Moment Generating Function," *Biometrika*, 69, 391-399.

Fiorentini, G., E. Sentana and G. Calzolari (2003a), "The Score of Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations, an LM Test for Multivariate Normality," *Journal of Business & Economic Statisitcs*, forthcoming.

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

Gallant, A.R. (1980), "Explicit Estimators of Parametric Functions in Nonlinear Regressions," *Journal of the American Statistical Association*, 75, 182-193.

Gallant, A.R. (1987), "Nonlinear Statistical Models," John Wiley and Sons.

Gallant, A.R. and H. White (1988), "A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models," Basil Blackwell.

Gallant, A.R. and G. Tauchen (1999), "The Relative Efficiency of Methods of Moments Estimators," *Journal of Econometrics*, 92, 149-172.

Gouriéroux, C. and A. Monfort (1996), *Simulation-Based Econometric Methods*, CORE Lectures, Oxford.

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

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Hansen, L.P. and J. Scheinkman (1995), "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica*, 63, 767-804.

Harvey, A.C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247-264.

Hong, Y. and H. Li (2002), "Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates," working paper, Cornell University.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371-389.

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

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

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

Kim, S., N. Shephard and S. Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

Kolmogorov, A.N. (1933), "Sulla Determinazione Empiricadi una Legge di Distribuzione," *Giorna. Ist. Attuari.*, 4, 83-91.

Koutrouvelis, I.A. (1980), "A Goodness-of-Fit Test of Simple Hypotheses Based on the Empirical Characteristic Function," *Biometrika*, 67, 238-240.

Koutrouvelis, I.A. and J. Kellermeier (1981), "A Goodness-of-Fit Test Based on the Empirical Characteristic Function when Parameters must be Estimated," *J. R. Statist. Soc. B*, 43, 173-176.

Lilliefors, H.W. (1967), "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance," *Journal of the American Statistical Association*, 62, 399-402.

MacKinnon, J.G. and L. Magee (1990), "Transforming the Dependent Variable in Regression Variables," *International Economic Review*, 31, 315-339.

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

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.

Newey W.K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.

Newey W.K. and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

Richardson, M. and T. Smith (1993), "A test for Multivariate Normality in Stock Returns," *Journal of Business*, 66, 295-321.

Schoutens, W. (2000), *Stochastic Processes and Orthogonal Polynomials*, Lecture Notes in Statistics 146, Springer-Verlag.

Smirnov, N.V. (1939), "Sur les Écarts de la Courbe de Distribution Empirique," *French/Russian Su Matematiceskii Sbornik N. S.*, 6, 3-26.

Stein, C. (1972), "A Bound for the Error in the Normal Approximation to the Distribution of a Sum of Dependant Random Variables," *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, 2, 583-602.

Taqqu, M. (1979), "Convergence of Iterated Processes of Arbitrary Hermite Rank," *Zeitschrift für Wahrscheinlichkeitstheorie under verwandte Gebiete*, 50, 53-83.

Tauchen, G.E. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415-443.

Taylor, S.J. (1986), *Modeling Financial Time Series*, John Wiley.

Thomakos, D.D. and T. Wang (2003), "Realized Volatility in the Future Markets," *Journal of Empirical Finance*, 10, 321-353.

Thompson, S. (2002), "Specification Tests for Continuous Time Models," working paper, Harvard University.

van der Klaauw, B. and R.H. Koning (2001), "Testing the Normality Assumption in the Sample Selection Model with an Application to Travel Demand," *Journal of Business & Economic Statisitcs*, forthcoming.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.

White, H. and G.M. MacDonald (1980), "Some Large-Sample Tests for Nonnormality in the Linear Regression Model," *Journal of the American Statistical Association*, 75, 16-28.

Wooldridge, J. (1990), "A Unified Approach to Robust, Regression-based Specification Tests," *Econometric Theory*, 6, 17-43.