Testing Distributional Assumptions: A GMM Approach*

Christian Bontemps† and **Nour Meddahi**‡

October 26, 2007

Abstract

In this paper, we consider testing distributional assumptions. Special cases that we consider are the Pearson's family like the normal, Student, gamma, beta and uniform distributions. The test statistics we consider are based on a set of moment conditions. This set coincides with the first moment conditions derived by Hansen and Scheinkman (1995) when one considers a continuous time model. By testing moment conditions, we treat in detail the parameter uncertainty problem when the considered variable is not observed but depends on estimators of unknown parameters. In particular, we derive moment tests that are robust against parameter uncertainty. We also consider the case where the variable of interest is serially correlated with unknown dependence by adopting a HAC approach for this purpose. This paper extends Bontemps and Meddahi (2005) who considered this approach for the normal case. Finite sample properties of our tests when the variable of interest is a Student are derived through a comprehensive Monte Carlo study. An empirical application to Student-GARCH model is presented.

Keywords: Pearson's distributions; Hansen-Scheinkman moment conditions; parameter uncertainty; serial correlation; HAC.

JEL codes: C12, C15.

---

*The first author acknowledges financial support from ANR, the second author from CRSNG, IFM2, and MITACS.

†Toulouse School of Economics & IDEI, 21 allée de Brienne, 31000 Toulouse, France. E-mail: bontemps@cict.fr.

‡Tanaka Business School, Imperial College London, Exhibition Road London SW7 2AZ. E-mail: n.meddahi@imperial.ac.uk

1 Introduction

Let x be a continuous random variable with a density function denoted by $q(.)$. Then, an integration by part leads to

$$E[\psi'(x) + \psi(x)(\log q)'(x)] = 0, \tag{1.1}$$

where the function $\psi(\cdot)$ follows some regularity conditions and constraints on the boundary support of x discussed later on. It turns out that we show than any zero mean smooth moment condition can be written as Eq. (1.1). Equation (1.1) is clearly important for modeling, estimation and specification testing purposes. The main goal of the paper is the use of Eq. (1.1) and the generalized method of moments (GMM) of Hansen (1982) for testing distributional assumptions. The paper extends Bontemps and Meddahi (2005) who used the same approach for testing normality. In this case, when one wants to test that x is a standard normal random variable, one has $\log(q)'(x) = -x$, and therefore Eq. (1.1) becomes $E[\psi'(x) - \psi(x)x] = 0$, which is known as the Stein equation (Stein, 1972).

Karl Pearson introduced a century ago in several papers the so-called Pearson class of distributions, where $(\log q)'(\cdot)$ is the ratio of an affine function over a quadratic one. This class contains as special cases the Gaussian, Student, Gamma, Beta, and the uniform distributions. By using (1.1) with polynomial test functions $\psi(\cdot)$, K. Pearson derived the moments of these distributions. In order to estimate the distributions parameters, K. Pearson also introduced the method of moments by matching some empirical moments with their theoretical counterpart, the number of moments being the number of unknown parameters. More recently, Cobb, Koppstein and Chen (1983) extended Pearson's modeling approach to generate multimodal distributions by taking a more general form of $(\log q)'(\cdot)$ than K. Pearson.

Wong (1964) made a connection between the Pearson distributions and diffusions processes, i.e., he provided stationary continuous time modes for which the marginal density is a Pearson distribution. This connection was used by Hansen and Scheinkman (1995), Aït-Sahalia (1996) and by Conley, Hansen, Luttmer and Scheinkman (1997), in order to model the short term interest rate whose marginal distribution are among the class of the generalized Pearson's distributions of Cobb, Koppstein and Chen (1983). It is worth noting that Hansen and Scheinkman (1995) derived two classes of moment conditions that characterize a diffusion process: one class related to its marginal distribution and a second one related to its conditional distribution. Importantly, the Hansen and Scheinkman (1995) first class of moments conditions coincide with one generated by Eq. (1.1).

The GMM is convenient for handling two potential problems: the serial correlation in the data and the parameter uncertainty when one uses estimated parameters. Two important examples of the recent development of the financial literature emphasize the importance of developing distributional specification test procedures that are valid in the case of a serial correlation in the data. The first one is modeling continuous time Markov models, particularly the short term interest rate. It turns out that the specification of a stationary scalar diffusion process through the drift and the diffusion terms characterizes its marginal distribution. Consequently, a leading specification test approach in the literature was developed by Aït-Sahalia (1996) and by Conley, Hansen, Luttmer and Scheinkman (1997) by testing whether the marginal distribution of the data coincides with the theoretical one implied by the specification of the scalar diffusion. Aït-Sahalia (1996) compared the nonparametric estimator of the density function with its theoretical counterpart while Conley, Hansen, Luttmer and Scheinkman used the moment conditions (1.1).

The evaluation of density forecasts approach developed by Diebold, Gunter and Tay (1998) in the univariate case and by Diebold, Hahn and Tay (1999) in the multivariate case also highlighted the importance of testing distributional assumption for serially correlated data. This evaluation is done by testing that some variables are independent and identically distributed (i.i.d.) and follow a uniform distribution on $[0, 1]$. However, the non independence and the non uniformness of these data mean different things about the specification of the model. Therefore, when one rejects the joint hypothesis, i.i.d. and uniform, one wants to know which assumptions are wrong (both or a unique). This is why Diebold, Tay and Wallis (1999) explicitly asked for the development of testing uniform distribution in the case of serial correlation by arguing that traditional tests (e.g., Kolmogorov-Smirnov) are valid under the i.i.d. assumption. Of course, one can use the bootstrap to get a correct statistical procedure as did Corradi and Swanson (2002).

The GMM is well suited for handling the serial correlation in the data by using the Heteroskedastic-Autocorrelation-Consistent (HAC) method of Newey and West (1987) and Andrews (1991). Using a HAC procedure in testing marginal distributions was already adopted by Richardson and Smith (1993), Bai and Ng (2005) and Bontemps and Meddahi (2005) for testing normality, and by Aït-Sahalia (1996), Conley et al. (1997), and Corradi and Swanson (2002) for testing marginal distributions of nonlinear scalar diffusions.

In general, the test statistics will involve an unknown parameter that should be estimated in order to get a feasible test statistic. This is the case if the true distribution of x depends on an unknown parameter, as well as if the variable x is not observed but is a function of the observable variables and an unknown parameter, like the residuals in a regression model. The dependence of the feasible test statistic on an estimated parameter has to be taken into account, given that in general the asymptotic distribution of the feasible test statistic will not equal one of the unfeasible test statistic. This problem leads Lilliefors (1967) to tabulate the Kolmogorov-Smirnov test statistic for testing normality when one estimates the mean and the variance of the distribution. In the linear homoskedastic model, White and MacDonald (1980) stated that various normality tests are robust against parameter uncertainty, particularly in tests based on moments that used standardized residuals. Dufour, Farhat, Gardiol and Khalaf (1998) developed Monte Carlo tests to take into account parameter uncertainty in the linear homoskedastic regression model in finite samples with normal errors. More recently, several solutions have been proposed in the literature for general distribution: Bai (2003) and Duan (2003) proposed transformations (as in Wooldridge, 1990) of their test statistics that are robust against parameter uncertainty; Thompson (2002) proposed upper bound critical values for his tests; Hong and Li (2002) used separate inference procedure by splitting the sample; while Corradi and Swanson (2002) used the bootstrap.

It turns out that the GMM setting is well suited for incorporating parameter uncertainty in testing procedures by using Newey (1985), Tauchen (1985), Gallant (1987), Gallant and White (1988), and Wooldridge (1990). Bontemps and Meddahi (2005) followed this approach for testing normality. In particular, in the context of a regression model (linear, nonlinear, dynamic), they characterized the test functions $\psi(\cdot)$ that are robust to the parameter uncertainty problem, i.e., the asymptotic distribution of the feasible test statistic based on an estimated parameter is identical to that of the test statistic based on the true (unknown) parameter. The Hermite polynomials are special examples of these robust functions, a result proved by Kiefer and Salmon (1983) for a nonlinear homoskedastic regression estimated by the maximum likelihood method; as pointed out in Bontemps and Meddahi (2005), Jarque and Bera (1980) is a special case of Kiefer and Salmon (1993).

It is well known that one gets a standard normal variable, $\mathcal{N}(0, 1)$, if one considers the variable y defined as $y \equiv \Phi^{-1}(Q(x))$, where $Q(\cdot)$ and $\Phi(\cdot)$ are the cumulative distributions functions of x and standard normal variable. Therefore, given that tests for normality are already studied in details, it is natural to do tests based on y. For instance, Diebold, Gunter and Tay (1998) and Lejeune (2002) followed this approach. A natural question is the usefulness of testing (1.1) on the variable x instead of the Stein equation or any normality test on the variable y. We can give actually several reasons. First, when one rejects the normality of y, one does not know how to modify the distribution of x to get a correct specification. For instance, under misspecification, one may have a correct specification of the mean of x but gets a nonzero mean for y. In other words, observing that the mean of y is nonzero does not imply that this is case for the mean of x. Note however that some characteristics of x remains in y; for instance if the true distribution of x is symmetric, it is also the case for one of y even if the function $Q(\cdot)$ is not the correct distribution function of x. Second, handling the parameter uncertainty problem may be easiest with x than y. Given that the function $Q(\cdot)$ will depend in general on the unknown parameter, tests based on y will be more difficult than those based on x. For instance, while one has the function $Q(\cdot)$, at least numerically, the distribution of the feasible test statistic will involve the derivative of $Q(\cdot)$ with respect to the parameter, which one does not get easily, even numerically. In addition, the characterization of the robust test functions $\psi(\cdot)$ based the on the tests on y will involve more conditions than ones based on x. It is worth noting that Bontemps and Meddahi (2005) characterized the robust functions in the case of regression models which does not include the nonlinear transform function $\Phi^{-1}(Q(\cdot))$. Finally, an important limitation of the transform method is that one can not do it for non continuous random variables, like discrete ones (Binomial, Poisson), or mixed ones (for instance $x = u$ if $u > 0$ and $x = 0$ if $u \leq 0$, where u is a continuous variable on the real line). It turns out that similar moment conditions like Eq. (1.1) hold in these cases. Similarly, if x is a multivariate random variable, it is difficult to transform it on a multivariate normal distribution. Interestingly, one can characterize an equation like Eq. (1.1) in the multivariate case by using Hansen and Scheinkman (1995) and Chen, Hansen and Scheinkman (2000). Note that Stein (1972) and Amemiya (1977) give this equation in the normal multivariate case. The treatment of the non continuous and multivariate cases is beyond the scope of the paper and is left for future research.

The rest of the paper is organized as follows. Section 2 introduces and studies the moment condition of interest. The connection to Pearson's family of distributions and to Hansen and Scheinkman (1995) is provided. The third section deals with the asymptotic distribution of the tests statistics when the parameter uncertainty problem does not hold. This problem is studied in Section 4. Section 5 provides Monte Carlo simulations to assess the performance of the tests, an empirical application is given in Section 6, while Section 7 concludes. All the proofs are provided in the Appendix.

2 Test functions

2.1 Moment conditions

Let x be a stationary random variable with density function denoted by $q(.)$. We assume that the support of x is (l, r), where l and r may be finite or not, and the function $q(\cdot)$ is differentiable on (l, r). Consider a differentiable function $\psi(.)$ such that its derivative function, denoted by $\psi'(.)$, is integrable with respect to the density function $q(.)$. Then, an integration by part leads to:

$$E[\psi'(x)] = [\psi(x)q(x)]_l^r - E[\psi(x)\frac{q'(x)}{q(x)}].$$

Hence, we get that

$$E[\psi'(x) + \psi(x)(\log q)'(x)] = 0, \tag{2.1}$$

under the following assumption (that we comment in few subsections):
Assumption A1: $\lim_{x\to l} \psi(x)q(x) = 0$ and $\lim_{x\to r} \psi(x)q(x) = 0$.

The general moment condition (2.1) gives a class of test functions that a random variable with a density function $q(\cdot)$ should follow. It will be the basis of our testing approach. It will be then natural to chose some specific (i.e. optimal) functions $\psi(\cdot)$ for some particular purposes (e.g., parameter uncertainty, power, etc.). Of course, assumption A1 should hold for the function $\psi(\cdot)$. This is not however a restrictive assumption when one knows the function $q(\cdot)$ (up to unknown parameters). For instance, in the case of a normal distribution, assumption A1 holds for any polynomial function and for any function dominated by $\exp(-x^2/2)$, (i.e., $q(x) = o(\exp(-x^2/2))$ when $|x|$ is large). We will study this assumption in the context of the Pearson's distributions in the next section.

As pointed out in the introduction, Karl Pearson used (2.1) to introduce his famous class of distributions as well as for deriving moment based estimator of the parameters. However, we did not find in the literature a systematic use of (2.1) for any distribution. However, it is implicitly suggested in Hansen (2001) in the case of scalar diffusion processes. In addition, Chen, Hansen and Scheinkman (2000) explicitly used this equality in the multivariate continuous time processes (see the equation that follows their Eq. (3), page 14).[1]

The moment condition (2.1) is written marginally; however it holds also conditionally on some variable z, i.e., if one assumes that the conditional distribution of x given z is $q(x, z)$, then one has

$$E\left[\frac{\partial\psi(x,z)}{\partial x} + \frac{\psi(x,z)}{q(x,z)}\frac{\partial q(x,z)}{\partial x} \mid z\right] = 0,$$

while feasible test statistics will be based on

$$E\left[g(z)\left(\frac{\partial\psi(x,z)}{\partial x} + \frac{\psi(x,z)}{q(x,z)}\frac{\partial q(x,z)}{\partial x}\right)\right] = 0,$$

where $g(z)$ is a square-integrable function of z.

In many cases, one has moment restrictions like

$$Em(x) = 0. \tag{2.2}$$

This is the case either because one has an economic model that implies (2.2) or because one computes explicit moments implied by the density function $q(\cdot)$. It is therefore of interest to characterize the relationship between the moment conditions (2.1) and (2.2). This is the purpose of the following proposition:

Proposition 2.1 *Let $m(\cdot)$ be a continuous and integrable function with respect to the density function $q(\cdot)$. Then a solution $\psi(\cdot)$ of the ordinary differential equation*

$$m(x) = \psi'(x) + \psi(x)(\log q)'(x). \tag{2.3}$$

[1]Strictly speaking, these authors did not use the fact that the variable of interest is a continuous time process. In a private discussion, Lars Hansen confirmed to us that he knew that (2.1) holds for any distribution. In addition, a reader of Eq. (3) in Chen, Hansen and Scheinkman (2000) may not see the direct connection with (2.1) because additional variables appear (namely a matrix $\Sigma(x)$ and a second function $\phi(x)$); however it is exactly (2.1) and therefore corresponds to the multivariate extension of (2.1); we are currently studying this extension to test multivariate distributions.

is given by

$$\psi(x) = \frac{1}{q(x)} \int_l^x m(u)q(u)du. \tag{2.4}$$

In addition, (2.2) holds if and only if assumption A1 holds for $\psi(\cdot)$.

Some remarks are in order. First, the connection in (2.3) holds without the expectation operator. Consequently, the statistical properties (size, power) of (2.1) coincide with those of (2.2). Second, the function $m(\cdot)$ should be continuous, otherwise the function $\psi(\cdot)$ defined in (2.4) is not differentiable. The continuity assumption of $m(\cdot)$ precludes quantile moment restrictions. Third, given that any moment condition (2.2) (where $m(\cdot)$ is continuous) can be written as (2.1), the informational content of the class of moment conditions (2.1) is huge. In particular, it encompasses the score function and therefore by considering the all class for estimation purpose, one gets an efficient estimator. It also encompasses the so-called information-matrix test moment conditions (White, 1982) as well as their generalization, i.e., the Bartlett identities tests (Chesher, Dhaene, Gouriéroux and Scaillet, 1999).

2.2 Transformed distributions

In many cases, it is convenient to transform the variable of interest in order to get a variable whose distribution is simple, e.g. for testing purpose. For instance, in their density forecast analysis, Diebold, Gunter and Tay (1998) transform the variable of interest onto a uniform one. Consequently, it is interesting to characterize the relationship between the classes of test functions associated with each random variable.

Proposition 2.2 *Let X and Y be two random variables such that $Y = G(X)$ where $G(\cdot)$ is a monotonic and one-to-one differentiable function. We denote by $q_X(\cdot)$ and $q_Y(\cdot)$ the density functions of X and Y and by (l_X, r_X) and (l_Y, r_Y) their supports. For any function $\psi_X(\cdot)$, define the function $\psi_Y(\cdot)$ by*

$$\psi_Y(y) = \left(G' \circ G^{-1}(y)\right)\left(\psi_X \circ G^{-1}(y)\right).$$

Then $\forall x,\ y$, with $y = G(x)$, we have

$$\psi_X'(x) + \psi_X(x)(\log q_X)'(x) = \psi_Y'(y) + \psi_Y(y)(\log q_Y)'(y). \tag{2.5}$$

In addition, we have

$$\lim_{x \to l_X} \psi_X(x)q_X(x) = \lim_{x \to r_X} \psi_X(x)q_X(x) = 0 \Longleftrightarrow \lim_{y \to l_Y} \psi_Y(y)q_Y(y) = \lim_{y \to r_Y} \psi_Y(y)q_Y(y) = 0. \tag{2.6}$$

Again, (2.5) holds without the expectation operator and therefore the statistical properties of tests based on the variable X coincide with those based on Y. Meanwhile, (2.6) means that assumption A1 holds for ψ_X if and only if it holds for ψ_Y. We will use this connection later when we study the parameter uncertainty problem.

2.3 Pearson's distributions and their generalizations

2.3.1 The Pearson family of distributions

At the end of the nineteenth century, Karl Pearson introduced his famous family distribution that extends the classical normal distribution. If a distribution with a density function $q(\cdot)$

on (l, r) belongs to the Pearson family, then $q'(\cdot)/q(\cdot)$ equals the ratio of two polynomials $A(\cdot)$ and $B(\cdot)$, where $A(\cdot)$ is affine and $B(\cdot)$ is quadratic and positive on (l, r), i.e.,

$$\frac{q'(x)}{q(x)} = \frac{A(x)}{B(x)} = \frac{-(x+a)}{c_0 + c_1 x + c_2 x^2}. \tag{2.7}$$

The Pearson's class of distributions include as special examples the Normal, Student, Gamma, Beta, and Uniform distributions; for more details, see Johnson, Kotz and Balakrishnan (1994). A major motivation for introducing this family is the simple way they can be estimated. By using (2.1) for $\psi_j(x) = x^j B(x)$, $i = 1, 2, ...,$, one gets this recursive equations

$$(c_2(j+2) - 1)E[X^{j+1}] = (a - c_1(j+1))E[X^j] - c_0 j E[X^{j-1}].$$

Pearson solved this system for $j = 1, ..., 4$, i.e., he derived $\theta = (a, c_0, c_1, c_2)^\top$ in terms of $E[X^j]$ and then provided an estimator for θ by using the empirical counterpart of $E[X^j]$ (under the assumption that these moments exist). This was the introduction of the method of moments; see Bera and Bilias (2002) for a historical review.

2.3.2 Orthonormal polynomials

For a given distribution, one can derive orthonormal polynomials whenever the corresponding squared moments are finite. A simple approach is based on the use of the Gram-Schmidt method. For the Pearson's family, these polynomials are actually genuine, i.e., they can be derived more directly from the density function $q(\cdot)$. More precisely, define the sequence of functions $\tilde{P}_n$ by the so-called Rodrigue's formula

$$\tilde{P}_n = \frac{1}{q(x)} \left[B^n(x) q(x)\right]^{(n)}. \tag{2.8}$$

where $f^{(n)}(\cdot)$ denotes the n-th derivative function of any function $f(\cdot)$. Then, one can show (see Chihara (1978)) that $\tilde{P}_n$ is a polynomial of degree n. In addition, its expectation equals zero for $n \geq 1$ whenever it exists. Finally, two different polynomials $\tilde{P}_n$ and $\tilde{P}_m$ are orthogonal whenever their variance is finite. In other words, the sequence of polynomials is orthogonal when their variances exist. The problem of infinite variance holds for distributions like the Student.

When this problem does not hold (as in the normal, gamma, beta or uniform case), then sequence of polynomials is infinite and indeed dense in $L^2(]l, r[)$, i.e., any square-integrable function may be expanded onto the polynomials $\tilde{P}_n$, $n = 0, 1, ...$; in this case, the density function of a random variable x equals $q(\cdot)$ if and only if

$$\forall n \geq 1, \;\; E[\tilde{P}_n(x)] = 0.$$

For a formal proof, see Gallant (1980, Theorem 3, page 192). This result means that for statistical inference purposes, in particular testing, one could use only orthogonal polynomials.

The polynomials $\tilde{P}_n(\cdot)$ are orthogonal but not orthonormal. The orthonormal ones are

$$P_n = \alpha_n \tilde{P}_n(x) \tag{2.9}$$

where

$$\alpha_n = \frac{(-1)^n}{\sqrt{(-1)^n n! d_n \int_l^r B^n(x) q(x) dx}}, \quad d_n = \prod_{k=0}^{n-1} \left(-1 + (n+k+1)c_2\right),$$

and they satisfy the recurrence relation

$$n \geq 1,\ P_{n+1}(x) = -\frac{1}{a_n}\left((b_n - x)P_n(x) + a_{n-1}P_{n-1}(x)\right),\ P_0(x) = P_{-1}(x) = 1, \tag{2.10}$$

where

$$a_n = \frac{\alpha_n d_n}{\alpha_{n+1} d_{n+1}},\ b_n = n\mu_n - (n+1)\mu_{n+1},\ \mu_n = \frac{-a + nc_1}{-1 + 2nc_2}.$$

2.3.3 Examples

We now consider some of the most popular examples among the Pearson's family, i.e., the normal, student, gamma, beta, and uniform distributions.

1) The Normal distribution. When $X \sim \mathcal{N}(\mu, \sigma^2)$, one has

$$q(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right),\ \frac{\partial(\log q)}{\partial x} = -\frac{x-\mu}{\sigma^2}.$$

In this case, Eq. (3.2) is known as the as the Stein equation and has been used by Bontemps and Meddahi (2005) to test normality. The corresponding orthonormal polynomials are $P_n(x, \mu, \sigma) = H_n\left((x-\mu)/\sigma\right)$, where $H_n(\cdot)$ is the normalized Hermite polynomial of degree n, i.e.,

$$n \geq 2,\ H_n(x) = \frac{1}{\sqrt{i}}(xH_{n-1}(x) - \sqrt{n-1}H_{n-2}(x)),\ H_0(x) = 1,\ H_1(x) = x.$$

2) The Student distribution. When the variable X follows a Student $T(\nu)$, with $\nu > 0$, one has

$$q(x, \nu) = \nu^{-1/2}\left(B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right)^{-1}\left[1 + \frac{x^2}{\nu}\right]^{-(\nu+1)/2},\ \frac{\partial(\log q)}{\partial x}(x, \nu) = -(\nu+1)\frac{x}{\nu + x^2},$$

where $B(\cdot,\cdot)$ denotes the Beta function, connected to the Gamma function $\Gamma(\cdot)$ by

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},\ \Gamma(\alpha) = \int_0^\infty \exp(-u)u^{\alpha-1}du,\ \alpha > 0.$$

Equation (2.10) defines the Romanovski polynomials. However, as mentioned above, higher order moments do not exist for a Student distribution. Consequently, only those with finite variance, i.e., $n < \nu/2$, can be called orthonormal polynomials. The remaining orthonormal functions are among the set of hypergeometric functions and actually they are a continuum of functions.

The Romanovski polynomials are given by

$$R_{n+1}(x, \nu) = \sqrt{\frac{(\nu - 2n)(\nu - 2n - 2)}{(n+1)\nu(\nu - n)}}xR_n(x, \nu) - \sqrt{\frac{n(\nu - n + 1)(\nu - 2n - 2)}{(n+1)(\nu - n)(\nu - 2n + 2)}}R_{n-1}(x, \nu)$$

The first three polynomials are given by

$$R_1(x, \nu) = \sqrt{\frac{\nu - 2}{\nu}}x,\ R_2(x, \nu) = \sqrt{\frac{\nu - 4}{2(\nu - 1)}}\left(\frac{\nu - 2}{\nu}x^2 - 1\right),$$

$$R_3(x, \nu) = \sqrt{\frac{(\nu - 2)(\nu - 6)}{6\nu(\nu - 1)}}\left(\frac{\nu - 4}{\nu}x^3 - 3x\right).$$

3) The Gamma distribution. When X follows a gamma (α, β, γ), for $\alpha > 0,\ \beta > 0;\ x > \gamma$, one has

$$q(x, \alpha, \beta, \gamma) = \frac{(x-\gamma)^{\alpha-1} \exp\left(-\frac{(x-\gamma)}{\beta}\right)}{\beta^\alpha \Gamma(\alpha)}, \quad \frac{\partial(\log q)}{\partial x}(x, \alpha, \beta, \gamma) = \frac{\alpha - 1 - \frac{(x-\gamma)}{\beta}}{x-\gamma}.$$

For this case, all the moments are finite and the set of orthonormal polynomials characterize the distribution. Equation (2.10) which defines the Laguerre polynomials becomes

$$\begin{aligned} P_n(x, \alpha, \beta, \gamma) = \frac{1}{\sqrt{n(\alpha+n-1)}} \Bigg(\left(\frac{x-\gamma}{\beta} - \alpha - 2n + 2 \right) P_{n-1}(x, \alpha, \beta, \gamma) \\ - \sqrt{(n-1)(\alpha+n-2)} P_{n-2}(x, \alpha, \beta, \gamma) \Bigg), \end{aligned}$$

while the first and second polynomials are given by

$$\begin{aligned} P_1(x, \alpha, \beta, \gamma) &= \frac{1}{\sqrt{\alpha}} \left(\frac{x-\gamma}{\beta} - \alpha \right), \\ P_2(x, \alpha, \beta, \gamma) &= \frac{1}{\sqrt{2\alpha(\alpha+1)}} \left(\left(\frac{x-\gamma}{\beta} \right)^2 - 2(\alpha+1) \left(\frac{x-\gamma}{\beta} \right) + \alpha(1+\alpha) \right). \end{aligned}$$

4) The Beta distribution. When X follows a standard beta distribution beta (α, β) with $0 \leq x \leq 1,\ \alpha > 0,\ \beta > 0$, one has

$$q(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad \frac{\partial(\log q)}{\partial x}(x, \alpha, \beta) = \frac{(-\alpha - \beta + 2)x + \alpha - 1}{x(1-x)}.$$

Like the normal and gamma cases, all the moments of a beta distribution are finite and the set of orthonormal polynomials characterize the distribution. Equation (2.10) defines the Jacobi polynomials. The corresponding coefficient a_n and b_n are

$$\begin{aligned} a_n &= \sqrt{\frac{(n+1)(\alpha+\beta+n-1)(\alpha+n)(\beta+n)}{(\alpha+\beta+2n)^2(\alpha+\beta+2n-1)(\alpha+\beta+2n+1)}}, \\ b_n &= \frac{\alpha^2 + \alpha\beta + 2(\alpha+\beta)n + 2n^2 - 2\alpha - 2n}{(\alpha+\beta+2n)(\alpha+\beta+2n-2)}, \end{aligned}$$

while the first and second polynomials are given by

$$\begin{aligned} P_1(x, \alpha, \beta) &= \sqrt{\frac{\alpha+\beta+1}{\alpha\beta}} \left((\alpha+\beta)x - \alpha \right) \\ P_2(x, \alpha, \beta) &= \frac{\Gamma(\alpha+\beta+3)}{\Gamma(\alpha+\beta)} \sqrt{\frac{(\alpha+\beta)(\alpha+\beta+3)}{2(\alpha\beta)(\alpha+1)(\beta+1)}} \left(x^2 - 2\frac{\alpha+1}{\alpha+\beta+2}x + \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta+2)} \right) \end{aligned}$$

5) The Uniform distribution. When X follows a uniform distribution on $[0, 1]$, the density function equals one for $x \in [0, 1]$ and $\log(q)'(x)$ equals zero. Therefore, Eq. (2.7) does not hold for a uniform distribution. Actually, the uniform distribution is within the Pearson's family as the limit of a beta(α, β) when both α and β goes to one. The orthonormal polynomials are the Legendre polynomials and they characterize the distribution. Equation (2.10) becomes

$$P_n(x) = \frac{\sqrt{2n+1}}{n}\left(\sqrt{2n-1}(2x-1)P_{n-1}(x) - \frac{n-1}{\sqrt{2n-3}}P_{n-2}(x)\right),$$

while the first three Legendre polynomials are given by

$$P_1(x) = \sqrt{3}(2x-1),\ P_2(x) = \sqrt{5}(6x^2-6x+1),\ P_3(x) = \sqrt{7}(20x^3-30x^2+12x-1).$$

2.3.4 Generalized Person's distributions

One limitation of the Pearson's distributions is the shape of their density functions: they can not have more than one mode. For this reason, Cobb, Koppstein and Chen (1983) extended Pearson's class of distributions by allowing $A(\cdot)$ in (2.7) to be a polynomial of degree higher than one and, hence, generated multimodal distributions. This extension has been exploited by Hansen and Scheinkman (1995), Aït-Sahalia (1996) and by Conley, Hansen, Luttmer and Scheinkman (1997), in order to model the short term interest rate whose marginal distribution looks like a bimodal distribution. These authors strongly rejected Pearson's unimodal distributions.

2.4 Stationary distribution of scalar diffusions

As we pointed out in the introduction, Wong (1964) made a connection between the Pearson distributions and diffusions processes, i.e., he provided stationary continuous time modes for which the marginal density is a Pearson distribution. This connection was used by Hansen and Scheinkman (1995), Aït-Sahalia (1996) and by Conley, Hansen, Luttmer and Scheinkman (1997), in order to model the short term interest rate whose marginal distribution are among the class of the generalized Pearson's distributions of Cobb, Koppstein and Chen (1983). In this subsection, we recap some results in Hansen and Scheinkman (1995) to show the interpretation of (2.1) in the diffusion case.

Assume that the random variable x_t is a stationary scalar diffusion process and characterized by the stochastic differential equation

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \tag{2.11}$$

where W_t is a scalar Brownian motion. The marginal distribution $q(.)$ is related to the functions $\mu(.)$ and $\sigma(.)$ by the following relationship

$$q(x) = K\sigma^{-2}(x)\exp\left(\int_z^x \frac{2\mu(u)}{\sigma^2(u)}du\right), \tag{2.12}$$

where z is a real number in (l, r) and K is a scale parameter such as the density integral equals one; see Aït-Sahalia, Hansen and Scheinkman (2003) for a review of all the properties of diffusion processes we consider in this paper.

Hansen and Scheinkman (1995) provided 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

$$E[\mathcal{A}g(x_t)] = 0, \tag{2.13}$$

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 (2.11), i.e.,

$$\mathcal{A}g(x) = \mu(x)g'(x) + \frac{\sigma^2(x)}{2}g''(x). \tag{2.14}$$

From (2.12), one gets easily

$$\frac{q'(x)}{q(x)} = \frac{2\mu(x) - (\sigma^2)'(x)}{\sigma^2(x)}. \tag{2.15}$$

As a consequence, by using (2.15) in (2.13), one gets after some manipulations

$$E[(g\sigma^2)'(x) + (\log q)'(x)(g\sigma^2)(x)] = 0, \tag{2.16}$$

which is exactly the general test function (2.1) applied to the function $\psi = (g\sigma^2)'$. Again, Hansen and Scheinkman (1995) assumed that the variable x_t is Markovian to derive (2.13) (and (2.16)) while we did not for deriving (2.1).

3 Asymptotic distribution of the test statistics

In this section, we discuss the asymptotic distribution of the test statistics based on (2.1). However, the study of the parameter uncertainty problem is postponed to the next section.

Consider a sample $x_1, ..., x_T$, of the variable of interest denoted by X. The observations may be independent or not. Let $\psi_1(\cdot)$, ..., $\psi_p(\cdot)$, be p differentiable functions such that assumption A1 holds for $\psi_i(\cdot)$. Let us denote $m(x)$ as the vector whose components are $\psi_i(x)' + \psi_i(x)(\log q)'(x)$, $i = 1, 2, ..., p$. Thus, by (2.1), we have

$$E[m(x)] = 0.$$

Throughout the paper, we assume the matrix Σ defined by

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

is finite and positive definite. In the context of time series, this assumption ruled out some long memory processes; see Bontemps and Meddahi (2005). Under some regularity conditions, we know that

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

while

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right)^\top \Sigma^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right) \sim \chi^2(p). \tag{3.2}$$

Feasibility test procedure needs the matrix Σ or a consistent estimator of it.

3.1 The cross-sectional case

In the context of cross-sectional observations where the observations are assumed to be independent and identically distributed (i.i.d.), we have

$$\Sigma = Var[m(x)] = E[m(x)m(x)^\top]. \tag{3.3}$$

Two cases may arise. One can explicitly compute the matrix Σ and, hence, one can use the test statistic (3.2). This is the case for Pearson's distributions discussed above when the component of $m(\cdot)$ are indeed the orthonormal polynomials associated to the distribution. In this case, we have $E[P_i(x)] = 0$ and $E[P_i(x)P_j(x)] = \delta_{i,j}$ where $\delta_{\cdot,\cdot}$ is the Kronecker symbol. Consequently, the matrix Σ will be the identity matrix, implying that the univariate test statistics based on $E[P_i(x)] = 0$ are asymptotically independent.

In the second case, computing Σ explicitly is not possible (or difficult). One can therefore use any consistent estimator $\hat{\Sigma}_T$ of Σ like

$$\hat{\Sigma}_T = \frac{1}{T}\sum_{t=1}^{T} m(x_t)m(x_t)^\top.$$

In this case, one can use the following test statistic

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

3.2 The serial correlation case

Assume now that the observations are correlated. Then without additional assumptions on the dependence structure, one can not explicitly compute the matrix Σ. For instance, knowing that the marginal distribution of a process is normal does not imply that its conditional distribution is normal and therefore one has not information about $E[m(x_t)m(x_{t-h})]$ in (3.1) for $h \neq 0$. When one does not have information about the dependence in the process x_t, one has to estimate Σ. 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. We will follow this approach as did Richardson and Smith (1993), Bai and Ng (2005) and Bontemps and Meddahi (2005) for testing normality, and by Aït-Sahalia (1996), Conley, Hansen, Luttmer and Scheinkman (1997), and Corradi and Swanson (2002) for testing marginal distributions of nonlinear scalar diffusion processes.

However, making additional assumption on the dependence structure may lead to simple estimates of the Σ. This is the case when one considers some Pearson's distribution and assumes that the components of $m(\cdot)$ are orthonormal polynomials. In the serial correlation case, we still have $E[P_n(x_t)P_m(x_t)] = 0$. However, without additional assumptions, one does not have

$$E[P_n(x_t)P_m(x_{t-h})] = 0,\ n \neq m,\ h \neq 0. \tag{3.4}$$

Several scalar diffusion processes have as the stationary distribution the normal $\mathcal{N}(0,1)$ distribution but (3.4) does not hold because it is related to the conditional distribution of the process $\{x_t\}$. In contrast, by assuming that the conditional distribution of x_t given its past values is Gaussian, one gets (3.4). For instance, when one assumes that the process x_t is a normal autoregressive process of order one, AR(1), that is

$$x_t = \gamma x_{t-1} + \sqrt{1-\gamma^2}\,\varepsilon_t, \quad \varepsilon_t \text{ is i.i.d. and } \sim \mathcal{N}(0,1), \quad \text{and} \quad |\gamma| < 1. \tag{3.5}$$

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

$$E[H_i(x_{t+1})|x_\tau, \tau \leq t] = \gamma^i H_i(x_t). \tag{3.6}$$

In this case, one has

$$\Sigma_{ij} = \sum_{h=-\infty}^{+\infty} E[H_i(x_t)H_j(x_{t-h})] = \frac{1+\gamma^i}{1-\gamma^i}\delta_{ij}. \tag{3.7}$$

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

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

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 Bontemps and Meddahi (2005) assessed this issue.

It is worth noting that Σ is also diagonal for other time series processes, in particular for scalar diffusions whose marginal distribution is among the Pearson's class and the drift is affine. This is the case of the square-root process of Cox, Ingersoll and Ross (1984) whose marginal distribution is gamma and quite popular in modeling the short term interest rate. This is also the case for the Jacobi diffusion (Karlin and Taylor (1975), page 335) whose marginal distribution is beta; see Gouriéroux and Jasiak (2006) for financial applications.

4 Parameter uncertainty

In general, the density function involved in (2.1) depends on unknown parameters. Moreover, the variable x may be not observable but can depend on unknown parameters like, e.g., residuals in a regression model. Therefore, one has to first estimate these parameters before implementing any distributional test procedure. However, it is well known that the asymptotic distribution of the feasible test statistic based on (3.2) is, in general, different from the unfeasible one that uses the true (unknown) parameter. The main purpose of this section is to derive sufficient conditions in order to avoid the parameter uncertainty problem, i.e., making the asymptotic distribution of the feasible and unfeasible test statistics coincide.

In this section, we assume that the probability density function depends on a parameter β and we denote by β^0 the true unknown value. In addition, we assume that the variable x_t is not necessarily observable. However, x_t is related to the observable variables, denoted by z_t, by the relationship

$$x_t = h(z_t, \beta^0, \theta^0), \tag{4.1}$$

where the function $h(\cdot)$ is a one-to-one known function (for a given vector $(\beta^\top, \theta^\top)^\top$) and θ^0 is an unknown parameter different from β^0. $h^{-1}(\cdot, \beta, \theta)$ denotes the inverse function of $h(\cdot, \beta, \theta)$. Observe that the function $h(\cdot)$ may depend on other variables like explanatory variables. For instance, a leading example is a non-linear regression model

$$x_t = \frac{z_t - m(\tilde{z}_t, \beta^0, \theta^0)}{\sigma(\tilde{z}_t, \beta^0, \theta^0)},$$

where $\tilde{z}_t$ is a vector of explanatory variables.

Our goal is to test:

$$H_0 : \text{The probability density function of } x_t \text{ is } q(x, \beta^0). \tag{4.2}$$

The test will be based on the moment condition (2.1). We allow $\psi(\cdot)$ to depend on both β^0 and θ^0, which leads to

$$E\left[\frac{\partial \psi}{\partial x}(x, \beta^0, \theta^0) + \psi(x, \beta^0, \theta^0)\frac{\partial \log q}{\partial x}(x, \beta^0)\right] = 0, \tag{4.3}$$

while we will use the notation

$$m(x, \beta, \theta) \equiv \frac{\partial \psi}{\partial x}(x, \beta, \theta) + \psi(x, \beta, \theta)\frac{\partial \log q}{\partial x}(x, \beta), \quad \tilde{m}(z_t, \beta, \theta) = m(x_t, \beta, \theta), \tag{4.4}$$

with $\psi(\cdot) = (\psi_1(\cdot), ..., \psi_p(\cdot))^\top$, where $\psi_i(\cdot)$, $i = 1, 2, ..., p$, are real functions for which assumption A1 holds. For notation convenience, for any function $g(x, \beta, \theta)$, $g^0(x)$ will denote $g(x, \beta^0, \theta^0)$; for instance, $\psi^0(x) = \psi(x, \beta^0, \theta^0)$ and $\dfrac{\partial \psi^0}{\partial \beta}(x) = \dfrac{\partial \psi}{\partial \beta}(x, \beta^0, \theta^0)$.

We assume that we have square-root T consistent estimators of β^0 and θ^0 denoted respectively by $\hat{\beta}_T$ and $\hat{\theta}_T$, which leads to the notation $\hat{x}_t = h(z_t, \hat{\beta}_T, \hat{\theta}_T)$. The main goal of the section is to derive sufficient conditions such that the asymptotic distributions of

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{m}(z_t, \hat{\beta}_T, \hat{\theta}_T) \quad \text{and} \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{m}^0(z_t)$$

coincide. In this case, we will say in the sequel that the test statistic based on (4.3) is robust against parameter uncertainty.

A Taylor expansion of $\tilde{m}(z_t, \hat{\beta}_T, \hat{\theta}_T)$ around (β^0, θ^0) yields to

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{m}(z_t, \hat{\beta}_T, \hat{\theta}_T) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{m}^0(z_t) + \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \tilde{m}^0}{\partial \beta^\top}(z_t)\right]\sqrt{T}(\hat{\beta}_T - \beta^0) \\ &+ \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \tilde{m}^0}{\partial \theta^\top}(z_t)\right]\sqrt{T}(\hat{\theta}_T - \theta^0) \ + \ o_p(1),\end{aligned}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{m}(z_t, \hat{\beta}_T, \hat{\theta}_T) = [I_p \ P_m] \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} m^0(z_t) \\ \sqrt{T}(\hat{\beta}_T - \beta^0) \\ \sqrt{T}(\hat{\theta}_T - \theta^0) \end{bmatrix} + o_p(1), \tag{4.5}$$

where I_p is the p×p identity matrix and $P_m = [P_{\psi\beta}\ P_{\psi\theta}]$ with

$$P_{\psi\beta} = E\left[\frac{\partial \tilde{m}^0}{\partial \beta^\top}(z_t)\right], \quad P_{\psi\theta} = E\left[\frac{\partial \tilde{m}^0}{\partial \theta^\top}(z_t)\right],$$

while the functions $m(\cdot)$ and $\psi(\cdot)$ are connected through (4.4).

Equation (4.5) implies that, in general, the asymptotic distribution of $T^{-1/2}\sum_{t=1}^{T}\tilde{m}(z_t,\hat{\beta}_T,\hat{\theta}_T)$ depends on the asymptotic distribution of the estimators $(\hat{\beta}_T,\hat{\theta}_T)$ and their covariance with $T^{-1/2}\sum_{t=1}^{T}\tilde{m}^0(z_t)$; see Newey (1985) and Tauchen (1985), as well as Gallant (1987), Gallant and White (1988), and Wooldridge (1990).

However, it is clear from (4.5) that a sufficient condition for the robustness of (4.3) against parameter uncertainty is

$$P_m = [P_{\psi\beta}\ P_{\psi\theta}] = 0. \tag{4.6}$$

In the sequel, we will propose two approaches that ensure (4.6). The first one will use the so-called generalized information matrix equality to characterize the functions $m(\cdot)$ such that (4.6) holds. The second one is due to Wooldridge (1990). We also make the connection between these two approaches.

4.1 First approach: orthogonality to the score function

When the expectation of a function $\tilde{m}(z_t,\gamma^0)$ equals zero, with $\gamma = (\beta,\theta)^\top$, one has the generalized information matrix equality, i.e.,

$$E\left[\frac{\partial \tilde{m}}{\partial \gamma^\top}(z_t,\gamma^0)\right] + E[\tilde{m}(z_t,\gamma^0)s(z_t,\gamma^0)^\top] = 0, \tag{4.7}$$

where $s(z_t,\gamma)$ is the score function of the variable z_t. Equation (4.7) have been used for instance in Newey and McFadden (1994). Consequently, the condition (4.6) which guarantees the robustness against parameter uncertainty of $\tilde{m}(\cdot)$ holds if and only if $\tilde{m}(z_t,\gamma^0)$ is orthogonal to the score, i.e.,

$$0 = E[\tilde{m}(z_t,\gamma^0)\ s(z_t,\gamma^0)^\top] = E[m(x_t,\gamma^0)\ s(h^{-1}(x_t,\gamma^0),\gamma^0)^\top]. \tag{4.8}$$

This result explains the finding of Bontemps and Meddahi (2005) who showed that Hermite polynomials, $H_i(\cdot)$, $i \geq 3$, are robust against parameter uncertainty when one tests that an observable variable (i.e., $z = x$) follows a $\mathcal{N}(\mu^0,(\sigma^0)^2)$ distribution. In this case, $\gamma^0 = (\mu^0,(\sigma^0)^2)^\top$ and the score function is given by

$$s(x,\gamma) = \begin{bmatrix} \dfrac{x-\mu}{\sigma^2} \\[2ex] \dfrac{(x-\mu)^2}{2\sigma^4} - \dfrac{1}{2\sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma}H_1\left(\dfrac{x-\mu}{\sigma}\right) \\[2ex] \dfrac{1}{\sqrt{2}\sigma^2}H_2\left(\dfrac{x-\mu}{\sigma}\right) \end{bmatrix},$$

where $H_1(\cdot)$ and $H_2(\cdot)$ are the first and second Hermite polynomials. However, the distribution of $(x-\mu^0)/\sigma^0$ is $\mathcal{N}(0,1)$. Consequently, the orthogonality of the Hermite polynomials implies that $\forall i \geq 3$, the test statistics based on $E[H_i((x-\mu^0)/\sigma^0)] = 0$ are robust against parameter uncertainty. Actually, Bontemps and Meddahi (2005) also showed this robustness result when x is not observable and is, for instance, the residual of a heteroskedastic and nonlinear regression model. In this case, one has to take into account the uncertainty in $h(z_t,\hat{\beta}_T,\hat{\theta}_T)$.

We now characterize the test-functions $\psi(\cdot)$ associated to $m(\cdot)$ such that (4.8) holds. Consider a function $g(\cdot)$, and assume that Assumption A1 holds for all the components of $\psi(\cdot)g(\cdot)^\top$, then applying (2.1) to $\psi(\cdot)g(\cdot)$ yields to

$$E[m(x_t,\gamma^0)g(x_t,\gamma^0)^\top] = -E[\psi(x_t,\gamma^0)\frac{\partial g}{\partial x}(x_t,\gamma^0)^\top].$$

Consequently, Eq. (4.8) is tantamount to

$$E[\psi(x_t,\gamma^0)\frac{\partial s}{\partial x}(h^{-1}(x_t,\gamma^0),\gamma^0)^\top] = 0, \tag{4.9}$$

when Assumption A1 holds for $\psi(x,\gamma)s(h^{-1}(x,\gamma),\gamma)^\top$. In the Appendix, we show that $\frac{\partial s}{\partial x}(h^{-1}(x_t,\gamma),\gamma) = (b_\beta(x,\gamma)^\top, b_\theta(x,\gamma)^\top)^\top$ where

$$\begin{aligned} b_\beta(x,\gamma) = &\frac{\partial^2 \log q}{\partial^2 x}(x,\beta)\frac{\partial h}{\partial \beta}(h^{-1}(x,\gamma),\gamma) + \frac{\partial \log q}{\partial x}(x,\beta)\frac{\partial^2 h}{\partial z\partial\beta}(h^{-1}(x,\gamma),\gamma)\frac{\partial h^{-1}}{\partial x}(x,\gamma) \\ &+ \frac{\partial^2 \log q}{\partial x\partial\beta}(x,\beta) + \frac{\partial^2 h^{-1}}{\partial^2 x}(x,\gamma)\frac{\partial^2 h}{\partial z\partial\beta}(h^{-1}(x,\gamma),\gamma) \\ &+ \left(\frac{\partial h^{-1}}{\partial x}(x,\gamma)\right)^2 \frac{\partial^3 h}{\partial^2 z\partial\beta}(h^{-1}(x,\gamma),\gamma), \end{aligned} \tag{4.10}$$

$$\begin{aligned} b_\theta(x,\gamma) = &\frac{\partial^2 \log q}{\partial^2 x}(x,\beta)\frac{\partial h}{\partial \beta}(h^{-1}(x,\gamma),\gamma) + \frac{\partial \log q}{\partial x}(x,\beta)\frac{\partial^2 h}{\partial z\partial\beta}(h^{-1}(x,\gamma),\gamma)\frac{\partial h^{-1}}{\partial x}(x,\gamma) \\ &+ \frac{\partial^2 h^{-1}}{\partial^2 x}(x,\gamma)\frac{\partial^2 h}{\partial z\partial\beta}(h^{-1}(x,\gamma),\gamma) + \left(\frac{\partial h^{-1}}{\partial x}(x,\gamma)\right)^2 \frac{\partial^3 h}{\partial^2 z\partial\beta}(h^{-1}(x,\gamma),\gamma). \end{aligned} \tag{4.11}$$

We are now able to state the following corollary that we write as a proposition:

Proposition 4.1 *Let $\psi(x,\beta,\theta)$ be a test-function such that Assumption A1 holds for $\psi(x,\beta^0,\theta^0)$ and $\psi(x,\gamma^0)s(h^{-1}(x,\gamma^0),\gamma^0)^\top$ where $s(\cdot,\gamma)$ is the score function of z_t, and*

$$E[\psi(x_t,\gamma^0)b_\beta(x,\gamma^0)^\top] = 0 \quad \text{and} \quad E[\psi(x_t,\gamma^0)b_\theta(x,\gamma^0)^\top] = 0. \tag{4.12}$$

Then, (4.3) *is robust against parameter uncertainty.*

In order to illustrate this proposition, consider the example where the variable x_t is observable, i.e., $x = z$ (there is no parameter θ and $h(z) = z$). Then, $b_\beta(x,\beta)$ becomes

$$b_\beta(x,\beta) = \frac{\partial^2 \log q}{\partial x\partial\beta}(x,\beta). \tag{4.13}$$

In the case of testing normality, Bontemps and Meddahi (2005) showed that (4.12) holds for Hermite polynomials $H_i(\cdot)$, $i \geq 3$, when one considers a regression-type model. However, while this orthogonality still holds for non-linear and heteroskedastic models (Bontemps and Meddahi (2005)), one can show that it does not necessarily hold when the variable x_t is transform of another distribution like in density forecasts.

Note also the robustness of orthonormal polynomials does not hold for other distributions; for instance, when x_t is observed and follows a $\mathrm{T}(\nu)$, $b_\beta(x,\beta)$ becomes

$$b_\nu(x,\nu) = \frac{x - x^3}{(\nu + x^2)^2}, \tag{4.14}$$

which is not a linear combination of Romanovski polynomials.

The remaining question in this subsection is the derivation of functions $\psi(\cdot)$ such that (4.12) holds. We adopt a regression approach: For a given $\psi(\cdot, \beta, \theta)$ function, define $\psi^{\perp}(\cdot, \beta, \theta)$ as

$$\psi^{\perp}(x, \beta, \theta) = \psi(x, \beta, \theta) - E[\psi(x, \beta, \theta)\zeta^{\top}(x, \beta, \theta)] \left(E[\zeta(x, \beta, \theta)\zeta^{\top}(x, \beta, \theta)]\right)^{-1} \zeta(x, \beta, \theta), \quad (4.15)$$

where

$$\zeta(x, \beta, \theta) = \left(b_{\beta}(x, \beta, \theta)^{\top}, b_{\theta}(x, \beta, \theta)^{\top}\right)^{\top}. \quad (4.16)$$

Then, the moment condition (4.3) associated to $\psi^{\perp}(\cdot)$ is robust against parameter uncertainty (when Assumption A1 holds for $\zeta(x, \beta, \theta)$).

In some cases like the GARCH example we consider in the simulation and empirical sections, one can compute analytically $\psi^{\perp}(x, \beta, \theta)$. However, if this is not the case, then one can do an empirical regression to get estimates of $E[\psi(x, \beta, \theta)\zeta^{\top}(x, \beta, \theta)]$ and $E[\zeta(x, \beta, \theta)\zeta^{\top}(x, \beta, \theta)]$. The corresponding test function $\psi^{\perp}(x, \beta, \theta)$ is indeed robust against parameter uncertainty.

We could have done the same treatment on $\dot{m}()$. Having a moment and projecting it on the orthogonal of the score will give to us a moment whose expectation is, under the null, equal to zero and robust by construction to the parameter uncertainty. One should first notice that, if the parameters of interest are estimated by maximum likelihood, the statistic in (3.2) gives exactly the same value than the one computed with the correction used in Newey (1985). it means that we will not loose any power by restricting our selves to moment which are robust.

Moreover it is sometimes easier to work with ψ than with m. in the Monte Carlo simulations, working on ψ allows us to have analytical expressions.

4.2 Wooldridge's approach

An alternative approach is obtained by transforming the moment (4.3). More precisely, let S be a matrix. Then one has $E[S\tilde{m}(z, \gamma^0)] = 0$ and these moments are robust against parameter uncertainty when $SP_m = 0$. This approach is not always possible. In particular, one needs that the dimension of $m(\cdot)$, i.e., p, exceeds the dimension of (β^0, θ^0), denoted k ($p > k$). In this case, when one assume that P_m has a full rank, a simple choice of S is

$$S = I_p - P_m[P_m^{\top} P_m]^{-1} P_m^{\top}. \quad (4.17)$$

This general approach is due to Wooldridge (1990). He provided this approach as well a conditional version of it, i.e., when the moments of interest are conditional ones. Interestingly, by using a Khmaladze (1981)'s transform to get a robust test against parameter uncertainty, Bai (2003) did the conditional approach of Wooldridge (in an infinite dimensional space).

Observe that the solution (4.17) is not unique, i.e., when one has more structure on the model, one can derive other matrices S such that $SP_m = 0$. Duan (2003) proposed robust moment tests by considering other choices than (4.17) to get $SP_m = 0$.

We conclude this section by mentioning that Eq. (4.8) implies that the robust moments of Wooldridge (1990), Bai (2003) and Duan (2003) are orthogonal to the score function.

5 A Monte Carlo Study

This section provides Monte Carlo simulations to assess the finite sample properties of our test procedures. All the simulations are based on 10,000 replications. Three sample sizes are considered: 100, 500 and 1,000. We focus on testing the Student distributional assumption which is important empirically; Bontemps and Meddahi (2005) studied the normal case.

We study various cases: the variable is observable, not observable, in a cross-section case or a time series case. Without having prior on the degrees of freedom ν of the Student distribution, it seems difficult to use polynomial for testing purposes. In addition, we focus on even moment conditions (which correspond to odd test functions ψ) given that one often uses Student specification for modeling symmetric distribution. Finally, our goal is to provide simple and systematic test procedures, i.e., valid tests whatever the value of ν. We therefore focus on the test-functions $\psi_\alpha(\cdot,\nu)$ for various values of α with

$$\psi_\alpha(x,\nu) = \frac{x}{(x^2+\nu)^\alpha}, \quad m_\alpha(x,\nu) = \frac{\nu - (2\alpha+\nu)x^2}{(\nu+x^2)^{\alpha+1}}. \tag{5.1}$$

where $m_\alpha(\cdot)$ are the corresponding moment conditions given in Eq. (2.1). We consider univariate tests $m_\alpha(\cdot)$ based on a particular set of positive values {0,1/2,1,2,3,4}. Given that the moments $m_\alpha(x_t)$ are somewhat highly correlated in practice, we only perform one joint test (denoted m_j) based on $m_0(\cdot)$ and $m_1(\cdot)$.

5.1 Observable variables

In this case, we know from Proposition 4.1 that a sufficient condition for having robustness against parameter uncertainty is to consider a test function $\psi(\cdot)$ orthogonal to $b_\nu(\cdot,\nu)$ defined in (4.14). Following (4.15), we transform $\psi_\alpha(\cdot)$ to get $\psi_\alpha^\perp(\cdot)$, where the corresponding moment condition denoted $m_\alpha^\perp(\cdot)$ equals

$$m_\alpha^\perp(x,\nu) = \frac{\nu - (2\alpha+\nu)x^2}{(\nu+x^2)^{\alpha+1}} - k_\alpha(\nu)\left(\frac{x^4(\nu+2) - 4x^2(\nu+1)+\nu}{(\nu+x^2)^3}\right) \tag{5.2}$$

where $k_\alpha(\nu)$, a proof of (5.2), and the variance of $m_\alpha^\perp(x,\nu)$ are given in the Appendix.

5.1.1 The cross-section case

We first study the size properties of our tests. We consider the cases when ν equals 5 (Panel A) and 20 (Panel B). In Table 1, we assume that ν is known in the left-hand side set of columns while it is estimated from the empirical second moment of the variable in the two other sets. When ν is known, we used test based on $m_\alpha(\cdot)$ defined in (5.1) and their robust form given in (5.2) where the variances used to derive the test statistics (3.2) are the theoretical ones (the results are given in the Appendix). We also compute the Kolmogorov-Smirnov test (denoted KS). The simulation results clearly show that the finite sample performance of the new tests are quite good and close to the nominal level, whatever the sample size and the value of α. There are also very small differences between $m_\alpha(x,\nu)$ and their robust forms $m_\alpha^\perp(x,\nu)$. The finite sample properties of the KS test are also quite good.

When ν is estimated,[2] two sub-cases are considered. In the first one (the second set of columns denoted "in population"), the variances used in the test-statistics are the theoretical ones whereas in the second one (the third set of columns denoted "in sample") they are

[2]The theoretical variance of a T(ν) distribution equals $\nu/(\nu-2)$, which is therefore higher than one. In practice, it happens few times that the empirical variance is smaller than one. This happens when the sample size is small or ν high. In these case, we set $\hat{\nu}$ equal to 500. The performance properties seem insensitive to the choice of this high value. The results are available upon request.

estimated. The results of Table 1 clearly show that the non-robust moments are quite sensitive to the parameter uncertainty and that they are not reliable. In contrast, robust tests are quite reliable. In addition, estimating the variance of the moments leads to a very small size distortion. Surprisingly, KS test is reliable, which is not the case for other distributions like the normal one (see Bontemps and Meddahi (2005)). It is worth noticing than estimating ν by the maximum likelihood method leads to similar results (they are available upon request).

We also perform the test developed in Bai (2003). As we mentioned in Section 4, this test, denoted in the tables S_{Bai}, presents size distortions when one tests normality which leads Bai (2003) to consider another statistic, denoted S^T_{Bai} in the tables, computed as the maximum over the 90% smallest values of the individual statistics. Surprisingly, the reverse holds for the Student case, i.e., the results reported in Table 1 show that S^T_{Bai} presents size distortions while S_{Bai} does not. The finite sample properties of S_{Bai} are similar to those of our new tests.

In Table 2, we study the power properties of our tests against an asymmetric distribution and against the mixture of two standard normals. The parameter ν is estimated by the second moment assuming that the data are i.i.d.; all the expectations are computed in the samples. We compare the power properties with those of the tests developed by Bai (2003).

The asymmetric distribution we consider is derived from a $\chi^2(7)$. More precisely, we consider an affine transform of the $\chi^2(7)$ such that it matches the mean and variance of a Student T(5) or T(20). However, we did not assume that we know the variance in our testing procedures. The results in Table 2 clearly show that our tests have very good power, similar to the power of Bai's test.

We consider three examples of mixture of two centered normals, i.e., we set the probability p to have one of the normal equal to 0.7, 0.8, and 0.9. For a given p, the variances of the two normal distributions are chosen such that the second and fourth moments of the mixture distribution equal those of a T(5) or T(20). When p increases, the sixth moment of the mixture distribution increases; we report in the Appendix the corresponding moments as well as the theoretical variances of each component of the mixture. The simulation results are reported in Table 2. The main finding is that the joint test denoted $m_J^\perp$ has the best power among all tests. In addition, most of our tests, have more power than the test of Bai (2003). The results show that our tests have a very good power whatever the sample size when the null is a T(5) distribution. The power is somewhat lower when p equals 0.9; the main reason is that the sixth moment of the mixture of normals equals 1,001, which is quite high. In contrast, the power decreases significantly when the null is a T(20) distribution. Again, the mixture of two normals is very close to a T(20) distribution; for instance, the first fifth moments are the same while the sixth ones are quite close (see the Appendix). Actually, we perform the likelihood ratio test where the critical values are computed by simulations for each sample size. We do know by Neyman-Pearson theorem that this test is the optimal one. The simulation rejection frequencies are 6.3%, 9.6%, and 12.7% when the sample size equals 100, 500, and 1,000 respectively.

5.1.2 The serial correlation case

We now study the finite sample properties of our tests when the variable of interest is serially correlated with unknown dependence structure. We use the same tests as previously. We use a HAC method to estimate the variances/covariances. The HAC method is developed by using the quadratic kernel with an automatic lag selection procedure *á la* Andrews (1991). However, we do not perform the test of Bai (2003) given that it is not valid under serial correlation.

The process x_t is defined as $x_t = u_t/\sqrt{s_t}$ where the variables u_t and s_t are independent, the distribution of u_t is $\mathcal{N}(0,1)$ while s_t follows a gamma (ν/2,2/ν,0) distribution, where ν equals

5 or 20 as in our previous simulations. Consequently, the unconditional distribution of x_t is $\mathrm{T}(\nu)$. However, there is a dependence in u_t while s_t is i.i.d.; we assumed that the conditional distribution of u_t given its past is $\mathcal{N}(\rho u_{t-1}, 1-\rho^2)$ where ρ equals 0.4 or 0.9. When we study the power of the tests, we assumed that x_t is an AR(1), $x_t = \rho x_{t-1} + \varepsilon_t$, whose innovation process ε_t is, as in the previous simulations, a mixture of two normals where p equals 0.7 or 0.9.

Again, ν is estimated with the second moment of x_t. The results reported in Table 3 show that the size of the tests is very good. The analysis of the power is mixed. When ρ equals 0.9, the process x_t is close to a normal distribution given the central limit theorem (see Bai and Ng (2005)). Consequently, our tests detect more easily the departure from a $\mathrm{T}(\nu)$ distribution, especially when ν equals 5. When ρ equals 0.4, the distribution of x_t is close to the one of ε_t. Therefore, we get similar results than for the cross-sectional case.

5.2 The residual case: Student GARCH models

We now study a more general model where x_t is an innovation process of a GARCH. More precisely, we consider Student GARCH(1,1) model of Bollerslev (1987), i.e.,

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sqrt{v_t}\, u_t, \quad v_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta v_{t-1}, \quad u_t = \sqrt{\frac{\nu-2}{\nu}} x_t, \quad x_t \sim \mathrm{T}(\nu), \tag{5.3}$$

where $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$ and $\beta = 0.8$. We use exactly the same distribution for x_t as in the cross-section examples when we study the size and power of the tests.

The parameter $\gamma \equiv (\mu, \omega, \alpha, \beta)^\top$ is estimated with a Gaussian-QMLE procedure which is known to be consistent provided that the conditional mean and variance process of y_t are correctly specified (Bollerslev and Wooldridge (1992)). We then construct an estimator of u_t by using $\hat{u}_t = (y_t - \hat{\mu})/\sqrt{\hat{v}_t}$. Under H_0, u_t is a linear transformation of a Student distribution. Given that the empirical variance of u_t is by construction almost one, we estimated ν by using the fourth moment of u_t, i.e. $Eu_t^4 = 3(\nu-2)/(\nu-4)$.

Following (4.15), one can show that when a function $\psi(\cdot)$ is orthogonal to the functions $x/(x^2+\nu)^2$, $(x^2-\nu)/(x^2+\nu)^2$, and $x^3/(x^2+\nu)^2$, the corresponding moment condition is robust against parameter uncertainty. In our simulations, we consider the functions $\psi_\alpha(\cdot)$. Observe that $\psi_2(\cdot)$ equals the first function while $\psi_1(\cdot)$ is a linear combination of the first and third functions. Therefore, these functions are not included in our tests. The projections are done analytically; the calculations are provided in the Appendix.

Table 5 reports the size results. One can notice that the size properties are quite good though we have small over-rejection in some cases. The performances are quite similar to the Bai test. In the analysis of the power, we used the same distributions as in the cross-section case. Table 4 reports the power results against the mixture of normals while Table 5 reports those against an asymmetric distribution. We observe qualitatively the same results than in the observable case with less power due to the fact that we estimate five parameters instead of one.

6 Empirical example

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

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

and the process u_t is assumed to be i.i.d. and $\mathcal{N}(0,1)$. Two important characteristics of GARCH models are that the kurtosis of y_t is higher than for a normal variable and the process exhibits a clustering effect. It turns out that financial returns share these two properties and therefore GARCH models describe financial data; for a survey on GARCH models, see, e.g., Bollerslev, Engle and Nelson (1994).

However, some empirical studies found that the implied kurtosis of a GARCH(1,1) is lower than empirical ones. These studies lead Bollerslev (1987) to assume that the standardized process u_t may follow a Student distribution given in (5.3) where $\mu = 0$. Under this assumption, GARCH(1,1) fits financial returns very well. Indeed, by using a Bayesian likelihood criterion, Kim, Shephard and Chib (1998) proved that a Student GARCH(1,1) outperforms another very popular volatility model, namely the log-normal stochastic volatility model of Taylor (1986) popularized by Harvey, Ruiz and Shephard (1994) and Jacquier, Polson and Rossi (1994). We already tested in Bontemps and Meddahi (2005) the normality of u_t and we strongly rejected it, corroborating the results of Kim, Shephard and Chib (1998). These authors estimated the T-GARCH model by the maximum likelihood method and find that the degree of freedom of the returns of FF-US$, UK-US$, SF-US$, and Yen-US$, equals 12.82, 9.71, 7.57, and 6.86 respectively. Our goal is to test the Student specification of the innovations x_t.

We consider the same data as Harvey, Ruiz and Shephard (1994), Kim, Shephard and Chib (1998), and Bontemps and Meddahi (2005),[3] i.e., observations of weekday close exchange rates from 1/10/81 to 28/6/85. The exchange rates are the U.K. Pound, French Franc, Swiss Franc and Japanese Yen, all versus the U.S. Dollar. After the QML estimation, we get the fitted residuals $\hat{u}_t$ and test their distribution. We use the same moments as in the monte carlo section. None of our tests reject the Student assumption. Likewise, Bai test does not reject the assumption. These results corroborate the finding of Kim, Shephard and Chib (1998).

7 Conclusion

When one specifies a distribution, one gets a general set of moment conditions. We used these moment conditions for testing purposes in a GMM setting. This setting allows us to study two important statistical issues. We deal in detail with the problem of parameter uncertainty by providing robust moments against this uncertainty. We also use the HAC method to handle the problem of potential serial correlation in the variable of interest. An extensive simulation study in the Student case shows that the finite sample properties of our tests are very good.

An important feature of our test method is its simplicity given that researchers are quite familiar with the method of moments. More importantly, the set of moment conditions we used holds also for discrete variable and for multivariate ones. These extensions are under investigation in Bontemps (2006) and Bontemps, Koumtingue and Meddahi (2006) respectively.

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

Appendix

Proof of Proposition 2.1. The continuity of $m(\cdot)$ and $q(\cdot)$ imply that $\psi(\cdot)$ defined in (2.4) is differentiable. By differentiating (2.4), one gets

$$\psi'(x) = m(x) - \frac{q'(x)}{q^2(x)} \int_l^x m(u)q(u)du = m(x) - \psi(x)(\log q)'(x),$$

i.e. (2.3). For any function $m(\cdot)$, we have $\lim_{x\to l} \psi(x)q(x) = 0$ and $\lim_{x\to r} \psi(x)q(x) = E[m(x)]$. Hence, (2.2) holds if and only if assumption A1 holds.■

Proof of Proposition 2.2. The functions $q_X(\cdot)$ and $q_Y(\cdot)$ are connected by the relation

$$q_Y(y) = (G^{-1})'(y)q_X(x) = \frac{1}{G' \circ G^{-1}(y)} q_X(x) = \frac{1}{G' \circ G^{-1}(y)} q_X(G^{-1}(y)).$$

Observe that

$$q_Y(y)\psi_Y(y) = q_X(G^{-1}(y))\psi_X(G^{-1}(y)). \tag{A.1}$$

By deriving the previous equality with respect to y, one gets:

$$\begin{aligned} q'_Y(y)\psi_Y(y) + q_Y(y)\psi'_Y(y) &= (G^{-1})'(y) \left(q'_X(G^{-1}(y))\psi_X(G^{-1}(y)) + q_X(G^{-1}(y))\psi'_X(G^{-1}(y))\right) \\ &= \frac{q_Y(y)}{q_X(x)} \left(q'_X(x)\psi_X(x) + q_X(x)\psi'_X(x)\right), \end{aligned}$$

which leads to (2.5) given that $q_Y(y) \neq 0$. Finally, (2.6) is implied by (A.1) and the continuity and monotonicity of $G(\cdot)$.■

Proof of (4.10) and (4.11): Given that the density function of x_t is $q(x,\beta)$, the density function of z_t is

$$q_z(z,\gamma) = \left|\frac{\partial h}{\partial z}(z,\gamma)\right| q(h(z,\gamma),\beta).$$

Hence,

$$s(z,\gamma) = \begin{pmatrix} \frac{\partial \log q}{\partial x}(h(z,\gamma),\beta)\frac{\partial h}{\partial \beta}(z,\gamma) + \frac{\partial \log q}{\partial \beta}(h(z,\gamma),\beta) + \left(\frac{\partial h}{\partial z}(z,\gamma)\right)^{-1} \frac{\partial^2 h}{\partial z \partial \beta}(z,\gamma) \\ \frac{\partial \log q}{\partial x}(h(z,\gamma),\beta)\frac{\partial h}{\partial \theta}(z,\gamma) + \left(\frac{\partial h}{\partial z}(z,\gamma)\right)^{-1} \frac{\partial^2 h}{\partial z \partial \theta}(z,\gamma) \end{pmatrix}.$$

By differentiating the equality $h(h^{-1}(x,\gamma),\gamma) = x$, one gets

$$s(z,\gamma) = \begin{pmatrix} \frac{\partial \log q}{\partial x}(x,\beta)\frac{\partial h}{\partial \beta}(z,\gamma) + \frac{\partial \log q}{\partial \beta}(x,\beta) + \frac{\partial h^{-1}}{\partial x}(x,\gamma)\frac{\partial^2 h}{\partial z \partial \beta}(z,\gamma) \\ \frac{\partial \log q}{\partial x}(x,\beta)\frac{\partial h}{\partial \theta}(z,\gamma) + \frac{\partial h^{-1}}{\partial x}(x,\gamma)\frac{\partial^2 h}{\partial z \partial \theta}(z,\gamma) \end{pmatrix}.$$

By differentiating the previous equation with respect to x, one gets (4.10) and (4.11).■

Computations of moments robust to parameter uncertainty in the Student case

Following (2.1) and (4.13), we know that we can obtain a moment whose expectation is zero with respect to the $T(\nu)$ distribution and robust to the parameter uncertainty problem if it is constructed from a test-function ψ_α orthogonal to the derivative of the score, *i.e.* $\frac{x - x^3}{(\nu + x^2)^2}$. Let define $\psi_\alpha^\perp$ the orthogonalized test-function:

$$\psi_\alpha^\perp(x,\nu) = \psi_\alpha(x,\nu) - \frac{E\left[\psi_\alpha(x,\nu)\frac{x-x^3}{(\nu+x^2)^2}\right]}{E\left[\frac{x-x^3}{(\nu+x^2)^2}\right]^2}\frac{x-x^3}{(\nu+x^2)^2}$$

A first expectation is of interest:

$$A_\alpha^\nu = E\left[\frac{1}{(x^2+\nu)^\alpha}\right] = \frac{1}{\nu^\alpha}\frac{\Gamma(\alpha+\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\alpha+\frac{\nu+1}{2})},$$

where the expectation (and the followings) is taken with respect to the $T(\nu)$ distribution.
Standard calculations lead to:

$$\begin{aligned}
E\left[\psi_\alpha(x,\nu)\frac{x-x^3}{(\nu+x^2)^2}\right] &= A_\alpha^\nu\left(-1+\frac{2\nu+1}{\nu}\frac{\alpha+\frac{\nu}{2}}{\alpha+\frac{\nu+1}{2}} - \frac{\nu+1}{\nu}\frac{\alpha+\frac{\nu}{2}}{\alpha+\frac{\nu+1}{2}}\frac{\alpha+1+\frac{\nu}{2}}{\alpha+1+\frac{\nu+1}{2}}\right)\\
E\left[\frac{x-x^3}{(\nu+x^2)^2}\right]^2 &= \left(\frac{1}{\nu+1} - \frac{3\nu+2}{\nu}\frac{\nu+2}{(\nu+1)(\nu+3)}\right.\\
&\quad\left.+\frac{3\nu+1}{\nu^2}\frac{(\nu+4)(\nu+2)}{(\nu+5)(\nu+3)} - \frac{\nu+1}{\nu^2}\frac{(\nu+6)(\nu+4)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)}\right)
\end{aligned}$$

The moment $m_\alpha^\perp(x,\nu)$ constructed from $\psi_\alpha(x,\nu)$ could be expressed as:

$$m_\alpha^\perp(x,\nu) = \underbrace{\frac{\nu-(2\alpha+\nu)x^2}{(\nu+x^2)^{\alpha+1}}}_{l_1(x)} - \underbrace{\frac{E\left[\psi_\alpha(x,\nu)\frac{x-x^3}{(\nu+x^2)^2}\right]}{E\left[\frac{x-x^3}{(\nu+x^2)^2}\right]^2}}_{k_\alpha}\underbrace{\left(\frac{x^4(\nu+2)-4x^2(\nu+1)+\nu}{(\nu+x^2)^3}\right)}_{l_2(x)}$$

Its variance can be computed using the equality:

$$Var[m_\alpha^\perp(x,\nu)] = Var[l_1(x)] + (k_\alpha)^2\,Var[l_2(x)] - 2k_\alpha Cov(l_1(x),l_2(x))$$

Like previously, we obtain:

$$\begin{aligned}
Var[l_1(x)] =& (2\alpha+\nu)^2A_{2\alpha}^\nu - 2\nu(2\alpha+\nu)(2\alpha+\nu+1)A_{2\alpha+1}^\nu + (2\alpha\nu+\nu+\nu^2)^2A_{2(\alpha+1)}^\nu\\
Var[l_2(x)] =& (\nu+2)^2A_2^\nu - 4(\nu+2)(\nu^2+4\nu+2)A_3^\nu\\
&+ \left(\left(4(\nu^2+4\nu+2)^2 + 2\nu(\nu+1)(\nu+2)(\nu+5)\right)A_4^\nu\right.\\
&- 4\nu(\nu+1)(\nu+5)(\nu^2+4\nu+2)A_5^\nu + \left(\nu(\nu+1)(\nu+5)\right)^2A_6^\nu\\
Cov(l_1(x),l_2(x)) =& (\nu+2)(2\alpha+\nu)A_{\alpha+1}^\nu + (\nu(\nu+2)(2\alpha+\nu+1) + 2(\nu+2\alpha)(\nu^2+4\nu+2))A_{\alpha+2}^\nu\\
&- \nu(2(\nu^2+4\nu+2)(\nu+2\alpha+1) + (2\alpha+\nu)(\nu+1)(\nu+5))A_{\alpha+3}^\nu\\
&+ \nu^2(\nu+1)(\nu+5)(\nu+2\alpha+1)A_{\alpha+4}^\nu.
\end{aligned}$$

Computations for the power case with mixture of normals.
The mixture of normals used in the Monte Carlo section are such that the first five moments are equal to the moments of a T(ν) distribution. Given the weights ((p, $1-p$), p=0.7, 0.8 and 0.9), the variances of the normal are function of p and ν. In fact:

$$\sigma_1^2 = \frac{\nu}{\nu-2}\left(1-\sqrt{\frac{1-p}{p}\frac{2}{\nu-4}}\right)$$
$$\sigma_2^2 = \frac{\nu}{\nu-2}\left(1+\sqrt{\frac{p}{1-p}\frac{2}{\nu-4}}\right)$$

The following Table gives the value of the moments as function of p for the two values of ν used in the simulations: 5 for Panel A and 20 for Panel B.

Moments of the mixtures and of the Student

Panel A: $\nu = 5$

		EX^2	EX^4	EX^6
$T(\nu)$		1.66	25	—
$pN(0,\sigma_1^2)+$ $(1-p)N(0,\sigma_2^2)$	$p=0.7$	1.66	25	657.6
	$p=0.8$	1.66	25	780.7
	$p=0.9$	1.66	25	1009.9

Panel B: $\nu = 20$

		EX^2	EX^4	EX^6
$T(\nu)$		1.11	4.16	29.76
$pN(0,\sigma_1^2)+$ $(1-p)N(0,\sigma_2^2)$	$p=0.7$	1.11	4.16	29.09
	$p=0.8$	1.11	4.16	29.66
	$p=0.9$	1.11	4.16	30.72

Computations for the Garch example.

In the Garch example the function which relates the observables y_t to the variable of interest ε_t is given by: $h(y_t,\mu,\omega,\tilde{\alpha},\beta,\nu) = \sqrt{\frac{\nu}{\nu-2}}\frac{y_t-\mu}{\sqrt{v_t}}$ where v_t is the volatility (which therefore depends on the parameters and the past-values of y_t).

Following (4.10) and (4.11), we obtain

$$\begin{aligned}
b_\nu^0(x) &= \frac{3\nu^0 x-(\nu^0-2)x^3}{(\nu^0+x^2)^2} = 4\nu^0\psi_2(x,\nu^0)-(\nu-2)\psi_1(x,\nu^0),\\
b_\mu^0(x) &= -\frac{1}{\sqrt{v_t}}\frac{(\nu^0+1)(x^2-\nu^0)}{(\nu^0+x^2)^2},\\
b_\theta^0(x) &= \frac{\partial v_t}{\partial\theta}\frac{1}{v_t}\frac{(\nu^0+1)\nu^0 x}{(x^2+\nu^0)^2} = \frac{\partial v_t}{\partial\theta}\frac{1}{v_t}(\nu^0+1)\nu^0\psi_1(x,\nu^0),
\end{aligned}$$

where $\theta = (\omega,\tilde{\alpha},\beta)$. At the end, the space spanned by these functions has indeed a dimension equal to 3. The ψ_α test-function is by construction orthogonal to $b_\mu^0(x)$. for symmetry reasons. ψ_α should therefore be projected only on the orthogonal space spanned by ψ_1 and ψ_2 or equivalently the orthogonal of the space spanned by ψ_2 and $\frac{x^3}{(\nu^0+x^2)^2}$. Denoting $< A(x), B(x) >_{q_\nu}$ the expectation of AB with respect to the $T(\nu)$ distribution, we can compute the coefficient of the projection of ψ_α on the space spanned by ψ_1 and ψ_2.

$$
\begin{aligned}
< \psi_\alpha(x,\nu), \psi_2(x,\nu) >_{q_\nu} &= A^\nu_{\alpha+1} - \nu A^\nu_{\alpha+2}, \\
< \psi_\alpha(x,\nu), \frac{x^3}{(\nu+x^2)^2} >_{q_\nu} &= A^\nu_{\alpha} - 2\nu A^\nu_{\alpha+1} + \nu^2 A^\nu_{\alpha+2}, \\
< \psi_2(x,\nu), \psi_2(x,\nu) >_{q_\nu} &= A^\nu_3 - \nu A^\nu_4, \\
< \psi_2(x,\nu), \frac{x^3}{(\nu+x^2)^2} >_{q_\nu} &= A^\nu_2 - 2\nu A^\nu_3 + \nu^2 A^\nu_4, \\
< \frac{x^3}{(\nu+x^2)^2}, \frac{x^3}{(\nu+x^2)^2} >_{q_\nu} &= A^\nu_1 - 3\nu A^\nu_2 + 3\nu^2 A^\nu_3 - \nu^3 A^\nu_4.
\end{aligned}
$$

The robust test function $\psi^\perp_\alpha(x,\nu)$ is thus equal to

$$
\psi^\perp_\alpha(x,\nu) = \psi_\alpha(x,\nu) - \left[\psi_2(x,\nu), \frac{x^3}{(\nu+x^2)^2}\right] P \left[\begin{array}{c} < \psi_\alpha(x,\nu), \psi_2(x,\nu) >_{q_\nu} \\ < \psi_\alpha(x,\nu), \frac{x^3}{(\nu+x^2)^2} >_{q_\nu}, \end{array} \right]
$$

where

$$
P = \left[\begin{array}{cc} < \psi_2(x,\nu), \psi_2(x,\nu) >_{q_\nu} & < \psi_2(x,\nu), \frac{x^3}{(\nu+x^2)^2} >_{q_\nu} \\ < \psi_2(x,\nu), \frac{x^3}{(\nu+x^2)^2} >_{q_\nu} & < \frac{x^3}{(\nu+x^2)^2}, \frac{x^3}{(\nu+x^2)^2} >_{q_\nu} \end{array} \right]^{-1}.
$$

At the end, the moment used for testing the student-GARCH assumption is, following (2.1) equal to:

$$
m^\perp_\alpha(x,\nu) = \frac{\partial}{\partial x}\left(\psi^\perp_\alpha(x,\nu)\right) - \frac{x}{\nu+x^2}\left(\psi^\perp_\alpha(x,\nu)\right).
$$

Table 1: Size of the tests

ν known.				ν estimated by the 2nd moment.							
				in population				in sample			
Panel A: $\nu = 5$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
m_0	4.6	5.0	5.0	m_0	0.6	0.7	0.9				
$m_{\frac{1}{2}}$	4.9	5.0	5.2	$m_{\frac{1}{2}}$	1.4	1.7	1.8				
m_0	5.1	5.1	5.2	m_1	2.3	2.5	2.6				
m_2	5.2	5.0	5.0	m_2	3.4	3.8	4.0				
m_3	5.2	4.9	5.2	m_3	3.8	4.3	4.7				
m_4	5.0	4.9	5.2	m_4	4.2	4.5	5.0				
m_j	4.6	5.1	4.9	m_j	2.3	2.4	2.4				
KS	4.9	5.0	4.8	KS	4.8	4.8	4.6				
$m_0^\perp$	5.1	5.0	5.1	$m_0^\perp$	5.0	5.0	5.1	$m_0^\perp$	5.1	5.0	5.1
$m_{\frac{1}{2}}^\perp$	5.1	4.9	5.2	$m_{\frac{1}{2}}^\perp$	5.2	4.9	5.2	$m_{\frac{1}{2}}^\perp$	5.1	5.0	5.2
$m_1^\perp$	5.1	4.9	5.1	$m_1^\perp$	5.1	5.0	5.1	$m_1^\perp$	5.0	5.0	5.0
$m_2^\perp$	5.1	4.9	5.1	$m_2^\perp$	5.1	5.0	5.1	$m_2^\perp$	5.0	5.0	5.0
$m_3^\perp$	5.1	4.9	5.3	$m_3^\perp$	5.3	4.9	5.3	$m_3^\perp$	5.1	5.0	5.1
$m_4^\perp$	5.1	4.9	5.1	$m_4^\perp$	5.1	4.8	5.2	$m_4^\perp$	5.2	4.9	5.2
$m_j^\perp$	5.0	4.9	5.0	$m_j^\perp$	4.8	5.1	5.0	$m_j^\perp$	6.0	5.5	5.1
				S_{Bai}^T	1.9	2.5	2.4				
				S_{Bai}	5.2	6.6	5.9				
Panel B: $\nu = 20$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
m_0	4.8	5.1	5.1	m_0	0.0	0.0	0.0				
$m_{\frac{1}{2}}$	4.9	5.2	5.0	$m_{\frac{1}{2}}$	0.0	0.0	0.0				
m_1	5.0	5.1	4.9	m_1	0.3	0.1	0.0				
m_2	5.1	5.1	4.9	m_2	0.8	0.3	0.2				
m_3	5.1	5.1	4.9	m_3	1.2	0.7	0.4				
m_4	5.0	5.2	5.1	m_4	1.7	1.1	0.8				
m_j	5.2	5.0	5.1	m_j	1.9	3.4	3.9				
KS	4.3	4.9	4.9	KS	4.1	4.4	4.4				
$m_0^\perp$	5.3	5.1	5.0	$m_0^\perp$	3.5	4.4	4.7	$m_0^\perp$	4.9	4.3	3.9
$m_{\frac{1}{2}}^\perp$	5.3	5.1	5.0	$m_{\frac{1}{2}}^\perp$	3.5	4.3	4.7	$m_{\frac{1}{2}}^\perp$	5.0	4.3	3.9
$m_1^\perp$	5.3	5.1	5.0	$m_1^\perp$	3.4	4.3	4.8	$m_1^\perp$	5.0	4.4	3.9
$m_2^\perp$	5.3	5.1	5.0	$m_2^\perp$	3.4	4.3	4.8	$m_2^\perp$	5.0	4.4	3.9
$m_3^\perp$	5.3	5.1	5.0	$m_3^\perp$	3.5	4.4	4.7	$m_3^\perp$	5.1	4.3	3.9
$m_4^\perp$	5.2	5.1	5.1	$m_4^\perp$	3.4	4.2	4.7	$m_4^\perp$	5.0	4.3	3.9
$m_j^\perp$	5.0	5.2	5.3	$m_j^\perp$	2.5	4.7	5.3	$m_j^\perp$	6.2	8.4	7.8
				S_{Bai}^T	1.4	2.6	2.3				
				S_{Bai}	3.9	6.8	6.3				

Note: The data are i.i.d. from a T(ν) distribution where $n = 5$ (Panel A) or $\nu = 20$ (Panel B); ν is either assumed known or estimated. The results are based on 10,000 replications. For each sample size, we provide the percentage of rejection at a 5% level. m_α corresponds to the moment test based on the test-function $\psi_\alpha(x, \nu)$, $m_\alpha^\perp$ is the moment robust against parameter uncertainty, KS is the Kolmogorov-Smirnov test. $m_\alpha^\perp$ is obtained either analytically (column denoted "in population") or empirically (column denoted "in sample"). The variances are computed theoretically or in the sample. $m_j^\perp$ corresponds to the joint test $m_0^\perp - m_1^\perp$. S_{Bai} is the statistic used by Bai (2003), S_{Bai}^T is the same statistic that one gets when one excludes 10% higher absolute values of the empirical process involved in the statistic.

Table 2: Power of the tests

	Asymetric Distribution			Mixture of normals $p=0.7$			Mixture of normals $p=0.8$			Mixture of normals $p=0.9$		
Panel A: $VX=\frac{5}{3}$.												
T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	42.6	98.0	100.0	100.0	100.0	100.0	63.4	99.8	100.0	7.0	8.8	10.0
$m_{\frac{1}{2}}^\perp$	42.8	97.8	100.0	100.0	100.0	100.0	73.0	100.0	100.0	7.8	17.4	27.4
$m_1^\perp$	42.4	97.5	100.0	100.0	100.0	100.0	76.8	100.0	100.0	8.7	23.4	38.9
$m_2^\perp$	42.4	97.5	100.0	100.0	100.0	100.0	76.8	100.0	100.0	8.7	23.4	38.9
$m_3^\perp$	42.4	97.6	100.0	100.0	100.0	100.0	72.4	100.0	100.0	7.9	18.5	29.5
$m_4^\perp$	41.9	97.5	100.0	100.0	100.0	100.0	65.6	100.0	100.0	7.3	13.8	20.9
$m_j^\perp$	35.0	95.8	99.9	100.0	100.0	100.0	73.3	100.0	100.0	19.0	76.8	97.8
S_{Bai}^T	29.9	91.8	99.8	52.8	100.0	100.0	6.6	48.0	84.1	2.0	3.1	4.4
S_{Bai}	42.9	94.2	99.8	81.7	100.0	100.0	28.0	85.7	98.6	9.2	22.7	33.4
Panel B: $VX=\frac{20}{18}$.												
T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	9.8	23.5	38.2	4.9	4.2	4.2	5.0	4.6	5.1	5.3	5.2	6.0
$m_{\frac{1}{2}}^\perp$	9.5	24.9	42.0	5.0	4.3	4.2	5.1	4.6	5.1	5.4	5.2	6.0
$m_1^\perp$	9.4	25.6	43.9	5.0	4.3	4.3	5.1	4.6	5.1	5.4	5.2	6.0
$m_2^\perp$	9.4	25.6	43.9	5.0	4.3	4.3	5.1	4.6	5.1	5.4	5.2	6.0
$m_3^\perp$	9.5	24.3	40.6	5.0	4.2	4.2	5.1	4.6	5.1	5.3	5.2	6.0
$m_4^\perp$	9.6	22.6	36.0	5.0	4.3	4.2	5.1	4.6	5.1	5.3	5.2	6.1
$m_j^\perp$	14.5	51.6	85.5	6.2	8.1	6.8	6.0	6.6	6.4	6.7	7.5	7.1
S_{Bai}^T	46.5	99.6	100.0	1.7	2.3	2.5	1.4	2.2	2.2	1.6	2.2	2.5
S_{Bai}	69.1	100.0	100.0	4.3	6.6	6.5	3.7	6.4	5.9	4.0	6.0	6.5

Note: The data are i.i.d. from a $\chi^2(7)$ distribution standardized in order to have a zero mean and a variance equal to the one of a T(ν) with $\nu=5$ (Panel A) or $\nu=20$ (Panel B). The data are also drawn from a mixture of two normal variables with weights p and $1-p$ where p equals 0.7, or 0.8, or 0.9. The standard deviations of the normal distributions are computed in order to match the second and fourth moments of a T(ν) with $\nu=5$ (Panel A) or $\nu=20$ (Panel B). ν is estimated by the second moment of x_t. The results are based on 10,000 replications. For each sample size, we provide the percentage of rejection at a 5% level. m_α, $m_\alpha^\perp$, $m_j^\perp$, S_{Bai}^T and S_{Bai} are defined in Table 1. The projections and the variances are computed in sample.

Table 3: Size and Power under serial correlation

Size properties

	$\nu = 5$						$\nu = 20$					
	$\rho = 0.4$			$\rho = 0.9$			$\rho = 0.4$			$\rho = 0.9$		
T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	4.8	5.2	5.3	5.9	8.4	8.5	3.2	5.3	5.2	4.5	4.0	4.7
$m_{\frac{1}{2}}^\perp$	4.8	5.4	5.1	6.6	9.1	9.0	3.3	5.4	5.3	4.9	4.1	4.7
$m_1^\perp$	4.8	5.5	5.2	6.8	9.2	9.1	3.4	5.5	5.3	5.0	4.1	4.7
$m_2^\perp$	4.8	5.5	5.2	6.8	9.2	9.1	3.4	5.5	5.3	5.0	4.1	4.7
$m_3^\perp$	4.9	5.3	5.2	6.7	9.1	9.0	3.4	5.4	5.3	5.0	4.2	4.8
$m_4^\perp$	4.9	5.2	5.3	6.6	8.7	8.8	3.3	5.3	5.2	5.1	4.2	4.9
$m_j^\perp$	5.7	5.4	5.5	5.6	8.0	7.9	4.8	6.7	6.3	2.8	4.5	4.9

Power against mixture of normals $p = 0.7$.

T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	79.9	100.0	100.0	7.2	7.4	6.1	9.2	33.5	58.9	12.4	21.9	32.5
$m_{\frac{1}{2}}^\perp$	75.4	100.0	100.0	22.1	72.3	94.3	10.4	36.6	63.0	51.3	99.0	100.0
$m_1^\perp$	73.0	100.0	100.0	38.2	93.2	99.8	10.9	38.1	64.5	72.7	100.0	100.0
$m_2^\perp$	73.0	100.0	100.0	38.2	93.2	99.8	10.9	38.1	64.5	72.7	100.0	100.0
$m_3^\perp$	73.9	100.0	100.0	33.5	86.8	98.9	10.5	36.6	62.9	62.9	99.7	100.0
$m_4^\perp$	73.2	100.0	100.0	28.8	77.2	96.2	9.9	34.4	59.7	52.1	98.8	100.0
$m_j^\perp$	67.9	100.0	100.0	55.0	100.0	100.0	11.6	41.8	68.8	91.6	100.0	100.0

Power against mixture of normals $p = 0.9$.

T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	8.0	24.1	41.9	10.1	9.4	9.4	10.6	40.1	67.6	12.8	21.2	32.1
$m_{\frac{1}{2}}^\perp$	8.4	22.1	37.9	41.3	96.5	100.0	11.7	43.3	70.9	50.2	99.2	100.0
$m_1^\perp$	8.2	20.8	35.4	67.6	99.9	100.0	12.3	44.5	72.1	71.7	100.0	100.0
$m_2^\perp$	8.2	20.8	35.4	67.6	99.9	100.0	12.3	44.5	72.1	71.7	100.0	100.0
$m_3^\perp$	8.5	21.3	36.4	57.9	99.5	100.0	11.8	43.1	70.6	62.5	99.9	100.0
$m_4^\perp$	8.5	21.1	35.8	47.1	97.8	100.0	11.2	40.6	67.8	51.3	99.0	100.0
$m_j^\perp$	7.7	19.7	35.4	89.0	100.0	100.0	13.1	46.6	73.0	91.8	100.0	100.0

Note: x_t is given by $x_t = u_t/\sqrt{s_t}$ where the variables u_t and s_t are independent, the distribution of u_t is $\mathcal{N}(0,1)$ while s_t follows a gamma ($\nu/2$,$2/\nu$,0) distribution, where ν equals 5 or 20. Consequently, the unconditional distribution of x_t is $\mathrm{T}(\nu)$; s_t is i.i.d. while the conditional distribution of u_t given its past is $\mathcal{N}(\rho u_{t-1}, 1-\rho^2)$ where ρ equals 0.4 or 0.9. For the power properties, $x_t = \rho x_{t-1} + \varepsilon_t$ where ε_t follows a mixture of two normal variables with weights p and $1-p$ for p equals 0.7 or 0.9. The standard deviations of the two normal variables are computed such that the second and fourth moment of a mixture matches those of a $\mathrm{T}(\nu)$ where ν equals 5 or 20. We test that the stationary distribution of x_t is a $\mathrm{T}(\nu)$. ν is estimated by the second moment of x_t. We take into account the serial correlation by estimating the variance matrix through a HAC procedure (Andrews (1991)). The results are based on 10,000 replications. For each sample size, we provide the percentage of rejection at a 5% level. The notations $m_\alpha^\perp$ and $m_j^\perp$ are defined in Table 1.

Table 4: Size and Power with GARCH(1,1) DGP

Size				Power against mixture of normals							
				$p = 0.7$				$p = 0.9$			
Panel A: $\nu = 5$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
$m_0^\perp$	6.8	7.3	6.2	$m_0^\perp$	92.4	100.0	100.0	$m_0^\perp$	6.8	28.8	60.9
$m_{\frac{1}{2}}^\perp$	6.5	6.9	6.2	$m_{\frac{1}{2}}^\perp$	90.9	100.0	100.0	$m_{\frac{1}{2}}^\perp$	6.8	30.5	62.5
$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—
$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—
$m_3^\perp$	5.8	6.1	5.8	$m_3^\perp$	86.0	100.0	100.0	$m_3^\perp$	6.4	23.1	45.9
$m_4^\perp$	5.6	6.1	5.8	$m_4^\perp$	84.4	100.0	100.0	$m_4^\perp$	6.3	19.5	38.3
$m_j^\perp$	4.9	8.3	7.1	$m_j^\perp$	90.7	100.0	100.0	$m_j^\perp$	6.5	22.8	50.6
S_{Bai}^T	2.0	4.9	6.2	S_{Bai}^T	45.5	98.9	100.0	S_{Bai}^T	4.7	13.1	18.1
S_{Bai}	2.8	8.3	11.0	S_{Bai}	48.3	99.2	100.0	S_{Bai}	5.9	17.2	23.8
Panel B: $\nu = 20$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
$m_0^\perp$	2.4	5.2	5.9	$m_0^\perp$	2.3	5.0	5.9	$m_0^\perp$	2.5	4.0	5.0
$m_{\frac{1}{2}}^\perp$	2.5	5.2	5.8	$m_{\frac{1}{2}}^\perp$	2.3	4.9	5.8	$m_{\frac{1}{2}}^\perp$	2.5	4.0	5.1
$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—
$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—
$m_3^\perp$	2.5	5.0	5.5	$m_3^\perp$	2.4	4.8	5.5	$m_3^\perp$	2.7	4.0	5.1
$m_4^\perp$	2.6	5.0	5.2	$m_4^\perp$	2.4	4.7	5.3	$m_4^\perp$	2.7	3.9	5.2
$m_j^\perp$	1.6	2.7	3.9	$m_j^\perp$	1.4	2.9	4.4	$m_j^\perp$	1.8	2.2	3.2
S_{Bai}^T	1.4	2.9	2.7	S_{Bai}^T	1.5	2.6	2.8	S_{Bai}^T	1.6	3.0	3.1
S_{Bai}	2.2	5.3	5.3	S_{Bai}	2.3	5.0	5.4	S_{Bai}	2.2	5.2	5.4

Note: x_t is a GARCH(1,1) process : $x_t = \mu + \sqrt{v_t}u_t$ with $v_t = \omega + \tilde{\alpha}(x_t - \mu)^2 + \beta v_{t-1}$ where $\mu = 0$, $\omega = 0.2$, $\tilde{\alpha} = 0.1$, and $\beta = 0.8$. For the size properties, u_t follows a T(ν) up to a scale parameter with ν equals 5 or 20 such that $Var[u_t] = 1$. For the power properties, u_t follows, up to a scale parameter, a mixture of two normal with weights p and $1 - p$ where p equals 0.7 or 0.9. The standard deviations of the two normal variables are computed in order to match the second and fourth moments of a T(ν) where $\nu = 5, 20$. The scale parameter guarantees that $Var[u_t] = 1$. μ, ω, $\tilde{\alpha}$ and β are estimated by a Gaussian QML method. ν is estimated using the fourth moment of the fitted residual $\hat{u}_t$. We test that $u_t\sqrt{\frac{\nu}{\nu-2}}$ follows a T(ν). The results are based on 10,000 replications. For each sample size, we provide the percentage of rejection at a 5% level. The notations m_α, $m_\alpha^\perp$, $m_j^\perp$, S_{Bai}^T and S_{Bai} are defined in Table 1.

Table 5: Power of the tests with GARCH(1,1) DGP against asymmetric innovations.

T	100	500	1000
$m_0^{\perp}$	12.3	78.0	98.5
$m_{\frac{1}{2}}^{\perp}$	12.8	79.8	98.8
$m_1^{\perp}$	—	—	—
$m_2^{\perp}$	—	—	—
$m_3^{\perp}$	13.8	82.0	98.9
$m_4^{\perp}$	13.6	80.6	98.5
$m_5^{\perp}$	13.3	78.5	97.6
$m_{10}^{\perp}$	11.3	62.2	85.1
$m_{20}^{\perp}$	9.0	35.8	49.8
$m_j^{\perp}$	11.5	78.8	98.9
S_{Bai}^T	7.2	81.5	99.7
S_{Bai}	8.8	87.4	99.9

Note: x_t is a GARCH(1,1) process : $x_t = \mu + \sqrt{v_t}u_t$ with $v_t = \omega + \tilde{\alpha}(x_t - \mu)^2 + \beta v_{t-1}$, where $\mu = 0$, $\omega = 0.2$, $\tilde{\alpha} = 0.1$, and $\beta = 0.8$. u_t follows a $\chi^2(7)$ up to a scale parameter, which guarantees that $V[u_t] = 1$. μ, ω, $\tilde{\alpha}$, and β are estimated by a Gaussian QML method. ν is estimated using the fourth moment of fitted residuals $\hat{u}_t$. We test that the distribution of $u_t\sqrt{\frac{\nu}{\nu-2}}$ is T(ν). The results are based on 10,000 replications. For each sample size, we provide the percentage of rejection at a 5% level. The notations m_α, $m_\alpha^{\perp}$, $m_j^{\perp}$, S_{Bai}^T and S_{Bai} are defined in Table 1.

Table 6: Testing the Student distributional assumption of fitted residuals for a GARCH(1,1) model

	UK-US$		FF-US$		SF-US$		Yen-US$	
$\hat{\nu}$	9.61		9.56		6.64		5.54	
$m_0^{\perp}$	0.09754	(0.75)	1.25273	(0.26)	0.00157	(0.97)	0.00323	(0.95)
$m_{\frac{1}{2}}^{\perp}$	0.12138	(0.73)	1.09922	(0.29)	0.01311	(0.91)	0.01353	(0.91)
$m_3^{\perp}$	0.22614	(0.63)	0.70084	(0.40)	0.45082	(0.50)	0.21660	(0.64)
$m_4^{\perp}$	0.23585	(0.63)	0.66540	(0.41)	0.71038	(0.40)	0.24267	(0.62)
$m_j^{\perp}$	0.40240	(0.82)	1.77873	(0.41)	1.81173	(0.40)	1.11926	(0.57)
S_{Bai}^T	0.69467	(—)	1.19929	(—)	2.19812	(—)	2.27336	(—)
S_{Bai}	1.03593	(—)	1.26185	(—)	2.31280	(—)	3.02817	(—)

Note: We tested the Student assumption of the standardized residuals. The volatility model is a GARCH(1,1) and is estimated by the Gaussian QML method. We reported the test statistics and their corresponding p-values in parentheses. The data are daily exchange rate returns used by Harvey, Ruiz and Shephard (1994) and Kim, Shephard and Chib (1998). The notations m_α, $m_\alpha^{\perp}$, $m_j^{\perp}$, S_{Bai}^T and S_{Bai} are defined in Table 1. The critical values of the Bai-statistics are respectively: 1.94 (1%), 2.22 (5%) and 2.80 (10%).

References

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

Aït-Sahalia, Y., L.P. Hansen and J. Scheinkman (2003), "Operator Methods for Continuous-Time Markov Models," in Y. Aït-Sahalia and L.P. Hansen (Eds.), *Handbook of Financial Econometrics,* forthcoming.

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

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

Bahadur, R.R. (1967), "Rates of Convergence of Estimates and Test Statistics," *Annals of mathematical Statistics*, 38, 303-324.

Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, 85, 531-549.

Bai, J. and S. Ng (2005), "Tests for Skewness, Kurtosis, and Normality for Time Series Data," *Journal of Business, Economics and Statistics*, .

Bera, A. and Y. Bilias (2002), "The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis," *Journal of Econometrics*, 107, 51-86.

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

Bontemps, C. (2006), "Moment based tests for discrete distributions", in progress.

Bontemps, C. Koumtigue and N. Meddahi (2005), "Moment Based Tests of Multivariate Distributions," in progress.

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

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

Chesher, A., G. Dhaene, C. Gouriéroux, and O. Scaillet (1999), "The Bartlett Identities Tests," working paper, CREST-Paris.

Chesher, A. and R.J. Smith (1997), "Likelihood Ratio Specification Test," *Econometrica*, 65, 627-646.

Chihara, T.S. (1978), *An Introduction to Orthogonal Polynomials*, Gordon and Breach Science Publishers.

Cobb, L., P. Koppstein, and N.H. Chen (1983), "Estimation and Moment Recursions for Multimodal Distributions of the Exponential Family," *Journal of the American Statistical Association*, 78, 124-130.

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

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

Cox, J.C., J. Ingersoll and S. Ross (1984), "A Theory of the Term Structure of Interest Rates", *Econometrica*, 53, 385-408.

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.

Diebold, F.X., J. Hahn, and A.S. Tay (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Echange," *Review of Economics and Statistics*, 81, 661-673.

Diebold, F.X., A.S. Tay, and K.F. Wallis (1999), "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters," in *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, R.F. Engle and H. Whited, Eds., pp. 76-90, Oxford University Press.

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

Duan, J.C. and M.K.P. So (2001), "A Goodness-of-Fit Test Using Relative Entropy," working paper, University of Toronto.

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

Elderton, W.P. and N.L. Johnson (1969), *Systems of Frequency Curves*, Cambridge University Press.

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

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

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

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

Geweke, J. (1981), "The Approximate Slopes of Econometric Tests," *Econometrica*, 49, 1427-1442.

Hall, A.R. (2000), "Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test," *Econometrica*, 68, 1517-1527.

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

Hansen, L.P. (2001), "Generalized Method of Moments Estimation: A Time Series Perspective," *International Encyclopedia of Social and Behavioral Sciences*.

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

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

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

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

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

Johnson, N.L, S. Kotz and N. Balakrishnan (1994), *Continuous Univariate Distributions*, Vol 1, Wiley.

Karlin and Taylor (1975)

Kendall, M.G., A. Stuart and J.K. Ord (1977), *The Advanced Theory of Statistics*, London Griffin.

Khmaladze, E.V. (1981), "Martingale Approach in the Theory of Goodness-of-Tests," *Theory of Probability and its Application*, 26, 240-257.

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

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

Kitamura, Y. and M. Stutzer (1997), "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861-874.

Lejeune, B. (2002), "A Diagnostic M-Test for Distributional Specification of Parametric conditional Heteroscedasticity Models for Financial Data," CORE discussion paper 2002/24.

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

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 D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, Vol IV.

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

Ord, J.K. (1972), *Families of Frequency Distributions*, Vol 30, Griffin's Statistical Monographs and Courses, Hafner Publishing Company, New York.

Premaratne, G. and A. Bera (2001), "Modeling Asymmetry and Excess Kurtosis in Stock Return Data," working paper, University of Illinois at Urbana Champain.

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.

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.

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

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.

Wong, E. (1964), "The Construction of a Class of Stationary Markoff Processes," in Belleman, R., (Ed.), Sixteenth Symposium in Applied Mathematics, Stochastic Processes in Mathematical Physics and Engineering, American Mathematical Society, Providence, 264-276.

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