Cahier 2002-14

Testing Normality : A GMM Approach

BONTEMPS, Christian
MEDDAHI, Nour

Université
de Montréal

Testing Normality: A GMM Approach[*]

Christian Bontemps[†] and Nour Meddahi[‡]

First version: March 2001

This version: June 2002

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 adopted 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, three applications to GARCH and realized volatility models are presented.

Key words: Normality, Stein-Hansen-Scheinkman equation, GMM, Hermite polynomials, parameter uncertainty, HAC, OPG regression.

[†]LEEA-CENA, 7 avenue Edouard Belin, 31055 Toulouse Cedex, FRANCE. E-mail: bontemps@cena.fr;

[‡]Département de sciences économiques, CRDE, CIRANO, Université de Montréal, and CEPR. Address: C.P. 6128, succursale Centre-ville, Montréal (Québec), H3C 3J7, Canada. E-mail: nour.meddahi@umontreal.ca.

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 non linear variables like volatility in the EGARCH model of Nelson (1991), or the high frequency realized volatility model of Andersen, Bollerslev, Diebold and Labys (2002, ABDL). This is also important when one evaluates density forecasts as in Diebold, Gunther and Tay (1998). In risk management literature (see Christoffersen, 1998, and Berkowitz, 2001), the most popular measure of risk, that is the Value at Risk (VaR), is based on quantiles and, hence, on distributional assumptions. 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 tests procedure 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 that 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 a very important literature on testing normality. This includes tests based on the cumulative distribution (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; Bera and Jarque, 1981), the Hermite polynomials (Kiefer and Salmon, 1983; 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 showed 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 polynomials. Interestingly, the Stein equation coincides with the first class of moment

---

[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, 2002).

conditions given by Hansen and Scheinkman (1995) for continuous time processes when one considers a normal process, that is the Ornstein-Uhlenbeck process.[2]

We use 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, it is easy in the GMM setting 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 for instance 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) showed that various tests are robust against parameter uncertainty, in particular tests based on moments that used standardized residuals. Finally, note that 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.

It turns out that the GMM setting is well suited for incorporating parameter uncertainty in testing procedures by using Newey (1985) and Tauchen (1985). In this paper, we show that some testing functions are robust to the parameter uncertainty problem, that is the distribution of the test statistic when one uses the true unknown parameter coincides with the feasible test statistic when one uses an estimator of the 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 non linear homoskedastic regression estimated by the maximum likelihood method. In contrast, our result holds whatever the model and the 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 that when the variable of interest is serially correlated, the GMM setting is also well suited to take into account this dependence by using 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 (2001) for testing normality, Ait-Sahalia (1996) and Conley, Hansen, Luttmer and Scheinkman (1997) for testing marginal distributions of non linear scalar diffusion processes.

---

[2]Hansen and Scheinkman (1995) gave two class of moment conditions related respectively to the 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.

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 three examples from the volatility literature while the last section concludes. 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 also 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. A simple one is the classical recursive moment formula. More precisely, if one considers the case of monomial functions, that is $f_i(X) = X^i$, $i = 0, 1, ...$, the Stein equation (2.1) implies that

$$E[X^{i+1}] = iE[X^{i-1}].$$

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 in the Stein equation is that by construction, the expectation of the considered function is zero.

---

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

Therefore, one has not to compute theoretical formula as one would do when using 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 the Stein equation is useful because it is not necessary to compute the theoretical formula. Instead, one must compare the empirical counterparts of $E[g(X)]$ and $E[XG(X)]$ where G is any primitive function of g. Of course, this means that one has to compute the primitive of g. An alternative 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

$$E[g(X)] = E[Xh(X)] = E[h'(X)].$$

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 orthonormalized Hermite polynomials H_i associated to the $\mathcal{N}(0,1)$ distribution are 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 following 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 $i > 0$.

4

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:

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

$$H_i^{''}(x) - xH_i^{'}(x) + iH_i(x) = 0. \tag{2.8}$$

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

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

By using (2.8), this equation implies (2.6). As a consequence, the Stein equation (2.1) implies (2.6). It turns out that the converse holds also:

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 $\mathcal{N}(0,1)$.
As a consequence, a random variable X is $\mathcal{N}(0,1)$ if and only if (2.6) holds.

This 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

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{2.9}$$

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

$$E[\mathcal{A}g(X_t)] = 0, \tag{2.10}$$

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 to the diffusion (2.9), that is:

$$\mathcal{A}g(x) = \mu(x)g'(x) + \frac{\sigma^2(x)}{2}g''(x). \tag{2.11}$$

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

$$dX_t = -kX_tdt + \sqrt{2k}dW_t, \quad k > 0, \quad X_0 \sim \mathcal{N}(0,1). \tag{2.12}$$

For this process, Hansen and Scheinkman (1995) moment condition (2.10) becomes:

$$E[-kX_t g'(X_t) + kg''(X_t)] = 0. \tag{2.13}$$

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 (2.10) 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,[4]

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

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.[5] 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 $\{Y_\tau, \tau \in I\!\!R^+\}$ 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 the 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. In this section, 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 (see 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,

---

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

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

..., f_p, be p differentiable functions such that f_i' is integrable. For a real x, define the vector $g(x)$ $\in \mathbb{R}^p$, which components are $(f_i'(x) - x f_i(x))$ for $i = 1, ..., p$. Thus, by the Stein equation (2.1), we have $E[g(X)] = 0$. Assume that any component of the 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 definite positive, 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 using the Stein equation (2.1), one can have a simple form for $E[g(X) g(X)^\top]$. More precisely, it is easy to 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 one, 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 explicitly Σ is not possible (or difficult), then one can use any consistent estimator of Σ and denoted by $\hat{\Sigma}_T$. Examples of consistent estimators of Σ_T are

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

7

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 showed 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. Besides, 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 one of Kiefer and Salmon (1982) who showed (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 (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 was already used by Smith and Richardson (1991) and was more recently and independently of our work highlighted by Bai and Ng (2001).

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, we think that $E[H_i(x_t)H_j(x_{t-h})]$ is in general nonzero for $h \neq 0$. Thus, the matrix Σ will be in general non diagonal.[6]

In contrast, this may be the case 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

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

In this case, any Hermite polynomial $H_i(x_t)$ is an AR(1) process which 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. 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.

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

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

---

[6]In their paper, Bai and Ng (2001) write in page 5 that the tests based on the excess skwness and excess kurtosis are asymptotically independent under normality even for time series data. They do not provide a proof however.

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

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

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

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

Note that this test statistic is different from one of Jarque and Bera (1980) 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.

3.5   OPG regression

Davidson and MacKinnon (1993) considered testing normality through skewness and excess kurtosis. In particular, they reformulated Jarque and Bera (1980) test in terms of outer-product-of-the-gradient (OPG) regression.

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

$$1 = bx_t + s(x_t^2 - \sigma^2) + \text{residual}. \tag{3.17}$$

Assume now that one is interested in testing skewness. Then, Davidson and MacKinnon (1993) propose to add in (3.17) the regressor x_t^3, i.e.,

$$1 = bx_t + s(x_t^2 - \sigma^2) + ax_t^3 + \text{residual}, \tag{3.18}$$

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.

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

which is, under normality and when T is large, very close to

$$\frac{1}{T}\sum_{t=1}^{T} x_t^3 - 3\sigma^2 \frac{1}{T}\sum_{t=1}^{T} x_t.$$

Davidson and MacKinnon (1993) showed that the variance of $x_t^3 - 3\sigma^2 x_t$ is $6\sigma^6$. Thus, the t statistic is very close to

$$\frac{\frac{1}{T}\sum_{t=1}^{T} x_t^3 - 3\sigma^2 \frac{1}{T}\sum_{t=1}^{T} x_t}{\sqrt{6\sigma^6}}. \tag{3.19}$$

Similarly, Davidson and MacKinnon (1993) showed that one can test excess kurtosis by doing the following OPG regression

$$1 = bx_t + s(x_t^2 - \sigma^2) + ax_t^3 + b(x_t^4 - 3\sigma^4) + \text{residual}, \tag{3.20}$$

and that the t-statistic related to the coefficient b is very close to

$$\frac{\frac{1}{T}\sum_{t=1}^{T} x_t^4 - 6\sigma^2 \frac{1}{T}\sum_{t=1}^{T} x_t^2 + 3\sigma^4}{\sqrt{24\sigma^8}}. \tag{3.21}$$

Observe that the test statistics (3.19) and (3.21) may be written as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\sqrt{6}}[(x_t/\sigma)^3 - 3(x_t/\sigma)] = \frac{1}{T}\sum_{t=1}^{T} H_3(x_t/\sigma) \tag{3.22}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\sqrt{24}}[(x_t/\sigma)^4 - 6(x_t/\sigma)^2 + 3] = \frac{1}{T}\sum_{t=1}^{T} H_4(x_t/\sigma). \tag{3.23}$$

Given that the variance of $H_3(x_t/\sigma)$ is one, the test statistic (3.19) is also the t-statistic of the coefficient $\tilde{a}$ in the two OPG regressions

$$1 = bx_t + s(x_t^2 - \sigma^2) + \tilde{a}H_3(x_t/\sigma) + \text{residual}, \text{and} \tag{3.24}$$

$$1 = \tilde{b}H_1(x_t/\sigma) + \tilde{s}H_2(x_t/\sigma) + \tilde{a}H_3(x_t/\sigma) + \text{residual}. \tag{3.25}$$

In addition, given that the variance of $H_4(x_t/\sigma)$ is one, the test statistic (3.21) is also the t-statistic of the coefficient $\tilde{b}$ in the two OPG regressions

$$1 = bx_t + s(x_t^2 - \sigma^2) + \tilde{b}H_4(x_t/\sigma) + \text{residual}, \text{and} \tag{3.26}$$

$$1 = \tilde{b}H_1(x_t/\sigma) + \tilde{s}H_2(x_t/\sigma) + \tilde{b}H_4(x_t/\sigma) + \text{residual}. \tag{3.27}$$

11

Finally, because $H_3(x_t/\sigma)$ and $H_4(x_t/\sigma)$ are orthogonal, the test statistics (3.19) and (3.21) are respectively the t-statistic of the coefficients $\tilde{a}$ and $\tilde{b}$ in the two OPG regressions

$$1 = bx_t + s(x_t^2 - \sigma^2) + \tilde{a}H_3(x_t/\sigma) + \tilde{b}H_4(x_t/\sigma) + \text{residual, and} \tag{3.28}$$

$$1 = \tilde{b}H_1(x_t/\sigma) + \tilde{s}H_2(x_t/\sigma) + \tilde{a}H_3(x_t/\sigma) + \tilde{b}H_4(x_t/\sigma) + \text{residual.} \tag{3.29}$$

In other words, given the orthonormality of Hermite polynomials, 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. It is also the case when one tests excess kurtosis by using the fourth Hermite polynomial and, more generally, for higher order Hermite polynomials.

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 non linear, homoskedastic or heteroskedastic, where the normality assumption is in general made on the (standardized) residuals. This is also the case for non linear 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 some 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 that one has to take into account the uncertainty of $\hat{\theta}_T$ in the testing procedure. This is known as the parameter uncertainty problem.

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

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

12

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

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

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:

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

Define the matrix P_g by

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

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.[7] As a consequence, the parameter uncertainty generally changes the asymptotic distribution of test statistic when one uses an estimator instead of the unknown parameter θ^0.

In general, the matrices that appear in the asymptotic distributions (4.2) and (4.3) are easily estimated. However, it is difficult is 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[8] as in the case of stochastic volatility and latent factor models.

An alternative method that adopted in this paper is to consider moment conditions such that the matrix P_g is 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.

---

[7] We assume that the asymptotic distribution of the right hand side of (4.6) is normal.
[8] See Gouriéroux and Monfort (1996) for a review.

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

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

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.

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

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

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

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

i.e., a non linear 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.,[9]

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

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

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 which 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

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

Observe that in all 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 Example 1, 2 or 3.

---

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

Proposition 4.1 *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) and $\hat{u}_t$ 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*

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

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

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

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 (2001). 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. Jarque and Bera (1980) did this correction by using a Lagrange multiplier test.

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*

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

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:

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). 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 non linear regression with homoskedastic and i.i.d. errors, that is under (4.8) and (4.10). Moreover, they showed 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. In addition, the results of Davidson and MacKinnon (1993) in their OPG regression where the consider parameter uncertainty are also a special case of the previous proposition.

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

$$\hat{u}_t = \frac{y_t - x_t^\top \hat{\beta}}{\hat{\alpha}} \tag{4.19}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the MLE of α and β. Then, Jarque and Bera (1980) showed that

$$\frac{1}{6} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{u}_t^3 \right)^2 + \frac{1}{24} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\hat{u}_t^4 - 3] \right)^2 \overset{n \to +\infty}{\longrightarrow} \chi^2(2). \tag{4.20}$$

In Jarque and Bera (1980), the constant is in the regressors. Hence, we have

$$\sum_{t=1}^{T} \hat{u}_t = 0. \tag{4.21}$$

In addition, $\hat{\alpha}^2$ is given by $\hat{\alpha}^2 = T^{-1} \sum_{t=1}^{T} (y_t - x_t^\top \hat{\beta})^2$. Therefore, we have

$$\frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 = 1. \tag{4.22}$$

Hence, when one combines the left-hand side of (4.20) with (4.21) and (4.22), one gets

$$\frac{1}{6} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\hat{u}_t^3 - \hat{u}_t] \right)^2 + \frac{1}{24} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\hat{u}_t^4 - 6\hat{u}_t^2 + 3] \right)^2 \text{ , i.e.,}$$

16

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} H_3(\hat{u}_t) \right)^2 + \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} H_4(\hat{u}_t) \right)^2 .$$

In other words, 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. 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.[10] 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.

5   A Monte Carlo study

In this section, we provide some Monte Carlo experiments to study the finite sample properties of the tests we proposed. We also compare our tests with the most popular ones, that is the Kolmogorov-Smirnov and the Jarque-Bera tests. These 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 denoted by M-KS in the tables.

All the test functions we consider are based on Hermite polynomials given 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 denominated by H_{3-i} for $i = 4, ..., 10$ in the tables.

In all the simulations 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% are similar to 5% and are omitted to save space.[11]

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, this is a good benchmark for the realistic cases where the parameters are unknown and have to be estimated. We reported the results in Table 1. Consider the tests based on individual

---

[10]These two methods are sufficient ones. In particular, Fiorentini, Sentana and Calzolari (2002) show that for some heteroskedastic models estimated by the ML method, test based on the fourth moment (as in Jarque and Bera, 1980) applied to the fitted residuals is still valid.

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

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 have also very good finite sample properties for the four sample sizes. Indeed, we do not observe the under rejection 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 some times better for the sample size 100 when one considers a test based on H_4. However the JB test does not work well and overreject the null. The main reason is that when the empirical mean of the sample is not zero, the asymptotic distribution of the Jarque and Bera (1980) test is not a $\chi^2(2)$.

In Table 2, we present the results of the same tests[12] 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 Table 2 with Table 1, 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 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}.[13]

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 six different degrees of freedom: a) $T(60)$; b) $T(30)$; c) $T(20)$; d) $T(10)$; e) $T(6)$; f) $T(3)$. 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 3. 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

---

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

[13]There is a small difference in Table 2 between 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.

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 following proposition their behavior for the third and fourth Hermite polynomials.

Proposition 5.1 *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:*

$$\sqrt{T}\left(\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T} H_3(x_t) \\ \frac{1}{T}\sum_{t=1}^{T} H_4(x_t) \end{array} - \begin{array}{c} 0 \\ \sqrt{\frac{3}{2}\frac{1}{\nu-4}} \end{array} \right) \xrightarrow{T\to+\infty} \left(\begin{array}{c} 0 \\ 0 \end{array}, \left(\begin{array}{cc} A(\nu) & 0 \\ 0 & B(\nu) \end{array} \right) \right), \tag{5.1}$$

where

$$A(\nu) = \frac{\nu^2 - \nu + 10}{(\nu - 6)(\nu - 4)} \quad \text{and} \quad B(\nu) = \frac{24\nu^3 + 1321\nu^2 + 708\nu - 1572}{(\nu - 8)(\nu - 6)(\nu - 4)}. \tag{5.2}$$

As a consequence:

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_3(x_t) \right)^2 \xrightarrow{T\to+\infty} A(\nu)\chi^2(1), \quad \text{and} \tag{5.3}$$

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_4(x_t) \right)^2 \xrightarrow{T\to+\infty} +\infty. \tag{5.4}$$

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

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_4(x_t) \right)^2 \sim \frac{1}{T}B(\nu)\chi^2(1, T^2 C(\nu)) \quad \text{where} \quad C(\nu) = \frac{3}{2}\frac{1}{(\nu-4)B(\nu)}. \tag{5.5}$$

Equation (5.3) 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

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_3(x_t) \right)^2 \xrightarrow{T\to+\infty} \chi^2(1).$$

In Table 4, 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 test that uses this polynomial, as the joint test based on the third and fourth polynomial. In contrast, equation (5.4) explains why a test based on the fourth Hermite polynomial has a good asymptotic power against Student distribution.

Consider now the power of the tests against a $\chi^2(1)$ and an exponential distribution, exp(1). The results are reported in Table 5. 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.

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 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 6a. They clearly mean 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 6b and Table 6c. These results are clearly good and similar to the ones provided in Table 2 for the independent case. We observe again an under rejection 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 6d. 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 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, we consider the same degree of freedom as in the i.i.d. case, i.e., 60, 30, 20, 10, 6 and 3. 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 7a, 7b, 7c, 7d, 7e and 7f. We also report in Table 7g the results of the example where the innovation is $T(5)$. We consider this example for comparison purposes with Bai and Ng (2001). The main results of the tables can be summarized as follows. The test 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 tests univariate, bivariate and trivariate tests based on the third Hermite polynomials (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 (2001), 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 (2001) when they test normality (for $\nu = 5$) since that they find that their test are powerful. The main reason is that we do not assume diagonality of the weighting matrix when we do joint tests while Bai and Ng (2001) do. As we pointed out in the third section, there is no reason to assume in general that the tests are asymptotically independent. However, it is the case for an AR(1) process, i.e., the setup we adopted as well as Bai and Ng (2001) in our simulation. Given that we are interested on the power of the tests, we did not take into account this asymptotic independence in our simulations.

6   Empirical examples

In this section we provide three empirical examples. The first one concerns GARCH models while the second and third ones deal with high frequency realized volatility.

6.1   First example: GARCH model

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

$$y_t = \sqrt{h_t}u_t \quad \text{with} \quad h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}, \text{ where } \omega \geq 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1, \quad (6.6)$$

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 well financial data.[14]

---

[14]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).

However, some empirical studies showed 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 very well financial returns. Indeed, by using Bayesian likelihood method, Kim, Shephard and Chib (1998) showed 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).

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),[15] 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 against 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 8. 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; this fact is in line with the results of Fiorentini, Sentana and Calzolari (2002) who showed that the test based on the fourth moment for GARCH models is still valid even if the parameters are estimated.[16] 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 than 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.

6.2   Second and third examples: 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 (2001) 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

---

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

[16]Recall that exchange rates returns are symmetric.

stochastic differential equation:

$$d \log(S_t) = m_t dt + \sigma_t dW_t \quad \text{with} \quad d\sigma_t^2 = \tilde{m}_t dt + \tilde{\sigma}_t d\tilde{W}_t, \tag{6.7}$$

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

$$IV_t \equiv \int_{t-1}^{t} \sigma_u^2 du \quad \text{and} \quad RV_t(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2}, \tag{6.8}$$

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.

6.2.1 Second example: distribution of standardized residuals

In their forecasting paper, ABDL (2002) presented empirical results suggesting that the standardized residuals $\varepsilon_t(h)$ defined by $\varepsilon_t(h) \equiv (r_t/\sqrt{RV_t(h)})$ where r_t is the daily return, i.e. $r_t = \log(S_t) - \log(S_{t-1})$, are $\mathcal{N}(0,1)$. These empirical results were based on the skewness, kurtosis, and nonparametric estimation of density of $\varepsilon_t(h)$. This is our second example.

We consider the same data as in ABDL (2002),[17] 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. For each variable, we define the standardized residuals and test their normality.

We start by testing that the standardized residuals are $\mathcal{N}(0,1)$, i.e., we assume as ABDL(2002) that the mean and variance of the normal distribution are known and equal to zero and one respectively. The results are provided in Table 9a. In Panel A, we provide the results when one assumes that the standardized residuals are independent. Given that this independence is not obvious, we also provided in Panel B results of the test statistics that do not assume the independence and use a HAC method for estimating the weighting matrix.

The results form Table 9a clearly reject the $\mathcal{N}(0,1)$ assumption, particularly test based on the fourth Hermite polynomial which is powerful. Note that the results in Panel B are close but different from ones of Panel B, suggesting that some feature (like dependence or heteroskedasticity) of the data are not taken into account in Panel A.

Another striking result in Panel A is the difference between the Jarque-Bera (JB) test and the test based on third and fourth Hermite polynomials (H_{3-4}). This difference suggests that the mean of the residuals is not zero or that their variance is not one. Therefore, we also

---

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

test normality of the residuals with unknown mean and variance. The results are reported in Panel A of Table 9b when one assumes independence and in Panel B when one estimates the weighting matrix by a HAC procedure. The main message from Table 9b is that normality is not rejected for standardized residuals based on five-minute realized volatility and rejected for ones based on thirty-minute realized volatility.

6.2.2 Third example: distribution of realized volatility

Another assumption made in ABDL (2002) is conditional normality of the log of realized volatility. Hence, log-realized volatility are also unconditionally normal. This is our third example. In Table 10, we provide the results of testing normality of log-realized volatility with unknown mean and variance. The weighting matrix is again 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.[18] Observe that the rejection is due to the asymmetry of the distribution. In this test, we assume that the weighting matrix is well defined. This is not necessarily the case. In particular, ABDL (2002) 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 (2002). which are based on the skewness, kurtosis, and nonparametric estimation of density of log-realized volatilities. Testing normality under long memory is more difficult and is left for future research.

7 Conclusion

In this paper, we consider testing marginal normal distributional assumptions for both cross-section and time series data. We use 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 provide simulation results that clearly show the usefulness of our approach. We also apply 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 one 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.

---

[18]In their study, ABDL (2002) used thirty-minute realized volatilities.

References

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

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

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 (2001), "Parametric and Nonparametric Measurements of Volatility," in Y. Ait-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 (2002), "Modeling and Forecasting Realized Volatility," unpublished manuscript, Northwestern University.

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. and S. Ng (2001), "Tests for Skewness, Kurtosis, and Normality for Time Series Data," unpublished manuscript, Boston College.

Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian OU based Models and some of their uses in Financial Economics," *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*, forthcoming.

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.

Berkowitz, J. (2001), "Testing Density Forecasts, With Applications to Risk Management," *Journal of Business & Economic Statistics*, 19, 465-474.

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.

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.

Christoffersen, P.F. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

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.

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.

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 (2002), "The Score of Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations, an LM Test for Multivariate Normality," unpublished manuscript, CEMFI.

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.

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.

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.

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

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.

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.

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.

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.

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.

Table 1: Size of the tests.
Mean and variance are known.

T	100	250	500	1000
H_1	.0508	.0514	.0525	.0500
H_2	.0511	.0485	.0498	.0486
H_3	.0549	.0542	.0556	.0515
H_4	.0475	.0470	.0494	.0498
H_5	.0362	.0431	.0483	.0504
H_6	.0205	.0266	.0320	.0361
H_7	.0128	.0156	.0199	.0239
H_8	.0135	.0116	.0126	.0142
H_9	.0188	.0168	.0143	.0128
H_{10}	.0122	.0160	.0184	.0158
H_{3-4}	.0582	.0562	.0565	.0531
H_{3-5}	.0599	.0624	.0632	.0599
H_{3-6}	.0544	.0580	.0609	.0611
H_{3-7}	.0503	.0525	.0552	.0558
H_{3-8}	.0504	.0499	.0513	.0512
H_{3-9}	.0480	.0488	.0500	.0487
H_{3-10}	.0445	.0458	.0484	.0474
KS	.0446	.0470	.0493	.0486
JB	.1498	.1632	.1743	.1741

Note: The data are i.i.d. from a $\mathcal{N}(0,1)$ distribution.
We test the $\mathcal{N}(0,1)$ assumption. Thus, we do not
estimate the mean and standard deviation. The
results are based on 50000 replications. For each
sample size, we provide the frequency 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.

Table 2: Size of the tests.
Mean and variance are estimated.

T	100	250	500	1000
H_3	.0429	.0475	.0490	.0497
H_4	.0308	.0378	.0427	.0456
H_5	.0242	.0360	.0432	.0478
H_6	.0121	.0214	.0288	.0342
H_7	.0068	.0122	.0168	.0217
H_8	.0077	.0086	.0101	.0130
H_9	.0123	.0143	.0112	.0105
H_{10}	.0060	.0126	.0166	.0151
H_{3-4}	.0410	.0455	.0463	.0480
H_{3-5}	.0418	.0513	.0532	.0543
H_{3-6}	.0361	.0473	.0526	.0553
H_{3-7}	.0331	.0423	.0469	.0506
H_{3-8}	.0325	.0396	.0429	.0464
H_{3-9}	.0309	.0389	.0422	.0440
H_{3-10}	.0284	.0368	.0406	.0428
KS[1]	.0001	.0002	.0001	.0002
M-KS[2]	.0519	.0537	.0567	.0596
JB[3]	.0433	.0473	.0470	.0485

Note: The data are i.i.d. from a $\mathcal{N}(0,1)$ distribution.
We test the normality assumption. Thus, we
estimate the mean and standard deviation. The
results are based on 50000 replications. For each
sample size, we provide the frequency 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.
M-KS is the modified Kolmogorov-Smirnov test.

Table 3: Power of the tests against Student distributions.

	Panel A: $\nu=60$					Panel B: $\nu=30$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0566	.0649	.0683	.0680	H_3	.0744	.0863	.0897	.0942
H_4	.0487	.0729	.0949	.1303	H_4	.0782	.1330	.1942	.3033
H_5	.0373	.0614	.0783	.0946	H_5	.0559	.0982	.1294	.1635
H_6	.0210	.0392	.0546	.0737	H_6	.0331	.0660	.0986	.1373
H_7	.0114	.0236	.0351	.0539	H_7	.0202	.0432	.0710	.1055
H_8	.0122	.0167	.0243	.0350	H_8	.0188	.0300	.0500	.0715
H_9	.0204	.0245	.0250	.0276	H_9	.0325	.0415	.0445	.0562
H_{10}	.0124	.0259	.0347	.0353	H_{10}	.0214	.0479	.0671	.0737
H_{3-4}	.0627	.0836	.1003	.1255	H_{3-4}	.0930	.1403	.1915	.2797
H_{3-5}	.0636	.0913	.1129	.1403	H_{3-5}	.0940	.1488	.2046	.2937
H_{3-6}	.0565	.0855	.1081	.1387	H_{3-6}	.0859	.1415	.1974	.2870
H_{3-7}	.0515	.0770	.0994	.1296	H_{3-7}	.0794	.1302	.1827	.2701
H_{3-8}	.0508	.0740	.0935	.1211	H_{3-8}	.0774	.1239	.1734	.2539
H_{3-9}	.0486	.0722	.0913	.1161	H_{3-9}	.0743	.1222	.1696	.2448
H_{3-10}	.0454	.0682	.0881	.1131	H_{3-10}	.0696	.1165	.1637	.2358
KS	.0001	.0001	.0001	.0003	KS	.0001	.0002	.0003	.0004
M-KS	.0532	.0602	.0641	.0689	M-KS	.0613	.0687	.0805	.1003
JB	.0658	.0854	.1015	.1263	JB	.0973	.1426	.1932	.2809
	Panel C: $\nu = 20$					Panel D: $\nu=10$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0961	.1172	.1231	.1296	H_3	.2508	.3078	.3345	.3599
H_4	.1117	.2109	.3423	.5396	H_4	.3224	.5921	.8272	.9722
H_5	.0771	.1420	.1980	.2526	H_5	.2058	.3671	.4945	.6014
H_6	.0489	.1031	.1579	.2271	H_6	.1591	.3158	.4629	.6181
H_7	.0297	.0719	.1210	.1862	H_7	.1094	.2368	.3780	.5452
H_8	.0270	.0538	.0858	.1369	H_8	.1053	.2056	.3230	.4793
H_9	.0470	.0670	.0795	.1104	H_9	.1450	.2353	.3079	.4237
H_{10}	.0340	.0773	.1113	.1378	H_{10}	.1143	.2486	.3700	.5021
H_{3-4}	.1290	.2148	.3265	.5001	H_{3-4}	.3374	.5793	.7997	.9602
H_{3-5}	.1285	.2229	.3348	.4996	H_{3-5}	.3219	.5666	.7858	.9526
H_{3-6}	.1189	.2126	.3232	.4876	H_{3-6}	.3086	.5567	.7811	.9527
H_{3-7}	.1109	.1977	.3034	.4632	H_{3-7}	.2920	.5308	.7556	.9427
H_{3-8}	.1084	.1897	.2881	.4418	H_{3-8}	.2813	.5113	.7372	.9323
H_{3-9}	.1045	.1862	.2815	.4264	H_{3-9}	.2702	.4976	.7227	.9232
H_{3-10}	.0985	.1788	.2731	.4127	H_{3-10}	.2579	.4810	.7068	.9155
KS	.0003	.0005	.0005	.0011	KS	.0047	.0093	.0214	.0698
M-KS	.0664	.0832	.1074	.1556	M-KS	.2003	.3081	.4624	.7087
JB	.1339	.2179	.3284	.5013	JB	.3448	.5826	.8009	.9605

Note: The data are i.i.d. from a $T(\nu)$ distribution. We test the normality assumption. Thus, we estimate the mean and variance. The results are based on 50000 replications. For each sample size, we provide the frequency 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. Continued.

	Panel E: $\nu=6$					Panel F: $\nu=3$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.3143	.4085	.4654	.5183	H_3	.6279	.7567	.8251	.8753
H_4	.5060	.8449	.9802	.9999	H_4	.8975	.9983	1.000	1.000
H_5	.3321	.5637	.7138	.8223	H_5	.6973	.9041	.9633	.9849
H_6	.2710	.5219	.7284	.8943	H_6	.6649	.9199	.9903	.9998
H_7	.2080	.4501	.6682	.8445	H_7	.5831	.8882	.9780	.9959
H_8	.1808	.3894	.6078	.8153	H_8	.5339	.8583	.9770	.9991
H_9	.2496	.3968	.5646	.7597	H_9	.5982	.8432	.9640	.9961
H_{10}	.2278	.4615	.6545	.8253	H_{10}	.6109	.8973	.9792	.9971
H_{3-4}	.5117	.8296	.9738	.9998	H_{3-4}	.8927	.9976	1.000	1.000
H_{3-5}	.5013	.8183	.9695	.9995	H_{3-5}	.8842	.9967	1.000	1.000
H_{3-6}	.4946	.8203	.9710	.9997	H_{3-6}	.8935	.9979	1.000	1.000
H_{3-7}	.4766	.8028	.9653	.9994	H_{3-7}	.8824	.9970	1.000	1.000
H_{3-8}	.4650	.7881	.9601	.9993	H_{3-8}	.8742	.9966	1.000	1.000
H_{3-9}	.4540	.7794	.9563	.9991	H_{3-9}	.8673	.9960	1.000	1.000
H_{3-10}	.4404	.7659	.9512	.9989	H_{3-10}	.8578	.9956	1.000	1.000
KS	.0110	.0338	.1152	.4167	KS	.2398	.6816	.9726	1.000
M-KS	.2450	.4787	.7661	.9701	M-KS	.7381	.9789	.9999	1.000
JB	.5195	.8322	.9742	.9998	JB	.8961	.9977	1.000	1.000

Table 4: Probability of rejection for Student distributions.

ν	$A(\nu)$	$P(A(\nu)\chi^2(1) > 3.84)$
3	–	1
6	–	1
10	4.16	.337
20	1.74	.137
30	1.41	.099
60	1.17	.071

Table 5: Power of the tests against asymmetric distributions.

	Panel A: $\chi^2(1)$ distribution.					**Panel B:** $\exp(1)$ distribution.			
T	100	250	500	1000	T	100	250	500	1000
H_3	1.000	1.000	1.000	1.000	H_3	1.000	1.000	1.000	1.000
H_4	.9887	1.000	1.000	1.000	H_4	.8964	.9980	1.000	1.000
H_5	.8423	.9779	.9993	1.000	H_5	.7001	.8755	.9699	.9981
H_6	.8214	.9855	.9997	1.000	H_6	.6161	.9094	.9909	.9999
H_7	.8620	.9970	1.000	1.000	H_7	.6281	.9434	.9977	1.000
H_8	.7812	.9593	.9952	1.000	H_8	.6366	.8766	.9533	.9924
H_9	.9162	.9720	.9926	.9994	H_9	.7427	.9316	.9630	.9871
H_{10}	.7859	.9400	.9879	.9988	H_{10}	.6159	.7908	.9102	.9762
H_{3-4}	1.000	1.000	1.000	1.000	H_{3-4}	1.000	1.000	1.000	1.000
H_{3-5}	1.000	1.000	1.000	1.000	H_{3-5}	1.000	1.000	1.000	1.000
H_{3-6}	1.000	1.000	1.000	1.000	H_{3-6}	1.000	1.000	1.000	1.000
H_{3-7}	1.000	1.000	1.000	1.000	H_{3-7}	1.000	1.000	1.000	1.000
H_{3-8}	1.000	1.000	1.000	1.000	H_{3-8}	1.000	1.000	1.000	1.000
H_{3-9}	1.000	1.000	1.000	1.000	H_{3-9}	.9999	1.000	1.000	1.000
H_{3-10}	1.000	1.000	1.000	1.000	H_{3-10}	.9999	1.000	1.000	1.000
KS	1.000	1.000	1.000	1.000	KS	.9191	1.000	1.000	1.000
M-KS	1.000	1.000	1.000	1.000	M-KS	1.000	1.000	1.000	1.000
JB	1.000	1.000	1.000	1.000	JB	1.000	1.000	1.000	1.000

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

Table 6a: Size of the tests under serial correlation which is ignored.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0415	.0466	.0480	.0501	H_3	.0595	.0735	.0773	.0812
H_4	.0300	.0393	.0421	.0451	H_4	.0266	.0415	.0502	.0568
H_5	.0234	.0352	.0424	.0479	H_5	.0199	.0344	.0433	.0491
H_6	.0119	.0214	.0283	.0354	H_6	.0093	.0193	.0261	.0333
H_7	.0063	.0121	.0167	.0218	H_7	.0058	.0118	.0156	.0210
H_8	.0076	.0092	.0103	.0141	H_8	.0068	.0085	.0100	.0127
H_9	.0111	.0134	.0109	.0111	H_9	.0084	.0124	.0108	.0111
H_{10}	.0058	.0128	.0162	.0147	H_{10}	.0043	.0115	.0147	.0152
H_{3-4}	.0390	.0458	.0454	.0474	H_{3-4}	.0465	.0606	.0686	.0750
H_{3-5}	.0409	.0518	.0531	.0540	H_{3-5}	.0455	.0625	.0701	.0753
H_{3-6}	.0357	.0473	.0519	.0548	H_{3-6}	.0388	.0555	.0652	.0712
H_{3-7}	.0323	.0420	.0462	.0513	H_{3-7}	.0342	.0481	.0567	.0633
H_{3-8}	.0316	.0397	.0426	.0468	H_{3-8}	.0325	.0445	.0509	.0567
H_{3-9}	.0303	.0392	.0420	.0446	H_{3-9}	.0293	.0418	.0481	.0521
H_{3-10}	.0279	.0362	.0405	.0436	H_{3-10}	.0259	.0383	.0449	.0490
KS	.0001	.0001	.0002	.0002	KS	.0004	.0002	.0004	.0004
M-KS	.0521	.0560	.0562	.0594	M-KS	.0644	.0713	.0744	.0758
JB	.0413	.0473	.0463	.0479	JB	.0493	.0623	.0696	.0755

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0972	.1348	.1497	.1611	H_3	.1920	.3009	.3595	.3906
H_4	.0297	.0593	.0854	.0992	H_4	.0789	.1964	.2617	.3030
H_5	.0233	.0366	.0530	.0646	H_5	.1134	.1373	.1599	.1794
H_6	.0119	.0198	.0298	.0394	H_6	.0667	.0959	.1111	.1265
H_7	.0062	.0129	.0175	.0232	H_7	.0337	.0513	.0730	.0917
H_8	.0063	.0102	.0123	.0160	H_8	.0228	.0378	.0534	.0694
H_9	.0059	.0104	.0125	.0145	H_9	.0102	.0256	.0377	.0509
H_{10}	.0032	.0073	.0126	.0145	H_{10}	.0077	.0179	.0263	.0362
H_{3-4}	.0655	.1067	.1395	.1586	H_{3-4}	.1519	.3401	.4423	.4984
H_{3-5}	.0692	.1044	.1306	.1493	H_{3-5}	.1983	.3351	.4207	.4825
H_{3-6}	.0582	.0909	.1161	.1354	H_{3-6}	.2117	.3423	.4167	.4693
H_{3-7}	.0515	.0788	.1016	.1179	H_{3-7}	.1964	.3196	.3983	.4600
H_{3-8}	.0454	.0698	.0912	.1038	H_{3-8}	.1849	.2994	.3759	.4385
H_{3-9}	.0392	.0621	.0819	.0939	H_{3-9}	.1645	.2780	.3558	.4184
H_{3-10}	.0353	.0555	.0736	.0849	H_{3-10}	.1490	.2547	.3328	.3958
KS	.0012	.0015	.0018	.0019	KS	.0310	.0506	.0606	.0631
M-KS	.1082	.1199	.1275	.1328	M-KS	.3212	.3971	.4298	.4477
JB	.0694	.1090	.1408	.1595	JB	.1621	.3465	.4457	.4998

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 test the normality assumption. We do not take into account the serial correlation in the tests. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j}, KS, M-KS and JB are defined in Table 5.

Table 6b: Size of the tests under serial correlation.
The serial correlation is known and taken into account; ρ is known.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0409	.0462	.0477	.0498	H_3	.0346	.0429	.0456	.0477
H_4	.0300	.0392	.0420	.0451	H_4	.0225	.0331	.0361	.0435
H_5	.0234	.0352	.0424	.0479	H_5	.0167	.0308	.0376	.0455
H_6	.0119	.0214	.0283	.0354	H_6	.0091	.0183	.0256	.0335
H_7	.0063	.0121	.0167	.0218	H_7	.0050	.0108	.0149	.0211
H_8	.0076	.0092	.0103	.0141	H_8	.0064	.0083	.0108	.0128
H_9	.0111	.0134	.0109	.0111	H_9	.0082	.0115	.0113	.0110
H_{10}	.0058	.0128	.0162	.0147	H_{10}	.0045	.0104	.0147	.0138
H_{3-4}	.0388	.0455	.0452	.0472	H_{3-4}	.0309	.0405	.0424	.0469
H_{3-5}	.0406	.0517	.0529	.0540	H_{3-5}	.0318	.0445	.0482	.0533
H_{3-6}	.0355	.0472	.0517	.0548	H_{3-6}	.0276	.0403	.0475	.0541
H_{3-7}	.0322	.0419	.0461	.0511	H_{3-7}	.0251	.0365	.0421	.0496
H_{3-8}	.0316	.0397	.0426	.0467	H_{3-8}	.0246	.0346	.0395	.0458
H_{3-9}	.0302	.0391	.0419	.0445	H_{3-9}	.0230	.0337	.0387	.0439
H_{3-10}	.0277	.0361	.0404	.0435	H_{3-10}	.0212	.0316	.0372	.0424

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0265	.0388	.0415	.0444	H_3	.0090	.0209	.0322	.0405
H_4	.0143	.0257	.0346	.0364	H_4	.0040	.0106	.0187	.0264
H_5	.0099	.0210	.0308	.0391	H_5	.0030	.0073	.0135	.0217
H_6	.0055	.0130	.0196	.0271	H_6	.0015	.0035	.0082	.0130
H_7	.0036	.0081	.0115	.0173	H_7	.0010	.0025	.0064	.0090
H_8	.0044	.0075	.0096	.0116	H_8	.0013	.0032	.0067	.0084
H_9	.0049	.0086	.0104	.0112	H_9	.0010	.0034	.0065	.0098
H_{10}	.0027	.0080	.0115	.0134	H_{10}	.0008	.0021	.0046	.0084
H_{3-4}	.0236	.0349	.0395	.0410	H_{3-4}	.0083	.0187	.0298	.0370
H_{3-5}	.0222	.0363	.0441	.0471	H_{3-5}	.0068	.0168	.0290	.0379
H_{3-6}	.0184	.0328	.0416	.0455	H_{3-6}	.0060	.0135	.0250	.0336
H_{3-7}	.0170	.0296	.0374	.0419	H_{3-7}	.0053	.0129	.0231	.0307
H_{3-8}	.0162	.0283	.0350	.0388	H_{3-8}	.0047	.0117	.0219	.0296
H_{3-9}	.0146	.0272	.0341	.0376	H_{3-9}	.0040	.0107	.0206	.0277
H_{3-10}	.0133	.0254	.0321	.0364	H_{3-10}	.0035	.0097	.0197	.0267

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 test the normality assumption. We take into account the serial correlation in the tests. We assume that we know the AR(1) dynamics and that we know ρ. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 6c: Size of the tests under serial correlation.
The serial correlation is known and taken into account; ρ is estimated.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0412	.0464	.0479	.0499	H_3	.0339	.0430	.0459	.0478
H_4	.0300	.0393	.0420	.0451	H_4	.0221	.0326	.0364	.0435
H_5	.0234	.0352	.0424	.0479	H_5	.0165	.0306	.0378	.0455
H_6	.0119	.0214	.0283	.0354	H_6	.0092	.0184	.0257	.0334
H_7	.0063	.0121	.0167	.0218	H_7	.0051	.0108	.0149	.0211
H_8	.0076	.0092	.0103	.0141	H_8	.0065	.0084	.0108	.0128
H_9	.0111	.0134	.0109	.0111	H_9	.0083	.0115	.0113	.0110
H_{10}	.0058	.0128	.0162	.0147	H_{10}	.0045	.0104	.0147	.0137
H_{3-4}	.0389	.0457	.0453	.0473	H_{3-4}	.0315	.0400	.0421	.0472
H_{3-5}	.0408	.0517	.0529	.0539	H_{3-5}	.0318	.0444	.0486	.0532
H_{3-6}	.0356	.0473	.0517	.0548	H_{3-6}	.0275	.0408	.0473	.0537
H_{3-7}	.0322	.0419	.0460	.0512	H_{3-7}	.0245	.0367	.0426	.0500
H_{3-8}	.0316	.0397	.0426	.0467	H_{3-8}	.0245	.0350	.0395	.0459
H_{3-9}	.0303	.0392	.0419	.0446	H_{3-9}	.0229	.0338	.0388	.0439
H_{3-10}	.0279	.0362	.0404	.0435	H_{3-10}	.0211	.0317	.0374	.0425

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0235	.0376	.0414	.0437	H_3	.0023	.0163	.0293	.0381
H_4	.0126	.0243	.0339	.0361	H_4	.0016	.0085	.0172	.0258
H_5	.0096	.0204	.0305	.0391	H_5	.0020	.0064	.0130	.0213
H_6	.0052	.0130	.0197	.0268	H_6	.0005	.0033	.0082	.0130
H_7	.0035	.0080	.0116	.0170	H_7	.0007	.0022	.0057	.0090
H_8	.0042	.0074	.0094	.0117	H_8	.0006	.0023	.0067	.0080
H_9	.0048	.0087	.0105	.0111	H_9	.0005	.0024	.0062	.0094
H_{10}	.0029	.0079	.0116	.0135	H_{10}	.0005	.0016	.0041	.0079
H_{3-4}	.0204	.0335	.0383	.0400	H_{3-4}	.0029	.0144	.0273	.0353
H_{3-5}	.0198	.0359	.0432	.0465	H_{3-5}	.0026	.0137	.0268	.0370
H_{3-6}	.0171	.0319	.0411	.0460	H_{3-6}	.0021	.0113	.0233	.0334
H_{3-7}	.0157	.0293	.0371	.0419	H_{3-7}	.0019	.0105	.0224	.0305
H_{3-8}	.0149	.0284	.0351	.0394	H_{3-8}	.0015	.0099	.0210	.0291
H_{3-9}	.0137	.0273	.0336	.0375	H_{3-9}	.0015	.0089	.0199	.0278
H_{3-10}	.0125	.0258	.0319	.0364	H_{3-10}	.0014	.0082	.0186	.0264

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 test the normality assumption. We take into account the serial correlation in the tests. We assume that we know the AR(1) dynamics but not ρ which is estimated by OLS. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 6d: Size of the tests under serial correlation.
The serial correlation is unknown; Σ is estimated by a HAC procedure.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0338	.0416	.0463	.0477	H_3	.0334	.0424	.0464	.0501
H_4	.0556	.0756	.0754	.0694	H_4	.0386	.0665	.0732	.0713
H_5	.0404	.0407	.0420	.0415	H_5	.0421	.0431	.0428	.0434
H_6	.0441	.0416	.0459	.0497	H_6	.0452	.0442	.0463	.0466
H_7	.0372	.0431	.0468	.0469	H_7	.0380	.0424	.0442	.0469
H_8	.0387	.0444	.0483	.0494	H_8	.0360	.0426	.0466	.0466
H_9	.0402	.0450	.0441	.0453	H_9	.0413	.0429	.0443	.0450
H_{10}	.0400	.0462	.0456	.0455	H_{10}	.0379	.0430	.0449	.0461
H_{3-4}	.0423	.0822	.0894	.0825	H_{3-4}	.0276	.0702	.0875	.0879
H_{3-5}	.0186	.0642	.1022	.1090	H_{3-5}	.0141	.0424	.0804	.1017
H_{3-6}	.0157	.0356	.1362	.2218	H_{3-6}	.0133	.0233	.0648	.1599
H_{3-7}	.0286	.0342	.0971	.2402	H_{3-7}	.0292	.0254	.0495	.1393
H_{3-8}	.1267	.0648	.0809	.2682	H_{3-8}	.1311	.0593	.0536	.1119
H_{3-9}	.2706	.1249	.1014	.2160	H_{3-9}	.2866	.1325	.0828	.1062
H_{3-10}	.6553	.3723	.2440	.2326	H_{3-10}	.6810	.3951	.2390	.1703

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0293	.0400	.0436	.0479	H_3	.0156	.0258	.0346	.0399
H_4	.0229	.0512	.0708	.0747	H_4	.0065	.0182	.0383	.0613
H_5	.0477	.0475	.0470	.0456	H_5	.0449	.0504	.0525	.0504
H_6	.0534	.0475	.0463	.0472	H_6	.0485	.0638	.0576	.0518
H_7	.0387	.0449	.0459	.0475	H_7	.0411	.0416	.0435	.0463
H_8	.0387	.0420	.0452	.0485	H_8	.0483	.0467	.0455	.0485
H_9	.0400	.0436	.0445	.0438	H_9	.0371	.0445	.0443	.0461
H_{10}	.0373	.0429	.0461	.0465	H_{10}	.0377	.0436	.0462	.0492
H_{3-4}	.0154	.0461	.0771	.0872	H_{3-4}	.0053	.0118	.0283	.0571
H_{3-5}	.0110	.0240	.0483	.0775	H_{3-5}	.0047	.0093	.0151	.0284
H_{3-6}	.0128	.0179	.0278	.0656	H_{3-6}	.0043	.0131	.0159	.0191
H_{3-7}	.0291	.0234	.0268	.0498	H_{3-7}	.0108	.0288	.0301	.0238
H_{3-8}	.1344	.0700	.0459	.0472	H_{3-8}	.0322	.1309	.1119	.0635
H_{3-9}	.2868	.1668	.0925	.0640	H_{3-9}	.0714	.2551	.2397	.1385
H_{3-10}	.6705	.4716	.2861	.1618	H_{3-10}	.1330	.5814	.5634	.3831

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 test the normality assumption. We take into account the serial correlation in the tests. We assume that we do not know the AR(1) dynamics. We use a HAC method. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 7a: Power of the tests under serial correlation against T(60) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0200	.0699	.1198	.2102	H_3	.0071	.0223	.0471	.0957
H_4	.8742	.8853	.9016	.8263	H_4	.9173	.9253	.9305	.7719
H_5	.8902	.8944	.8962	.6667	H_5	.9293	.9352	.9362	.6891
H_6	.8844	.8930	.9049	.6985	H_6	.9274	.9347	.9385	.7109
H_7	.8772	.8839	.8925	.6633	H_7	.9234	.9311	.9343	.6728
H_8	.8670	.8749	.8820	.6002	H_8	.9178	.9267	.9300	.6267
H_9	.8566	.8669	.8732	.5557	H_9	.9114	.9224	.9262	.5713
H_{10}	.8532	.8619	.8680	.5285	H_{10}	.9079	.9186	.9227	.5396
H_{3-4}	.0297	.0574	.0843	.1185	H_{3-4}	.0124	.0236	.0388	.0826
H_{3-5}	.0037	.0039	.0083	.0627	H_{3-5}	.0013	.0018	.0026	.0287
H_{3-6}	.8434	.8620	.8638	.4546	H_{3-6}	.8968	.9155	.9194	.4587
H_{3-7}	.8665	.8763	.8871	.5990	H_{3-7}	.9168	.9270	.9307	.6120
H_{3-8}	.8664	.8737	.8803	.5929	H_{3-8}	.9168	.9263	.9292	.6077
H_{3-9}	.8591	.8679	.8741	.5539	H_{3-9}	.9131	.9236	.9269	.5803
H_{3-10}	.8501	.8611	.8669	.5062	H_{3-10}	.9075	.9194	.9235	.5312

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0019	.0056	.0114	.0330	H_3	.0002	.0007	.0008	.0014
H_4	.9414	.9493	.9525	.7186	H_4	.0025	.9756	.9815	.0401
H_5	.9510	.9570	.9584	.7218	H_5	.0017	.9793	.9848	.0359
H_6	.9511	.9573	.9598	.7137	H_6	.0057	.9800	.9850	.0634
H_7	.9487	.9555	.9584	.6979	H_7	.0208	.9796	.9846	.0862
H_8	.9449	.9532	.9563	.6617	H_8	.0018	.9782	.9839	.0355
H_9	.9392	.9506	.9541	.6043	H_9	.0002	.9764	.9830	.0126
H_{10}	.9331	.9480	.9521	.5406	H_{10}	.0002	.9742	.9821	.0060
H_{3-4}	.0064	.0093	.0138	.0500	H_{3-4}	.0055	.0025	.0021	.0376
H_{3-5}	.0006	.0012	.0013	.0125	H_{3-5}	.0001	.0002	.0004	.0017
H_{3-6}	.9123	.9438	.9489	.4083	H_{3-6}	.0014	.9548	.9783	.0077
H_{3-7}	.9421	.9526	.9562	.6140	H_{3-7}	.0004	.9764	.9832	.0143
H_{3-8}	.9434	.9526	.9561	.6371	H_{3-8}	.0013	.9773	.9834	.0285
H_{3-9}	.9404	.9511	.9546	.6142	H_{3-9}	.0004	.9764	.9830	.0163
H_{3-10}	.9352	.9488	.9529	.5630	H_{3-10}	.0002	.9750	.9824	.0078

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim T(60)$. We test the normality assumption. We take into account the serial correlation in the tests. We assume that we do not know the AR(1) dynamics. We use a HAC method. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 7b: Power of the tests under serial correlation against T(30) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0497	.1404	.2063	.2465	H_3	.0161	.0512	.0932	.1607
H_4	.8225	.8693	.9181	.9494	H_4	.8811	.9009	.9180	.8512
H_5	.8406	.8529	.8586	.7845	H_5	.8979	.9072	.9085	.6610
H_6	.8278	.8615	.8959	.8228	H_6	.8935	.9097	.9210	.7100
H_7	.8161	.8442	.8681	.8110	H_7	.8881	.9021	.9105	.6877
H_8	.7995	.8261	.8491	.7411	H_8	.8783	.8930	.9014	.6167
H_9	.7829	.8114	.8325	.7081	H_9	.8674	.8847	.8929	.5648
H_{10}	.7750	.8026	.8195	.6710	H_{10}	.8600	.8787	.8860	.5293
H_{3-4}	.0576	.0939	.1286	.0756	H_{3-4}	.0249	.0434	.0657	.0880
H_{3-5}	.0048	.0108	.0356	.0893	H_{3-5}	.0021	.0031	.0070	.0730
H_{3-6}	.7671	.7947	.7928	.5632	H_{3-6}	.8460	.8727	.8759	.4129
H_{3-7}	.7990	.8309	.8638	.7403	H_{3-7}	.8761	.8928	.9051	.6047
H_{3-8}	.7988	.8217	.8429	.7462	H_{3-8}	.8764	.8916	.8989	.6119
H_{3-9}	.7866	.8122	.8331	.6941	H_{3-9}	.8697	.8859	.8941	.5666
H_{3-10}	.7720	.8000	.8161	.6363	H_{3-10}	.8597	.8795	.8869	.5081

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0044	.0129	.0296	.0704	H_3	.0004	.0007	.0013	.0036
H_4	.9120	.9269	.9351	.6338	H_4	.0072	.9669	.9717	.0465
H_5	.9269	.9374	.9416	.5083	H_5	.0037	.9727	.9765	.0491
H_6	.9265	.9380	.9444	.5239	H_6	.0152	.9736	.9772	.0609
H_7	.9234	.9350	.9415	.5138	H_7	.0427	.9729	.9765	.0624
H_8	.9170	.9306	.9377	.4549	H_8	.0050	.9709	.9751	.0378
H_9	.9082	.9266	.9339	.3856	H_9	.0007	.9678	.9739	.0193
H_{10}	.8977	.9222	.9299	.3371	H_{10}	.0003	.9639	.9723	.0135
H_{3-4}	.0123	.0173	.0285	.0848	H_{3-4}	.0081	.0034	.0037	.0942
H_{3-5}	.0008	.0017	.0019	.0586	H_{3-5}	.0001	.0003	.0004	.0163
H_{3-6}	.8597	.9145	.9244	.2423	H_{3-6}	.0020	.9243	.9672	.0218
H_{3-7}	.9123	.9293	.9376	.4066	H_{3-7}	.0010	.9676	.9741	.0391
H_{3-8}	.9145	.9295	.9367	.4394	H_{3-8}	.0032	.9694	.9747	.0324
H_{3-9}	.9098	.9272	.9346	.3989	H_{3-9}	.0012	.9681	.9739	.0246
H_{3-10}	.9023	.9237	.9316	.3473	H_{3-10}	.0003	.9654	.9728	.0149

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

Table 7c: Power of the tests under serial correlation against T(20) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0790	.1961	.2767	.3733	H_3	.0262	.0781	.1322	.1828
H_4	.7912	.8757	.9494	.9932	H_4	.8508	.8902	.9243	.9676
H_5	.7981	.8185	.8396	.8701	H_5	.8701	.8872	.8884	.8972
H_6	.7882	.8481	.9053	.9616	H_6	.8637	.8962	.9163	.9465
H_7	.7689	.8189	.8673	.9213	H_7	.8544	.8832	.8983	.9225
H_8	.7442	.7940	.8418	.8980	H_8	.8411	.8727	.8854	.9087
H_9	.7192	.7713	.8185	.8713	H_9	.8248	.8598	.8738	.8951
H_{10}	.7059	.7570	.8004	.8493	H_{10}	.8140	.8499	.8627	.8833
H_{3-4}	.0810	.1161	.1521	.2133	H_{3-4}	.0366	.0579	.0836	.1152
H_{3-5}	.0070	.0257	.0896	.2328	H_{3-5}	.0028	.0053	.0191	.0663
H_{3-6}	.6958	.7253	.7344	.7578	H_{3-6}	.7943	.8371	.8355	.8374
H_{3-7}	.7425	.8036	.8685	.9404	H_{3-7}	.8367	.8737	.8941	.9290
H_{3-8}	.7410	.7864	.8316	.8899	H_{3-8}	.8375	.8684	.8811	.9044
H_{3-9}	.7229	.7719	.8177	.8722	H_{3-9}	.8280	.8619	.8751	.8979
H_{3-10}	.7016	.7516	.7924	.8412	H_{3-10}	.8144	.8510	.8643	.8836

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0071	.0230	.0499	.0824	H_3	.0007	.0013	.0024	.0051
H_4	.8871	.9111	.9223	.9370	H_4	.0137	.9559	.9660	.9678
H_5	.9055	.9226	.9257	.9255	H_5	.0071	.9641	.9717	.9730
H_6	.9046	.9238	.9328	.9418	H_6	.0251	.9652	.9723	.9738
H_7	.9007	.9196	.9269	.9330	H_7	.0599	.9641	.9715	.9730
H_8	.8921	.9141	.9218	.9277	H_8	.0094	.9612	.9704	.9718
H_9	.8789	.9075	.9168	.9220	H_9	.0014	.9566	.9687	.9706
H_{10}	.8638	.9009	.9103	.9154	H_{10}	.0005	.9516	.9663	.9688
H_{3-4}	.0169	.0272	.0418	.0592	H_{3-4}	.0112	.0046	.0052	.0079
H_{3-5}	.0013	.0016	.0031	.0089	H_{3-5}	.0001	.0003	.0006	.0006
H_{3-6}	.8050	.8905	.9009	.8998	H_{3-6}	.0032	.8901	.9593	.9643
H_{3-7}	.8843	.9119	.9225	.9317	H_{3-7}	.0016	.9563	.9687	.9711
H_{3-8}	.8879	.9119	.9204	.9258	H_{3-8}	.0067	.9589	.9696	.9712
H_{3-9}	.8812	.9084	.9173	.9227	H_{3-9}	.0023	.9570	.9687	.9706
H_{3-10}	.8698	.9029	.9127	.9180	H_{3-10}	.0006	.9532	.9670	.9694

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

Table 7d: Power of the tests under serial correlation against T(10) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.1531	.3184	.4744	.6205	H_3	.0575	.1500	.2404	.3716
H_4	.7713	.9417	.9946	.9999	H_4	.7975	.9017	.9714	.9983
H_5	.7066	.7774	.8496	.9234	H_5	.7886	.8184	.8451	.8839
H_6	.7126	.8650	.9565	.9955	H_6	.7889	.8729	.9349	.9829
H_7	.6798	.8198	.9162	.9762	H_7	.7683	.8419	.8957	.9504
H_8	.6349	.7762	.8839	.9611	H_8	.7407	.8185	.8787	.9382
H_9	.5945	.7404	.8535	.9423	H_9	.7059	.7934	.8542	.9176
H_{10}	.5760	.7130	.8226	.9210	H_{10}	.6801	.7729	.8329	.8982
H_{3-4}	.1165	.1302	.1677	.2382	H_{3-4}	.0612	.0818	.1082	.1515
H_{3-5}	.0260	.1079	.2833	.4745	H_{3-5}	.0094	.0296	.1041	.2614
H_{3-6}	.5396	.6042	.6884	.8079	H_{3-6}	.6347	.6939	.7078	.7406
H_{3-7}	.6331	.8015	.9271	.9919	H_{3-7}	.7304	.8284	.9060	.9733
H_{3-8}	.6264	.7602	.8734	.9536	H_{3-8}	.7302	.8033	.8600	.9213
H_{3-9}	.5979	.7382	.8519	.9463	H_{3-9}	.7117	.7967	.8586	.9235
H_{3-10}	.5631	.6960	.8060	.9104	H_{3-10}	.6813	.7706	.8287	.8931

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0159	.0548	.1023	.1602	H_3	.0010	.0024	.0059	.0138
H_4	.8106	.8784	.9210	.9692	H_4	.0297	.9276	.9441	.9520
H_5	.8382	.8741	.8809	.8906	H_5	.0201	.9412	.9532	.9576
H_6	.8326	.8862	.9188	.9541	H_6	.0486	.9427	.9545	.9599
H_7	.8241	.8756	.9009	.9287	H_7	.0801	.9415	.9528	.9585
H_8	.8048	.8654	.8904	.9196	H_8	.0228	.9368	.9502	.9566
H_9	.7734	.8512	.8776	.9049	H_9	.0058	.9278	.9473	.9539
H_{10}	.7339	.8356	.8636	.8913	H_{10}	.0024	.9167	.9431	.9506
H_{3-4}	.0304	.0478	.0650	.0873	H_{3-4}	.0210	.0084	.0111	.0159
H_{3-5}	.0036	.0057	.0178	.0612	H_{3-5}	.0003	.0006	.0006	.0010
H_{3-6}	.6122	.8001	.8083	.8070	H_{3-6}	.0052	.7310	.9293	.9422
H_{3-7}	.7802	.8615	.8958	.9379	H_{3-7}	.0067	.9261	.9470	.9545
H_{3-8}	.7927	.8580	.8813	.9087	H_{3-8}	.0178	.9323	.9485	.9552
H_{3-9}	.7788	.8532	.8795	.9095	H_{3-9}	.0085	.9284	.9471	.9539
H_{3-10}	.7495	.8407	.8671	.8931	H_{3-10}	.0032	.9201	.9444	.9516

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

Table 7e: Power of the tests under serial correlation against T(6) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.2102	.4324	.5973	.7244	H_3	.0957	.2330	.4013	.6008
H_4	.8263	.9878	.9999	1.0000	H_4	.7719	.9512	.9969	1.0000
H_5	.6667	.8287	.9336	.9791	H_5	.6891	.7714	.8544	.9259
H_6	.6985	.9196	.9901	.9998	H_6	.7109	.8853	.9721	.9982
H_7	.6633	.8888	.9745	.9956	H_7	.6728	.8347	.9323	.9796
H_8	.6002	.8352	.9541	.9964	H_8	.6267	.8050	.9182	.9818
H_9	.5557	.8062	.9374	.9904	H_9	.5713	.7626	.8865	.9651
H_{10}	.5285	.7663	.9137	.9856	H_{10}	.5396	.7297	.8633	.9541
H_{3-4}	.1185	.1095	.1148	.1315	H_{3-4}	.0826	.0859	.0990	.1320
H_{3-5}	.0627	.2002	.3715	.5162	H_{3-5}	.0287	.1016	.2532	.4411
H_{3-6}	.4546	.6419	.8304	.9327	H_{3-6}	.4587	.5644	.6791	.8125
H_{3-7}	.5990	.8667	.9802	.9996	H_{3-7}	.6120	.8296	.9524	.9965
H_{3-8}	.5929	.8324	.9524	.9896	H_{3-8}	.6077	.7744	.8918	.9591
H_{3-9}	.5539	.7977	.9379	.9941	H_{3-9}	.5803	.7686	.8965	.9752
H_{3-10}	.5062	.7457	.9042	.9828	H_{3-10}	.5312	.7195	.8520	.9467

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0330	.0964	.1761	.3087	H_3	.0014	.0060	.0144	.0333
H_4	.7186	.8727	.9586	.9959	H_4	.0401	.8695	.9167	.9400
H_5	.7218	.7883	.8163	.8519	H_5	.0359	.8941	.9260	.9346
H_6	.7137	.8539	.9311	.9828	H_6	.0634	.8955	.9309	.9473
H_7	.6979	.8222	.8886	.9449	H_7	.0862	.8940	.9267	.9412
H_8	.6617	.8055	.8831	.9460	H_8	.0355	.8823	.9222	.9390
H_9	.6043	.7732	.8521	.9217	H_9	.0126	.8580	.9145	.9330
H_{10}	.5406	.7352	.8282	.9067	H_{10}	.0060	.8217	.9048	.9265
H_{3-4}	.0500	.0651	.0748	.0966	H_{3-4}	.0376	.0153	.0204	.0270
H_{3-5}	.0125	.0249	.0796	.2158	H_{3-5}	.0017	.0018	.0017	.0037
H_{3-6}	.4083	.6198	.6407	.6738	H_{3-6}	.0077	.4973	.8678	.8952
H_{3-7}	.6140	.8006	.9019	.9729	H_{3-7}	.0143	.8469	.9145	.9363
H_{3-8}	.6371	.7770	.8500	.9121	H_{3-8}	.0285	.8691	.9163	.9335
H_{3-9}	.6142	.7789	.8629	.9342	H_{3-9}	.0163	.8597	.9144	.9333
H_{3-10}	.5630	.7451	.8283	.9008	H_{3-10}	.0078	.8343	.9077	.9279

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ T(6). We test the normality assumption. We take into account the serial correlation in the tests. We assume that we do not know the AR(1) dynamics. We use a HAC method. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 7f: Power of the tests under serial correlation against T(3) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.2465	.4043	.5058	.5951	H_3	.1607	.3311	.4694	.5726
H_4	.9494	.9998	1.0000	1.0000	H_4	.8512	.9972	1.0000	1.0000
H_5	.7845	.9683	.9938	.9977	H_5	.6610	.9060	.9808	.9947
H_6	.8228	.9899	.9999	1.0000	H_6	.7100	.9637	.9986	1.0000
H_7	.8110	.9859	.9990	.9999	H_7	.6877	.9472	.9946	.9995
H_8	.7411	.9704	.9991	1.0000	H_8	.6167	.9153	.9915	.9999
H_9	.7081	.9591	.9970	1.0000	H_9	.5648	.8899	.9844	.9994
H_{10}	.6710	.9402	.9959	1.0000	H_{10}	.5293	.8645	.9773	.9992
H_{3-4}	.0756	.0309	.0135	.0056	H_{3-4}	.0880	.0487	.0296	.0170
H_{3-5}	.0893	.1807	.2523	.3223	H_{3-5}	.0730	.1557	.2371	.3142
H_{3-6}	.5632	.9090	.9789	.9900	H_{3-6}	.4129	.7762	.9425	.9804
H_{3-7}	.7403	.9795	.9997	1.0000	H_{3-7}	.6047	.9356	.9973	1.0000
H_{3-8}	.7462	.9732	.9974	.9993	H_{3-8}	.6119	.9164	.9888	.9982
H_{3-9}	.6941	.9583	.9985	1.0000	H_{3-9}	.5666	.8919	.9879	.9998
H_{3-10}	.6363	.9331	.9951	.9999	H_{3-10}	.5081	.8530	.9760	.9988

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	n	100	250	500	1000
H_3	.0704	.1917	.3516	.5074	H_3	.0036	.0183	.0449	.0914
H_4	.6338	.9452	.9978	1.0000	H_4	.0465	.6330	.8661	.9644
H_5	.5083	.7425	.8995	.9773	H_5	.0491	.6044	.7306	.7803
H_6	.5239	.8596	.9812	.9997	H_6	.0609	.6041	.8310	.9427
H_7	.5138	.8247	.9596	.9951	H_7	.0624	.6004	.7863	.8870
H_8	.4549	.7854	.9476	.9961	H_8	.0378	.5582	.7873	.9107
H_9	.3856	.7306	.9185	.9904	H_9	.0193	.4771	.7350	.8723
H_{10}	.3371	.6859	.8964	.9870	H_{10}	.0135	.3971	.6777	.8501
H_{3-4}	.0848	.0687	.0527	.0397	H_{3-4}	.0942	.0358	.0373	.0361
H_{3-5}	.0586	.1169	.1970	.2834	H_{3-5}	.0163	.0281	.0348	.0717
H_{3-6}	.2423	.5219	.7737	.9325	H_{3-6}	.0218	.1973	.4509	.5501
H_{3-7}	.4066	.7886	.9672	.9993	H_{3-7}	.0391	.4498	.7428	.9112
H_{3-8}	.4394	.7655	.9356	.9895	H_{3-8}	.0324	.5073	.7209	.8356
H_{3-9}	.3989	.7463	.9329	.9943	H_{3-9}	.0246	.4834	.7425	.8891
H_{3-10}	.3473	.6911	.8968	.9852	H_{3-10}	.0149	.4220	.6901	.8410

Note: The data follow an AR(1) process: $x_t = \rho x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ T(3). We test the normality assumption. We take into account the serial correlation in the tests. We assume that we do not know the AR(1) dynamics. We use a HAC method. The results are based on 50000 replications. For each sample size, we provide the frequency of rejection at a 5% level. The notations H_{i-j} are defined in Table 5.

Table 7g: Power of the tests under serial correlation against T(5) innovations.

	Panel A: $\rho = .1$					Panel B: $\rho = .5$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.2288	.4567	.6103	.7263	H_3	.1111	.2677	.4592	.6404
H_4	.8575	.9951	1.000	1.000	H_4	.7775	.9679	.9992	1.000
H_5	.6746	.8674	.9614	.9885	H_5	.6500	.7749	.8874	.9581
H_6	.7142	.9404	.9955	1.000	H_6	.6840	.8994	.9842	.9996
H_7	.6845	.9178	.9881	.9980	H_7	.6458	.8492	.9536	.9904
H_8	.6167	.8710	.9774	.9989	H_8	.5918	.8157	.9404	.9926
H_9	.5725	.8481	.9639	.9965	H_9	.5328	.7730	.9147	.9823
H_{10}	.5406	.8099	.9487	.9948	H_{10}	.5062	.7384	.8908	.9766
H_{3-4}	.1128	.0928	.0834	.0821	H_{3-4}	.0875	.0816	.0880	.1018
H_{3-5}	.0761	.2228	.3642	.4884	H_{3-5}	.0422	.1328	.2879	.4450
H_{3-6}	.4488	.7056	.8921	.9585	H_{3-6}	.4121	.5661	.7380	.8829
H_{3-7}	.6148	.9004	.9912	.9999	H_{3-7}	.5773	.8417	.9706	.9992
H_{3-8}	.6118	.8729	.9751	.9950	H_{3-8}	.5732	.7883	.9224	.9791
H_{3-9}	.5677	.8385	.9665	.9984	H_{3-9}	.5412	.7785	.9225	.9892
H_{3-10}	.5148	.7891	.9422	.9939	H_{3-10}	.4901	.7256	.8813	.9723

	Panel C: $\rho = .7$					Panel D: $\rho = .9$			
T	100	250	500	1000	T	100	250	500	1000
H_3	.0395	.1133	.2158	.3999	H_3	.0019	.0076	.0203	.0428
H_4	.6805	.8820	.9733	.9989	H_4	.0424	.8341	.9067	.9366
H_5	.6541	.7473	.7962	.8602	H_5	.0403	.8623	.9049	.9119
H_6	.6489	.8448	.9426	.9905	H_6	.0639	.8643	.9172	.9427
H_7	.6310	.7999	.8927	.9537	H_7	.0794	.8625	.9094	.9312
H_8	.5874	.7842	.8888	.9637	H_8	.0370	.8454	.9058	.9292
H_9	.5230	.7394	.8525	.9383	H_9	.0154	.8057	.8943	.9204
H_{10}	.4588	.6939	.8228	.9263	H_{10}	.0080	.7507	.8802	.9114
H_{3-4}	.0581	.0684	.0737	.0893	H_{3-4}	.0497	.0179	.0240	.0311
H_{3-5}	.0211	.0448	.1260	.2903	H_{3-5}	.0029	.0030	.0036	.0066
H_{3-6}	.3406	.5429	.5934	.6874	H_{3-6}	.0107	.4078	.8109	.8456
H_{3-7}	.5348	.7793	.9121	.9850	H_{3-7}	.0186	.7872	.8949	.9268
H_{3-8}	.5607	.7443	.8446	.9214	H_{3-8}	.0306	.8223	.8943	.9162
H_{3-9}	.5331	.7490	.8669	.9542	H_{3-9}	.0191	.8084	.8951	.9219
H_{3-10}	.4786	.7041	.8223	.9170	H_{3-10}	.0102	.7689	.8841	.9123

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

Table 8: Testing $\mathcal{N}(0,1)$ of fitted residuals for a GARCH(1,1) model.

	UK-US\$	FF-US\$	SF-US\$	Yen-US\$
H_3	1.86 (.173)	1.24 (.264)	51.0 (.000)	17.5 (.000)
H_4	42.6 (.000)	38.7 (.000)	189 (.000)	577 (.000)
H_5	9.53 (.002)	15.5 (.000)	590 (.000)	3713 (.000)
H_6	46.8 (.000)	126 (.000)	2562 (.000)	28553(.000)
H_7	26.9 (.000)	8.7 (.003)	10956 (.000)	181020 (.000)
H_8	45.9 (.000)	8.5 (.000)	35135 (.000)	945122 (.000)
H_9	17.2 (.000)	2.35 (.125)	88029.039 (.000)	4186878.386 (.000)
H_{10}	9.13 (.003)	42.9 (.000)	177511 (.000)	15683206 (.000)
H_{3-4}	44.4 (.000)	39.9 (.000)	240 (.000)	594 (.000)
H_{3-5}	54.1 (.000)	55.4 (.000)	830 (.000)	4308 (.000)
H_{3-6}	100 (.000)	182 (.000)	3393 (.000)	32861 (.000)
H_{3-7}	127 (.000)	191 (.000)	14349 (.000)	213882 (.000)
H_{3-8}	173 (.000)	271 (.000)	49485 (.000)	1159004 (.000)
H_{3-9}	191 (.000)	273 (.000)	137514 (.000)	5345882 (.000)
H_{3-10}	200 (.000)	316 (.000)	315026 (.000)	21029088 (.000)
KS	1.01 (.010)	.754 (1.00)	1.29 (.010)	1.07 (.010)
JB	44.6 (.000)	4.1 (.000)	240 (.000)	597 (.000)

Note: We test the $\mathcal{N}(0,1)$ assumption of the standardized residuals. The volatility model is a GARCH(1,1) and is estimated by the Guassian QML method. We report the test statistics and their corresponding p-values in parenthese. 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.

Table 9a: Testing $\mathcal{N}(0,1)$ of stantardized residuals by realized volatility.

Panel A: i.i.d. assumption.

	DM-US$-5	DM-US$-30	Yen-US$-5	Yen-US$-30	Yen-DM-5	Yen-DM-30
H_3	.259 (.6106)	.450 (.5023)	.019 (.8910)	.759 (.3837)	.172 (.6780)	.076 (.7822)
H_4	6.57 (.0104)	6.69 (.0097)	3.83 (.0503)	8.91 (.0028)	1.99 (.1577)	11.0 (.0009)
H_5	1.24 (.2649)	.507 (.4764)	.523 (.4696)	.435 (.5095)	.074 (.7859)	.004 (.9504)
H_6	1.24 (.2661)	2.38 (.1228)	.296 (.5865)	2.26 (.1325)	.019 (.8916)	6.18 (.0129)
H_7	1.29 (.2566)	.274 (.6006)	.590 (.4424)	.157 (.6923)	.039 (.8427)	.035 (.8510)
H_8	.710 (.3994)	.041 (.8397)	.153 (.6961)	.003 (.9593)	.057 (.8106)	1.31 (.2526)
H_9	.533 (.4653)	.074 (.7851)	.207 (.6491)	.111 (.7390)	.038 (.8446)	.007 (.9337)
H_{10}	.501 (.4791)	.125 (.7239)	.223 (.6370)	.354 (.5517)	.143 (.7055)	.035 (.8511)
H_{3-4}	6.82 (.0330)	7.14 (.0281)	3.85 (.1458)	9.67 (.0080)	2.17 (.3382)	11.1 (.0040)
H_{3-5}	8.07 (.0447)	7.65 (.0539)	4.37 (.2238)	1.1 (.0177)	2.24 (.5237)	11.1 (.0114)
H_{3-6}	9.30 (.0540)	1.0 (.0399)	4.67 (.3229)	12.4 (.0148)	2.26 (.6880)	17.2 (.0017)
H_{3-7}	1.6 (.0602)	1.3 (.0671)	5.26 (.3850)	12.5 (.0283)	2.30 (.8063)	17.3 (.0040)
H_{3-8}	11.3 (.0795)	1.3 (.1109)	5.41 (.4921)	12.5 (.0512)	2.36 (.8841)	18.6 (.0049)
H_{3-9}	11.8 (.1062)	1.4 (.1660)	5.62 (.5848)	12.6 (.0815)	2.40 (.9347)	18.6 (.0096)
H_{3-10}	12.3 (.1369)	1.5 (.2289)	5.84 (.6649)	13.0 (.1122)	2.54 (.9599)	18.6 (.0170)
KS	2.52 (.0100)	.598 (1.000)	2.43 (.0100)	.519 (1.000)	1.99 (.0100)	.788 (1.000)
JB	.998 (.6073)	11.9 (.0026)	.819 (.6640)	13.5 (.0011)	2.14 (.3428)	16.4 (.0003)

Panel B: serial correlation.

	DM-US$-5	DM-US$-30	Yen-US$-5	Yen-US$-30	Yen-DM-5	Yen-DM-30
H_3	.623 (.4300)	.889 (.3459)	.046 (.8303)	1.43 (.2319)	.425 (.5146)	.138 (.7105)
H_4	17.8 (.0000)	17.7 (.0000)	8.99 (.0027)	19.4 (.0000)	5.65 (.0175)	21.2 (.0000)
H_5	3.87 (.0492)	1.30 (.2533)	1.66 (.1975)	1.09 (.2973)	.224 (.6362)	.009 (.9240)
H_6	4.13 (.0420)	6.83 (.0090)	.963 (.3265)	6.22 (.0127)	.060 (.8069)	15.2 (.0001)
H_7	4.77 (.0290)	.913 (.3392)	2.04 (.1531)	.457 (.4991)	.142 (.7059)	.095 (.7583)
H_8	2.87 (.0904)	.140 (.7087)	.590 (.4426)	.008 (.9272)	.209 (.6472)	3.52 (.0606)
H_9	2.24 (.1346)	.259 (.6105)	.817 (.3660)	.364 (.5463)	.157 (.6924)	.024 (.8777)
H_{10}	2.24 (.1344)	.495 (.4817)	.947 (.3304)	1.25 (.2640)	.595 (.4405)	.114 (.7351)
H_{3-4}	19.9 (.0000)	18.6 (.0001)	1.1 (.0064)	22.1 (.0000)	6.23 (.0445)	22.5 (.0000)
H_{3-5}	23.2 (.0000)	19.1 (.0003)	11.9 (.0079)	22.2 (.0001)	6.26 (.0996)	24.5 (.0000)
H_{3-6}	32.2 (.0000)	2.0 (.0005)	19.7 (.0006)	24.2 (.0001)	15.2 (.0043)	24.8 (.0001)
H_{3-7}	32.8 (.0000)	2.0 (.0013)	19.7 (.0014)	25.3 (.0001)	15.7 (.0076)	25.2 (.0001)
H_{3-8}	65.1 (.0000)	23.1 (.0007)	46.3 (.0000)	25.5 (.0003)	47.5 (.0000)	25.3 (.0003)
H_{3-9}	76.0 (.0000)	23.6 (.0013)	46.5 (.0000)	27.7 (.0002)	48.1 (.0000)	25.4 (.0006)
H_{3-10}	178. (.0000)	73.9 (.0000)	66.4 (.0000)	33.3 (.0001)	146. (.0000)	25.4 (.0013)

Note: We test the $\mathcal{N}(0,1)$ assumption of the standardized residuals. The volatility is the daily realized volatility computed with five-minute and thirty-minute returns. We report the test statistics and their corresponding p-values in parenthese. We use the same dat as ABDL (2002). In Panel A, we assume that the residuals are i.i.d., while we allow them to be serially correlated in Panel B. In this case, we use a HAC method of Andrews (1991) to estimate the weighting matrix. H_{i-j} is the joint test based on H_k, $i \leq k \leq j$.

Table 9b: Testing normality of stantardized residuals by realized volatility.

Panel A: i.i.d. assumption.

	DM-US$-5	DM-US$-30	Yen-US$-5	Yen-US$-30	Yen-DM-5	Yen-DM-30
H_3	.789 (.3744)	.550 (.4584)	.142 (.7068)	.930 (.3348)	.154 (.6947)	.128 (.7203)
H_4	.011 (.9151)	11.7 (.0006)	.172 (.6785)	11.5 (.0007)	1.99 (.1581)	14.8 (.0001)
H_5	3.31 (.0690)	.600 (.4387)	1.76 (.1847)	.334 (.5634)	.040 (.8422)	.061 (.8051)
H_6	1.27 (.2601)	1.29 (.2559)	.173 (.6778)	1.66 (.1981)	.177 (.6738)	5.42 (.0199)
H_7	.684 (.4081)	.232 (.6298)	.312 (.5767)	.108 (.7423)	.023 (.8784)	.093 (.7600)
H_8	.600 (.4385)	.094 (.7586)	.317 (.5731)	.049 (.8242)	.608 (.4355)	.695 (.4044)
H_9	.141 (.7073)	.027 (.8705)	.091 (.7629)	.111 (.7392)	.061 (.8049)	.008 (.9290)
H_{10}	.165 (.6842)	.308 (.5787)	.536 (.4641)	.540 (.4625)	.138 (.7107)	.006 (.9372)
H_{3-4}	.800 (.6702)	12.2 (.0022)	.313 (.8550)	12.5 (.0020)	2.15 (.3419)	15.0 (.0006)
H_{3-5}	4.11 (.2500)	12.8 (.0050)	2.07 (.5574)	12.8 (.0051)	2.19 (.5346)	15.0 (.0018)
H_{3-6}	5.38 (.2508)	14.1 (.0069)	2.25 (.6907)	14.5 (.0059)	2.36 (.6692)	2.4 (.0004)
H_{3-7}	6.06 (.3004)	14.4 (.0135)	2.56 (.7678)	14.6 (.0123)	2.39 (.7934)	2.5 (.0010)
H_{3-8}	6.66 (.3534)	14.5 (.0250)	2.87 (.8244)	14.6 (.0234)	2.99 (.8095)	21.2 (.0017)
H_{3-9}	6.80 (.4498)	14.5 (.0433)	2.97 (.8881)	14.7 (.0396)	3.06 (.8798)	21.2 (.0034)
H_{3-10}	6.97 (.5402)	14.8 (.0634)	3.50 (.8990)	15.3 (.0540)	3.19 (.9216)	21.3 (.0065)
KS	.797 (1.000)	.682 (1.000)	.646 (1.000)	.713 (1.000)	.485 (1.000)	.894 (.0100)
JB	.801 (.6699)	12.3 (.0022)	.314 (.8548)	12.5 (.0019)	2.15 (.3413)	15.0 (.0006)

Panel B: serial correlation.

	DM-US$-5	DM-US$-30	Yen-US$-5	Yen-US$-30	Yen-DM-5	Yen-DM-30
H_3	.863 (.3528)	.874 (.3498)	.160 (.6896)	1.53 (.2158)	.204 (.6513)	.205 (.6505)
H_4	.014 (.9043)	27.0 (.0000)	.21 (.6471)	23.1 (.0000)	3.35 (.0671)	24.8 (.0000)
H_5	5.93 (.0149)	1.44 (.2294)	2.65 (.1038)	.778 (.3779)	.081 (.7754)	.127 (.7219)
H_6	2.70 (.1000)	3.15 (.0760)	.341 (.5592)	4.20 (.0405)	.360 (.5487)	12.3 (.0005)
H_7	1.46 (.2272)	.668 (.4136)	.653 (.4190)	.284 (.5939)	.053 (.8174)	.233 (.6296)
H_8	1.29 (.2557)	.291 (.5893)	.645 (.4219)	.147 (.7016)	1.56 (.2112)	1.70 (.1918)
H_9	.335 (.5626)	.079 (.7789)	.198 (.6563)	.346 (.5562)	.172 (.6785)	.026 (.8728)
H_{10}	.439 (.5077)	1.07 (.2998)	1.35 (.2451)	1.76 (.1849)	.396 (.5292)	.019 (.8907)
H_{3-4}	.968 (.6163)	27.9 (.0000)	.309 (.8568)	26.2 (.0000)	3.65 (.1609)	26.8 (.0000)
H_{3-5}	6.31 (.0973)	28.6 (.0000)	3.60 (.3077)	26.8 (.0000)	3.66 (.3009)	3.0 (.0000)
H_{3-6}	14.6 (.0057)	39.1 (.0000)	5.98 (.2005)	32.6 (.0000)	12.2 (.0157)	3.1 (.0000)
H_{3-7}	18.7 (.0021)	39.2 (.0000)	8.04 (.1538)	35.8 (.0000)	12.2 (.0315)	32.1 (.0000)
H_{3-8}	21.0 (.0018)	4.4 (.0000)	9.98 (.1254)	36.6 (.0000)	19.0 (.0041)	35.4 (.0000)
H_{3-9}	21.4 (.0032)	4.7 (.0000)	11.8 (.1057)	37.2 (.0000)	19.3 (.0074)	36.0 (.0000)
H_{3-10}	22.7 (.0038)	47.3 (.0000)	18.4 (.0185)	38.0 (.0000)	19.8 (.0112)	36.8 (.0000)

Note: We test the normality assumption of the standardized residuals. The volatility is the daily realized volatility computed with five-minute and thirty-minute returns. We report the test statistics and their corresponding p-values in parenthese. We use the same dat as ABDL (2002). In Panel A, we assume that the residuals are i.i.d., while we allow them to be serially correlated in Panel B. In this case, we use a HAC method of Andrews (1991) to estimate the weighting matrix. H_{i-j} is the joint test based on H_k, $i \leq k \leq j$.

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	1.6 (.0011)	8.31 (.0039)	1.2 (.0014)	6.99 (.0082)	8.67 (.0032)	3.58 (.0585)
H_4	7.46 (.0063)	5.91 (.0151)	3.61 (.0572)	3.41 (.0646)	5.17 (.0230)	3.37 (.0664)
H_5	.149 (.6998)	2.92 (.0872)	1.14 (.2858)	2.19 (.1393)	.749 (.3869)	.002 (.9685)
H_6	.000 (.9966)	.035 (.8517)	.732 (.3923)	.997 (.3181)	1.77 (.1831)	.997 (.3181)
H_7	8.94 (.0028)	.827 (.3633)	1.01 (.3163)	1.65 (.1984)	.742 (.3891)	2.66 (.1030)
H_8	.708 (.4000)	.476 (.4903)	.919 (.3377)	1.29 (.2558)	1.02 (.3116)	1.66 (.1974)
H_9	1.46 (.2265)	.563 (.4532)	.007 (.9323)	4.00 (.0455)	8.14 (.0043)	.868 (.3514)
H_{10}	.562 (.4534)	1.05 (.3046)	2.19 (.1390)	4.02 (.0448)	1.06 (.3039)	.417 (.5182)
H_{3-4}	16.9 (.0002)	1.5 (.0052)	16.8 (.0002)	9.63 (.0081)	11.0 (.0040)	7.05 (.0295)
H_{3-5}	17.6 (.0005)	1.5 (.0145)	17.4 (.0006)	9.79 (.0204)	17.3 (.0006)	7.74 (.0518)
H_{3-6}	17.7 (.0014)	15.2 (.0044)	19.2 (.0007)	16.0 (.0030)	17.3 (.0017)	7.83 (.0981)
H_{3-7}	24.0 (.0002)	15.5 (.0084)	26.2 (.0001)	24.5 (.0002)	18.6 (.0023)	1.1 (.0715)
H_{3-8}	25.6 (.0003)	15.5 (.0165)	28.2 (.0001)	25.9 (.0002)	18.9 (.0043)	1.2 (.1157)
H_{3-9}	25.9 (.0005)	17.5 (.0147)	28.3 (.0002)	26.4 (.0004)	25.6 (.0006)	12.1 (.0968)
H_{3-10}	27.0 (.0007)	19.9 (.0108)	28.5 (.0004)	27.9 (.0005)	26.2 (.0010)	12.5 (.1323)

Note: We test the normality assumption of the log of realized volatility. The realized volatility is computed with five-minute and thirty-minute returns. We report the test statistics and their corresponding p-values in parenthese. We use the same data as ABDL (2002). We use a HAC method of Andrews (1991) to estimate the weighting matrix. H_{i-j} is the joint test based on H_k, $i \leq k \leq j$.

Appendix

Proof of Proposition 2.1. Let assume that a random variable x is such that $E[H_i(x)] = 0$. We will prove that this variable follows the Stein equation (2.1) to prove the proposition. Consider a polynomial f of degree n; it can be written as a linear combination of the first n Hermite polynomials, i.e., it exist $(\alpha_0, ..., \alpha_n)$ such as: $f(x) = \sum_{i=0}^n \alpha_i H_i(x)$. Thus, using (2.3) and (2.7), one gets:

$$\begin{aligned}
f'(x) - xf(x) &= \sum_{i=0}^n \alpha_i H_i'(x) - \sum_{i=0}^n \alpha_i (xH_i(x)) \\
&= \sum_{i=1}^n \alpha_i \sqrt{i} H_{i-1}(x) - \alpha_0 x - \sum_{i=1}^n \alpha_i \left(\sqrt{i} H_{i-1}(x) + \sqrt{i+1} H_{i+1}(x) \right) \\
&= -\alpha_0 H_1 - \sum_{i=1}^n \alpha_i \sqrt{i+1} H_{i+1}(x) = - \sum_{i=0}^n \alpha_i \sqrt{i+1} H_{i+1}(x),
\end{aligned}$$

which implies that $E[f'(x) - xf(x)] = 0$. Consider now a continuously differentiable function f whose derivative function is continuous (f is $\mathcal{C}^1$), such that $E|f'(x)| < \infty$ and let ε be a positive real number. $\forall (a, b) \in \mathbb{R}^2$, we have:

$$\int_{-\infty}^{+\infty} (f'(x) - xf(x)) \phi(x)dx = f(a)\phi(a) - f(b)\phi(b) + \int_a^b (f'(x) - xf(x)) \phi(x)dx \qquad \text{(A.1)}$$

where $\phi(x)$ is the density function of a $\mathcal{N}(0, 1)$ random variable. Given that f' is continuous function on the compact set $[a, b]$, the Stone-Weierstrass Theorem implies that:

$$\exists N \in \mathbb{N} \ \forall n \geq N, \exists P_n \in \mathbb{R}^n[X], \ \|f' - P_n\|_\infty < \varepsilon,$$

where $\|f' - P_n\|_\infty = \sup_{x \in [a,b]} |f'(x) - P_n(x)|$. and $\mathbb{R}^n[X]$ is the set of polynomials with degree $m \leq n$. Define the polynomial $Q_n(x)$ by $Q_n(x) = \int_a^x P_n(t)dt + f(a)$. Observe that Q_n is a polynomial of degree $n+1$, $Q_n(a) = f(a)$ and $Q_n(b) \overset{n \to +\infty}{\longrightarrow} f(b)$. Hence, given that $E(P_n(x) - xQ_n(x)) = 0$, one gets by integration by part:

$$0 = Q_n(a)\phi(a) - Q_n(b)\phi(b) + \int_a^b (P_n(x) - xQ_n(x)) \phi(x)dx.$$

Hence, we have:

$$E(f'(x) - xf(x)) = E(f'(x) - xf(x)) - E(P_n(x) - xQ_n(x))$$

$$= (Q_n(b) - f(b)) \phi(b) + \int_a^b \eta(x)\phi(x)dx - \int_a^b x \int_a^x \eta(t)dt\phi(x)dx, \qquad \text{(A.2)}$$

where $\eta(x) = f'(x) - P_n(x)$. The first term of (A.2) tends to 0 when $n \to +\infty$. The second term of (A.2) is bounded by $\varepsilon \int_{-\infty}^{+\infty} \phi(x)dx$ while the third one is bounded by $\varepsilon \int_{-\infty}^{+\infty} x^2\phi(x)dx$. Hence, we get that $E(f'(x) - xf(x)) = 0$, which achieves the proof of the proposition. ∎

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 $E\left[\dfrac{\partial g}{\partial \theta}(z_t, \theta^0)\right] = 0$, 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:

$$\frac{\partial g}{\partial \theta^\top}(z_t, \theta) = \tilde{g}'(u_t(\theta))\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \tilde{g}'(u_t(\theta))u_t(\theta)\frac{1}{\sigma}\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{and } E\left[\frac{\partial g}{\partial \theta}(z_t, \theta^0)\right] = 0 \text{ under (4.15)}.$$

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 (2.7) prove the proposition.∎

Proof of Proposition 5.1. 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 $E[H_3(X)] = Cov(H_3(X), H_4(X)) = 0$, $E[H_4(X)] = \sqrt{\dfrac{3}{2}\dfrac{1}{\nu - 4}}$.
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 (5.1).
The results (5.3), (5.4) and (5.5) are implied by (5.1).∎