

Testing Distributional Assumptions: A GMM Approach*

Christian Bontemps[†] and Nour Meddahi[‡]

Toulouse School of Economics (GREMAQ, IDEI)

March 21, 2011

Abstract

We consider testing distributional assumptions by using moment conditions. A general class of moment conditions satisfied under the null hypothesis is derived and connected to existing moment-based tests. The approach is simple and easy-to-implement, yet reasonably powerful. In addition, we provide moment tests that are robust against parameter estimation error uncertainty in the general case which includes the case of serial correlation. In particular, we consider the location-scale model for which we derive robust moment tests, regardless of the forms of the conditional mean and variance. We study in detail the Student and Inverse Gaussian distributions. Simulation experiments are conducted to assess the finite sample properties of the tests. We provide two empirical examples on foreign exchange rates by testing the Student distributional assumption of T-GARCH daily returns and on daily realized variance by testing the Inverse Gaussian distributional assumption.

Keywords: Moment-based tests; parameter uncertainty; location-scale models; serial correlation; HAC; T-GARCH model; Pearson distributions.

JEL codes: C12, C15.

*The authors would like to thank Torben Andersen, Manuel Arellano, Russell Davidson, Jean-Marie Dufour, John Geweke, Lars Hansen, Enrique Sentana, the editor and two anonymous referees for helpful comments and participants at various conferences and seminars for helpful discussions. The first author acknowledges financial support from ANR-05-BLAN-0024-01, while the second author thanks MITACS and CREST. The two authors benefited from the financial support of the chair "Marché des risques et création de valeur" Fondation du risque/SCOR.

[†]Toulouse School of Economics (GREMAQ, IDEI), 21 Allée de Brienne, 31000 Toulouse, France; phone: +33 5 61 12 85 95, email: bontemps@cict.fr.

[‡]Toulouse School of Economics (GREMAQ, IDEI), 21 Allée de Brienne, 31000 Toulouse, France; phone: +33 5 61 12 85 63, email: meddahi@cict.fr.

1 Introduction

Recent developments in financial econometrics emphasize the importance of developing testing procedures of general distributional assumptions. These developments include Value-at-Risk calculations (Christoffersen, 1998), density forecasts (Diebold, Gunter and Tay, 1998), continuous time modeling of interest rates (Aït-Sahalia, 1996, and Conley, Hansen, Luttmer and Scheinkman, 1997), and modeling realized volatility (Forsberg and Bollerslev, 2002). The main goal of the paper is to develop simple and easy-to-implement, yet reasonably powerful, tests of continuous distributions, when one faces statistical issues like parameter estimation error uncertainty and possible serial correlation of the data.

A common and popular approach to test normality of economic variables is to test whether some *ad hoc* empirical moments of the data, often the third and fourth moments, fit well with their theoretical counterparts. The method of moments leads to a statistic which, in the case of the third and fourth moments, is asymptotically $\chi^2(2)$ distributed. However, the variable of interest is often unobservable, e.g. the disturbance errors in a regression model. Consequently one often uses the fitted residuals instead of the true unknown error terms in the statistic. The asymptotic distribution of this skewness-kurtosis test is generally no longer $\chi^2(2)$ distributed (Durbin's problem, 1973). The literature on the method of moments (e.g. Newey, 1985a, and Tauchen, 1985) provides a correction that takes into account this parameter uncertainty. In a regression context, it coincides with the celebrated Jarque and Bera (1980) test for normality. The Jarque-Bera test has been extensively used because it is easy to interpret, simple to implement, and powerful against standard alternatives. This test can be used with residuals but not with observed data for which the empirical mean and variance does not equal zero and one respectively; see Bontemps and Meddahi (2005).¹ These authors prove that Hermite polynomials are robust against parameter estimation error uncertainty when one considers a location-scale model and tests normality. The test statistic based on the third and fourth Hermite polynomials is asymptotically $\chi^2(2)$ distributed whether one uses the (generally unknown) error terms or the fitted residuals. Likewise, serial correlation can be considered by computing the long-run variance matrix of the moments in a GMM framework (Hansen 1982), as in Richardson and Smith (1992).

The goal of this paper is to extend Bontemps and Meddahi (2005) to any continuous distribution. Let x be a continuous random variable with some assumed probability density function that one wants to test. Moment techniques will try to figure out whether the empirical counterpart of $E[h(x)]$ equals (asymptotically) its theoretical value, for a function $h(\cdot)$ chosen by the econometrician, like the third and fourth moments in the normal case. Of course, one needs to compute the expected value of $h(x)$ in order to conduct the test. It can be done theoretically or by simulations depending on the complexity of the considered function. In

¹Jarque-Bera test is however valid for some examples studied in Fiorentini, Sentana, and Calzolari (2004).

this paper, we derive, under mild regularity assumptions, a class of moment conditions for which the expectation equals zero by construction. Importantly, this class encompasses any regular moment and, hence, all moments traditionally used by empirical researchers. Moreover, this class of moments coincides with one derived by Hansen and Scheinkman (1995) when the variable of interest is a continuous time process and is related to Pearson's contribution (see Hansen, 2001, and Section 2).

We consider the case where the variable of interest is unobservable and/or where its distribution may involve some unknown parameters that have to be estimated. We know that, like for the normal case, the parameter estimation error uncertainty will affect the asymptotic distribution of the test. Usually, one corrects the asymptotic variance of the moments. In our paper, we provide conditions under which the asymptotic distribution of our moments does not depend on the fact that one uses the unknown true values or the estimated ones (provided that the estimators are square-root consistent). We call such moments, robust moments, (to the parameter estimation error uncertainty). We also propose a solution to transform any moment into a robust one. This consists of the projection of the original moment on the orthogonal of the space spanned by the score function.

Interestingly, a robust moment in an i.i.d. context is also robust in a serially correlated one. This is an attractive feature of our approach as the solution which consists in correcting the variance of the moment taken at the estimated parameters can be cumbersome. Moreover, it is generally difficult to analytically compute the long-run variance matrix in this case. One can, however, use the Heteroskedastic-Autocorrelation-Consistent (HAC) methods of Newey and West (1987) and Andrews (1991) to estimate it. Finally, we show that robust moments in a location-scale model with constant mean and variance, for which the score function can be computed easily, are indeed robust whatever the specification of the conditional mean and variance, including ARMA/GARCH forms.

An alternative method to test a continuous distribution is to transform it (under the null hypothesis) into a normal one (e.g. Lejeune, 2002, and Duan, 2003) or a uniform one (Diebold, Gunter, and Tay, 1998). This method has several drawbacks. A rejection of the null hypothesis does not inform how one should change the model.² More importantly, handling the parameter estimation error uncertainty is more cumbersome with the transformed data. For instance, Hermite polynomials are no longer robust when one uses the normal transformation.

There is a trade-off between simplicity and consistency, i.e. having power against any alternative. The Jarque-Bera test has become popular because of its simplicity. However, like any M-test, it is inconsistent. It does not have any power against a distribution which has the same first four moments as those of the standard normal distribution. Our tests are also

²For instance, if one rejects normality of the transformed variable due to the presence of skewness, one cannot derive a general conclusion about the asymmetry of the original variable.

inconsistent as we base our test statistic on a finite number of moments.³ In order to assess the power properties of our tests, we consider two important examples from financial econometrics: the Student distribution (in a GARCH framework) and the Inverse Gaussian distribution (in a realized variance setting). Simulations suggest that the small sample properties are good (both in terms of size and power) and that the proposed robust moments are powerful against common alternative assumptions.

The rest of this paper is organized as follows. Section 2 provides a literature review. Section 3 introduces and studies the moment conditions of interest. The parameter estimation error uncertainty is studied in Section 4. In particular we pay attention to the location-scale model which is considered in a separate subsection. Section 5 provides simulations to assess the performance, simplicity and easy-to-implement properties of our tests for the Student and Inverse Gaussian cases. Two empirical examples are provided in Section 6, while Section 7 concludes. All the proofs and calculations are provided in the Appendices.

2 Literature review

2.1 Pearson family of distributions and their generalizations

In this subsection, we briefly review how moment-based tests have been used for financial applications in relation to Pearson distributions.

2.1.1 The Pearson family of distributions

At the end of the nineteenth century Karl Pearson introduced his famous family of distributions that extends the classical normal distribution. If a distribution with a probability density function (p.d.f. hereafter) $q(\cdot)$ 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) :

$$\frac{q'(x)}{q(x)} = \frac{A(x)}{B(x)} = \frac{-(x+a)}{c_0 + c_1x + c_2x^2}. \quad (2.1)$$

The Pearson family includes, as special examples, the Normal, Student, Gamma, Beta, and Uniform distributions.⁴

An important result derived by Pearson is the following recursive formula involving the moments of the distribution:

$$(c_2(j+2) - 1)E[X^{j+1}] = (a - c_1(j+1))E[X^j] - c_0jE[X^{j-1}], \quad \forall j \geq 1. \quad (2.2)$$

Pearson uses Eq. (2.2), for $j = 1, \dots, 4$, to write the parameters a , c_0 , c_1 and c_2 as functions of $E[X^j]$ and then provides an estimator using the empirical counterpart of the moments (under

³One can extend our approach to test an infinite number of moments and, then, get consistent tests. This extension is beyond the contribution of this paper which focuses on simple and easy-to-implement methods.

⁴For more details, see Johnson, Kotz and Balakrishnan (1994), pages 15-25.

the assumption that these moments exist). This is the introduction of the method of moments (see Bera and Biliias, 2002, for a historical review).

Eq. (2.2) could also be used for testing purposes. Stein (1972), for example, uses it to characterize the standard normal distribution (see Bontemps and Meddahi, 2005).

2.1.2 Scalar diffusions

Wong (1964) makes a connection between Pearson distributions and some diffusion processes. He provides stationary continuous time models for which the marginal density is a Pearson distribution. We recap here some results from Hansen and Scheinkman (1995). Assume that the random variable x_t is a stationary scalar diffusion process characterized by the stochastic differential equation

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \quad (2.3)$$

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

$$q(x) = K\sigma^{-2}(x) \exp\left(\int_z^x \frac{2\mu(u)}{\sigma^2(u)} du\right), \quad (2.4)$$

where z is a real number in (l, r) and K is a scale parameter such that the function $q(\cdot)$ integrates to one (see also Aït-Sahalia, Hansen and Scheinkman, 2010, for a review of all the properties of the diffusion processes considered here).

Hansen and Scheinkman (1995) provide two sets of moment conditions related to the marginal and conditional distributions of x_t respectively. For the marginal distribution, they show that

$$E[\mathcal{A}g(x_t)] = 0, \quad (2.5)$$

where the function g is assumed to be twice differentiable and square-integrable with respect to the marginal distribution of x_t and \mathcal{A} is the infinitesimal generator associated with the diffusion (2.3),

$$\mathcal{A}g(x) = \mu(x)g'(x) + \frac{\sigma^2(x)}{2}g''(x). \quad (2.6)$$

One limitation of the Pearson distributions is the shape of their p.d.f., as they cannot have more than one mode. For this reason, Cobb, Koppstein and Chen (1983) extend the Pearson system by allowing $A(\cdot)$ in Eq. (2.1) to be a polynomial of degree higher than one and, hence, generate multimodal distributions. This extension has been exploited by Hansen and Scheinkman (1995), Aït-Sahalia (1996) and by Conley, Hansen, Luttmmer and Scheinkman (1997), for modeling the short-term interest rate whose marginal distribution looks like a bimodal distribution. These authors strongly reject Pearson unimodal distributions.

2.2 Orthogonal Polynomials

For any distribution, one can build orthogonal (or orthonormal) polynomials by using a Gram-Schmidt method. In the case of Pearson distributions, these polynomials have a simple form that one can get from the so-called Rodrigue's formula

$$P_n(x) = \alpha_n \frac{1}{q(x)} [B^n(x)q(x)]^{(n)}, \quad (2.7)$$

where $f^{(n)}(\cdot)$ denotes the n -th derivative function of any function $f(\cdot)$ and α_n is a scaling parameter, which could be chosen to normalize the variance of P_n for any n .

Interestingly, the polynomials $P_n(x)$ in Eq. (2.7) are also eigenfunctions of the infinitesimal operator \mathcal{A} in Eq. (2.6). When all the polynomials P_n are square-integrable with respect to the p.d.f. $q(\cdot)$, like for the Normal, Gamma, Beta or Uniform distributions,⁵ one can prove that this sequence is dense in $L^2([l, r])$, i.e. any square-integrable function may be expanded onto the polynomials P_n , $n = 0, 1, 2$, etc. In this case, the p.d.f. of a random variable x equals $q(\cdot)$ if and only if⁶

$$\forall n \geq 1, E[P_n(x)] = 0.$$

This result means that for testing purposes, one could focus on these orthogonal polynomials. Appendix A provides a summary of the orthonormal polynomial families for the following well-known distributions: Normal, Student, Gamma, Beta and Uniform; see Schoutens (2000) for more details.

2.3 Serial Correlation

Two leading examples of the recent development in the financial literature emphasize the importance of developing distributional test procedures that are valid in the presence of serial correlation in the data.

The first one is modeling continuous time Markov models, particularly the short term interest rate. Aït-Sahalia (1996) and Conley, Hansen, Luttmer and Scheinkman (1997) develop a specification test by testing whether the marginal distribution of the data coincides with the one implied by the specification of the scalar diffusion. Aït-Sahalia (1996) compares the nonparametric estimator of the density function with its theoretical counterpart while Conley, Hansen, Luttmer and Scheinkman use the moment conditions (2.5). Both papers use a HAC procedure (Newey and West, 1987; Andrews, 1991) in the implementation of their tests. Using such a procedure for testing serially correlated data has been done by Richardson and Smith (1993), Bai and Ng (2005), Bontemps and Meddahi (2005), and Lobato and Velasco (2004) in the context of normality.

⁵The problem of non-existence of such a family could occur for heavy-tailed distributions. The Student distribution is one example.

⁶For a formal proof, see Gallant (1980, Theorem 3, page 192).

The second example is the evaluation of density forecasts developed by Diebold, Gunter and Tay (1998) in the univariate case, and by Diebold, Hahn and Tay (1999) in the multivariate case. These papers highlight the importance of testing distributional assumptions for serially correlated data. This evaluation is indeed 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 have different implications for the specification of the model. When the model is rejected, one would like to have test procedures which can detect which assumption is wrong (both or only one).

3 Test functions

3.1 Moment conditions

Let x be a random variable with a p.d.f. denoted $q(\cdot)$. We assume that the support of x is (l, r) , where l and r may be finite or not, and that the function $q(\cdot)$ is differentiable on (l, r) . Consider a differentiable function $\psi(\cdot)$ (we call it from now a test function) such that its derivative function, $\psi'(\cdot)$, is integrable with respect to $q(\cdot)$.

Assumption L(imits): $\lim_{x \rightarrow l} \psi(x)q(x) = 0$ and $\lim_{x \rightarrow r} \psi(x)q(x) = 0$.

Assumption L is not very restrictive. In the normal case, any polynomial function satisfies this assumption.

Proposition 3.1 *Let $m(\cdot)$ be the function defined by*

$$m(x) = \psi'(x) + \psi(x)(\log q)'(x). \quad (3.1)$$

Under assumption L,

$$E[m(x)] = E[\psi'(x) + \psi(x)(\log q)'(x)] = 0. \quad (3.2)$$

Conversely, let $m(\cdot)$ be an integrable function with respect to the density function $q(\cdot)$ such that

$$Em(x) = 0. \quad (3.3)$$

Then, the function $\psi(\cdot)$ defined by

$$\psi(x) = \frac{1}{q(x)} \int_l^x m(u)q(u)du \quad (3.4)$$

satisfies assumption L and is the test function that generates $m(x)$ in Eq. (3.1).

Proof. The proof is given in Appendix B.1 ■

The first part of the proposition gives a class of restrictions that a random variable with a density function $q(\cdot)$ should satisfy. It is the basis of our testing approach.

The second part of the proposition shows that one does not lose any generality by focusing on moments defined by Eq. (3.1). Given that any integrable moment condition which satisfies Eq. (3.3) can be written as in Eq. (3.1) with some particular test function $\psi(\cdot)$, the informational content of the class of moment conditions (3.2) is substantial. In particular, it encompasses the score and quantile functions, the moment conditions related to the so-called information-matrix test (White, 1982) and its generalizations, i.e. the Bartlett identities tests (Chesher, Dhaene, Gouriéroux and Scaillet, 1999).

The moment condition (3.2) is written marginally but it holds also when one considers a conditional model given a variable z :

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

where $\psi(x, z)$ is a test function that satisfies Assumption L and $q(x, z)$ is the conditional p.d.f. of x given z . A feasible test statistic can be based on unconditional moments of the form

$$E \left[w(z) \left(\frac{\partial \psi(x, z)}{\partial x} + \frac{\psi(x, z)}{q(x, z)} \frac{\partial q(x, z)}{\partial x} \right) \right] = 0, \quad (3.5)$$

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

It is worth noting that Eq. (2.2) from Karl Pearson is exactly Eq. (3.2) with $\psi(x) = x^j B(x)$. We haven't found in the literature a systematic use of Eq. (3.2) for any distribution except for Chen, Hansen and Scheinkman (2009) who explicitly use it in the multivariate continuous time processes, and in Hansen (2001) who implicitly uses it in the case of scalar diffusion processes. In this context, Eq. (2.5) of Hansen-Scheinkman is simply Eq. (3.2) with the test function $\psi = (g'\sigma^2)$.

We propose in this paper to choose some particular test functions $\psi(\cdot)$ and to use the moments $m(\cdot)$, derived by Eq. (3.2), for testing our distributional assumption. The optimal choice of $w(\cdot)$ in (3.5) and $\psi(\cdot)$ is beyond the scope of this paper and is studied in Bontemps and Meddahi (2011). Our tests, like every M-tests, are not consistent, though optimal against some given directions (see Chesher and Smith, 1997). This paper however highlights moment tests that are simple to implement, but still have good power properties against common alternative models, as corroborated by our simulation results.

3.2 Asymptotic distribution of the test statistic

We discuss now the asymptotic distribution of the test statistic derived from Eq. (3.2). Consider a sample x_1, \dots, x_T of the variable of interest, x_t , where there may exist some serial correlation. The process $(x_t)_{t \in \mathbb{Z}}$ is assumed to be a stationary process. Let $\psi_1(\cdot), \dots, \psi_p(\cdot)$,

be p differentiable test functions satisfying assumption L. Let $m(x_t)$ be the p -vector whose components are $\psi'_i(x_t) + \psi_i(x_t)(\log q)'(x_t)$, $i = 1, 2, \dots, p$. Eq. (3.2) implies

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

Throughout the paper, we assume that the long-run variance matrix of $m(x_t)$, Σ , given by $\Sigma = \sum_{h=-\infty}^{+\infty} E[m(x_t)m(x_{t-h})^\top]$, is well-defined and positive definite (throughout the paper, $^\top$ denotes the matrix transpose operator). In the context of time series, this assumption rules out long memory processes. Under some regularity conditions (Hansen, 1982), we know that

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T m(x_t) \right)^\top \Sigma^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T m(x_t) \right) \xrightarrow[T \rightarrow \infty]{d} \chi^2(p). \quad (3.6)$$

A feasible test procedure requires the knowledge of the matrix Σ or a consistent estimator. There are cases where one can explicitly compute the matrix Σ . When the data are i.i.d. and are distributed according to a Pearson distribution, particular choices for $m(\cdot)$ are the orthonormal polynomials associated with the distribution (see Section 2.2 and Appendix A). In this case, Σ is simply the identity matrix (see Bontemps and Meddahi, 2005, for the normal case with Hermite polynomials). When the data are dependent, Σ can also be diagonal for some particular time series processes, in particular for any scalar diffusion process whose marginal distribution is among the Pearson family and whose drift is affine. This is the case for the AR(1) Normal model (Bontemps and Meddahi, 2005), the square-root process of Cox, Ingersoll and Ross (1984), for which the marginal distribution is a Gamma one, and for the Jacobi diffusion (Karlin and Taylor, 1981, page 335) associated to the Beta distribution (see Gouriéroux and Jasiak (2006) for financial applications).

However, in some i.i.d. cases and in most of the serially correlated cases, deriving Σ explicitly is difficult. One can therefore use any consistent estimator $\hat{\Sigma}_T$ of Σ like the HAC estimator proposed by Newey-West (1987) or Andrews (1991).

4 Parameter estimation error uncertainty

A probability density function generally involves some unknown parameters. Moreover the variable of interest x may be unobservable and the function which relates it to observable variables may involve some unknown parameters. All these parameters need to be estimated before testing the distributional assumption.

It is well known from the GMM literature that the asymptotic distribution of the feasible test statistic based on (3.6) is generally different from the infeasible one that uses the true (unknown) parameters. The problem is traditionally solved by correcting the test statistic (see Newey, 1985b, Tauchen, 1985).

An alternative solution is to transform the moments into robust ones, moments for which the asymptotic distribution of the feasible and infeasible test statistics coincide. There exist different transformations like the solution proposed by Wooldridge (1990) in a conditional context or the one proposed by Duan (2003). Bontemps and Meddahi (2005) prove that Hermite polynomials are robust when one considers location-scale models in the Gaussian case.

In this paper we follow the last approach. We characterize robust moments and propose a transformation to build from any moment such a robust one.

We assume that the p.d.f. depends on a parameter β . In addition, we assume that the variable of interest x_t is related to the observable variables, y_t , through a one-to-one function h_t which can depend on some parameter⁷ $\phi = (\theta, \beta)$ where θ is an additional parameter:

$$x_t = h_t(y_t, \phi^0). \quad (4.1)$$

The function h_t is indexed by t in order to summarize the possibility of having explanatory variables z_t which can be part of the relation between x_t and y_t , such as in a regression model.

Assume one wants to use a moment $m(\cdot, \beta)$ such that $E[m(x_t, \beta^0)] = 0$ to test the distributional assumption. In practice, one uses the estimates $\hat{\beta}$ and $\hat{\theta}$ of the unknown true values to work with the function

$$g_t(y_t, \hat{\phi}) = m(h_t(y_t, \hat{\phi}), \hat{\beta}). \quad (4.2)$$

Assuming that the estimators are square-root consistent ones, we now characterize the conditions under which the moment $g_t(\cdot, \phi)$ (or equivalently $m(\cdot, \beta)$) is robust and propose a construction to transform a given moment into a robust one.

4.1 Orthogonality to the score function

From now, we assume, without loss of generality, that h_t involves some conditioning variables z_t (which can be exogeneous variables and/or past values of y_t). The case without z_t is treated in Appendix B.4. A standard Taylor expansion proves that a moment $g_t(\cdot, \phi)$ is robust when

$$P_g = E \left[\frac{\partial g_t}{\partial \phi^\top}(y_t, \phi^0) \right] = 0, \quad (4.3)$$

where the expectation is taken with respect to the joint distribution of y_t and z_t (see Eq. (B.4) in Appendix B.2). The next proposition uses the generalized information equality to characterize moments which satisfy Eq. (4.3).

Proposition 4.1 *Let $s_t(y_t, \phi)$ be the conditional (on z_t) score function of the observable y_t . A moment $g_t(\cdot, \phi)$ is robust when it is orthogonal to this score function:*

$$E[g_t(y_t, \phi^0) s_t^\top(y_t, \phi^0)] = 0 \quad (4.4)$$

⁷In the following, the superscript 0 is our notation for the unknown true value.

In practice, one uses the moment $m(\cdot)$. Given that

$$E[g_t(y_t, \phi^0) s_t^\top(y_t, \phi^0)] = E[m(x_t, \beta^0) s_t^\top(h_t^{-1}(x_t, \phi^0), \phi^0)],$$

Eq. (4.4) implies that the moment $m(\cdot)$ is robust when

$$E[m(x_t, \beta^0) s_t^\top(h_t^{-1}(x_t, \phi^0), \phi^0)] = 0.$$

The next proposition addresses the issue when Eq. (4.4) does not hold.

Proposition 4.2 *Let $m(\cdot, \beta)$ a moment whose expectation under the null equals zero. A robust moment to the parameter estimation error uncertainty is given by*

$$m^\perp(x_t, \phi) = m(x_t, \beta) - E_t[m(x_t, \beta) s_t^\top(h_t^{-1}(x_t, \phi), \phi)] (V_t s_t^\top(h_t^{-1}(x_t, \phi), \phi))^{-1} s_t(h_t^{-1}(x_t, \phi), \phi), \quad (4.5)$$

where E_t and V_t denote the expectation and variance relative to the conditional distribution $y_t|z_t$.

This proposition states that when the moment of interest $m(\cdot)$ (or $g_t(\cdot)$) is not robust, one can transform it into a robust one by projecting it on the score function and by taking the residual as the new moment. It is of interest to note that Bai's (2003) method, which uses the martingale approach of Khmaladze (1981) to transform a process into a martingale one, is a similar one.

4.2 Wooldridge's approach and related methods

There are many transformations, in the literature, of the original moments $m(\cdot)$ which can lead to robust moments. For example, let S be a matrix such that $SP_g = 0$ and define the new moment $n(x_t, \beta) = Sm(x_t, \beta)$. A similar argument than the one in the previous section proves that this new moment $n(\cdot)$ is robust. This approach is however not always possible. In particular, the dimension of $m(\cdot)$ should exceed the dimension of ϕ . In this case, when one assumes that P_g has a full rank, Wooldridge (1990) proposes⁸

$$S = I_p - P_g[P_g^\top P_g]^{-1} P_g^\top. \quad (4.6)$$

Eq. (4.4) proves that any robust moment is orthogonal to the score function. We propose here a particular projection different from the transformations proposed by Wooldridge (1990) or Duan (2003). When h_t , the link between y_t and x_t , involves some conditioning variables z_t , Wooldridge changes the instruments (a function of z_t) to ensure the orthogonality condition (4.4). For our part, we project the moment onto the orthogonal space to the one spanned by the conditional score.

⁸Observe that the solution Eq. (4.6) is not unique, i.e. when one has more structure on the model, one can derive other matrices S such that $SP_g = 0$ like in Duan (2003).

In a context of MLE, our test statistic is exactly the correction derived in Newey (1985b) and Tauchen (1985) (see Appendix B.5). However, in other contexts, correcting the statistic can be difficult to do because of complications in calculating the likelihood function. Our approach can, in these cases, provide a statistic which is simple to derive and which takes into account the parameter estimation error uncertainty, though there may exist a loss of power with respect to the methods of Newey (1985b) and Tauchen (1985); see Khmaladze and Koul (2004). This attractiveness is now characterized in the special context of location-scale models. We prove that robust moments of constant location-scale models are still robust when one has a varying conditional mean or variance as in ARMA/GARCH models.

4.3 Location-scale model

Let us assume that we want to test whether y_t is distributed according to a given parametric distribution with p.d.f. $q(\cdot, \theta)$ up to a location and a scale parameter:

$$\exists \mu^0, \sigma^0, \theta^0 \in \mathbb{R}^2 \times \Theta \text{ such that } y_t = \mu^0 + \sigma^0 x_t,$$

where the p.d.f. of x_t is $q(x_t, \theta^0)$.

Let $m(\cdot, \theta)$ be a moment such that

$$Em(x_t, \theta^0) = 0.$$

We apply Proposition 4.2 to transform any moment into a robust one to the estimation of μ , σ and θ .

It shows that $m(\cdot, \theta)$ is robust if it is orthogonal to the three functions $\frac{\partial \log q}{\partial x}(x_t, \theta)$, $x \frac{\partial \log q}{\partial x}(x_t, \theta)$, $\frac{\partial \log q}{\partial \theta}(x_t, \theta)$ (see Appendix B.6). Otherwise, we can build a robust moment $m^\perp(x_t, \theta)$ by projecting $m(x_t, \theta)$ onto the orthogonal space to the one spanned by the last three functions.

Interestingly, the moment $m^\perp(\cdot, \theta)$ is also robust for any specification of the location and scale. Assume, for example, that $y_t = m_t(\phi) + \sigma_t(\phi)x_t$, where θ is part of ϕ . Under differentiability assumptions on $m_t(\phi)$ and $\sigma_t(\phi)$, we can derive similarly the score function $s(y, \phi)$:

$$s(y, \phi) = -\frac{\frac{\partial \mu_t}{\partial \phi}(\phi)}{\sigma(\phi)} \frac{\partial \log q}{\partial x}(x, \theta) - \frac{\partial \log \sigma_t}{\partial \phi}(\phi) \left(1 + x \frac{\partial \log q}{\partial x}(x, \theta)\right) + \frac{\partial \log q}{\partial \phi}(x, \theta). \quad (4.7)$$

This score function is a linear combination (up to the constant term $-\frac{\partial \log \sigma}{\partial \phi}(\phi)$) of the three functions used in the previous location-scale model. A moment orthogonal to these three functions is therefore orthogonal to this new score function.

In the Normal case, $q(\cdot)$ is the p.d.f. of the standard normal distribution $\frac{\partial \log q}{\partial x}(x, \theta) = -x$. Any moment orthogonal to x and $x^2 - 1$ is therefore robust to the parameter estimation error

uncertainty, independently of the parametric specification of μ and σ . It is for example the case of the Hermite Polynomials of order greater or equal to 3 considered in Bontemps and Meddahi (2005).⁹

The result of this section is particularly important when one wants to test a distributional assumption on some residuals, like the estimated innovations in general ARMA-GARCH processes. In many cases, the model is estimated by Gaussian QMLE for simplicity or tractability. The exact correction can be difficult to compute. Our result states that we can use the robust moments in the constant mean and variance case to test the distribution. This moment is robust to any specification of the ARMA and GARCH parts.

We provide in Appendix C.2 moments which are robust in a location-scale model for testing the Student distributional assumption. These moments can therefore be used for testing any T-GARCH process independently of the specification of the volatility process.

4.4 Robust test functions

The general class of moment conditions (3.2) is given in terms of the test function $\psi(\cdot)$. It is therefore interesting to characterize the functions $\psi(\cdot)$ that lead to robust moments (which we call robust test functions). For the sake of simplicity, we omit in this subsection the dependence of ψ on the parameters β or ϕ .

Proposition 4.3 *Let $\psi(\cdot)$ be a test function which satisfies Assumption L. Assume in addition that it is also satisfied for $\psi(x_t)s_t^\top(h_t^{-1}(x_t, \phi), \phi)$ and $\frac{\partial}{\partial x}s_t(h_t^{-1}(x_t, \phi), \phi)s_t^\top(h_t^{-1}(x_t, \phi), \phi)$. Define $M = E_t[\frac{\partial}{\partial x}s_t(h_t^{-1}(x_t, \phi), \phi)\frac{\partial}{\partial x}s_t^\top(h_t^{-1}(x_t, \phi), \phi)]$ and $G = E_t[\psi(x_t)\frac{\partial}{\partial x}s_t^\top(h_t^{-1}(x_t, \phi), \phi)]$. Then, the function $\psi^*(\cdot)$ defined by*

$$\psi^*(x_t, \phi) = \psi(x_t) - GM^{-1}\frac{\partial}{\partial x}s_t(h_t^{-1}(x_t, \phi)), \quad (4.8)$$

is a robust test function.

Proposition 4.3 is the counterpart of Proposition 4.2 for $\psi(\cdot)$. A moment $m(\cdot)$ derived from Eq. (3.2) is robust if $\psi(\cdot)$ is orthogonal to the derivative of the conditional score function ($G = 0$ in Eq (4.8)). Otherwise, it is possible to project $\psi(\cdot)$ on this derivative and take the residual function for testing purposes. The advantage of using Proposition 4.3 compare to Proposition 4.2 is that, in some particular cases, it is easier to numerically handle the derivative of the score than the score itself, like in the Student case considered in Section 5.1.

4.5 Transformation and parameter uncertainty

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 purposes. For instance, in their density forecast

⁹This result was proved in the special case of the Hermite family using local expansions, specific to these polynomials, and without the general argument provided here.

analysis, Diebold, Gunter and Tay (1998) transform the variable of interest into a Uniform one. Duan (2003) and Kalliovirta (2006) transform the variable of interest into a Normal one. However, this solution has several drawbacks.

First, it is important to notice that testing some specific moment on the transformed variable has a very difficult interpretation in terms of the original variable. Furthermore, the conditions for having robustness with respect to the parameter estimation error uncertainty depend also on the transformation itself. A moment which is robust for an observable variable is generally no longer robust when the variable is the result of a transformation. Assume for example that y_t is observable and follows a distribution whose c.d.f. (conditional or unconditional) is $Q_t(\cdot, \theta^0)$ and whose p.d.f. is $q_t(\cdot, \theta^0)$. Without loss of generality, assume that we transform y_t in a standard normally distributed variable x_t :

$$x_t = \Phi^{-1} \circ Q_t(y_t, \theta^0).$$

If we know that under the null hypothesis, $Em(x_t) = 0$, the matrix P_g in Eq. (4.3) can be written:

$$P_g = E \left[-m(x_t) \frac{\partial \log q_t}{\partial \theta^\top} (Q_t^{-1}(\Phi(x_t), \theta^0), \theta^0) \right]. \quad (4.9)$$

We know from Bontemps and Meddahi (2005) that the Hermite polynomials $H_i(\cdot)$, $i \geq 3$, are robust in the case of a general regression context. However this is no longer the case when one uses the general transformation given above. It seems very complicated to derive explicitly P_g in many cases. Simulations in the Monte Carlo section highlight that estimating P_g in the sample could give poor small sample size properties.

5 Monte Carlo Evidence

In this section we provide Monte Carlo simulations to study the performances of our test procedures. We focus on two examples: the Student and the Inverse Gaussian distributions. These two distributions are also considered in the empirical section (see Section 6).

This section has several objectives. The first one is to illustrate the simplicity of the test procedures. The second one is to provide the implementation of the tests, in particular the construction of robust moments, when one uses either a test function $\psi(\cdot)$ (Proposition 4.3) or a moment condition $m(\cdot)$ (Proposition 4.2). The third objective is to study the small sample properties in terms of size and power. All the simulations are based on 10,000 replications. Three sample sizes are considered: 100, 500 and 1,000. In all of the tables, we report the rejection frequencies for a 5% significance level test.

5.1 The Student distribution

We first study the Student distribution which is often used in financial applications due to its thick tail property. Without having any prior knowledge about the degrees of freedom, ν , it seems difficult to use polynomials since we need our moments to be square-integrable. For instance, in the empirical section, the lowest value for ν is 5.54, which implies that any polynomial of degree three or higher has an infinite variance. A moment whose expectation (with respect to the Student distribution) is zero expands mainly onto rational functions (see Wong, 1964, for details). We will therefore focus on the class of moments built from the test function

$$\psi_{\alpha,\beta}(x) = \frac{x^\beta}{(x^2 + \nu)^\alpha}.$$

The corresponding moments are:

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

Observe that even values of β lead to odd moments $m_{\alpha,\beta}(\cdot)$, and, conversely. Considering even moment conditions for symmetric alternatives or odd moment conditions for asymmetric alternatives increases the power of the tests.

We consider univariate moments $m_{\alpha,\beta}(\cdot)$ based on a particular set of values of α and β . The simulation results show that most of the even/odd moments are highly correlated. The percentages of rejections are therefore quite similar in a given family. To avoid too many redundancies, we only display the results for seven moments; three are even moments with (α, β) being equal to $(0, 1)$, $(1/2, 1)$ and $(5/2, 1)$, three are odd moments with values $(1/2, 0)$, $(1, 0)$ and $(1, 2)$. The last moment is the joint moment m_j which has one even component, $m_{5/2,1}$, and one odd component, $m_{1,2}$.

5.1.1 Location-scale model

We first assume that we observe n realizations of a random variable y , y_1, \dots, y_n , which are assumed to be i.i.d. and we want to test that y follows a T-distribution in a constant location-scale model like the one in Section 4.3. The true values for μ and σ are respectively 0 and 1.

We use the first, second and fourth moments of y to estimate μ , σ and ν . The results derived from a ML estimation of these parameters are similar and therefore not provided here.

From Section 4, we know that $m_{\alpha,\beta}(x, \nu)$ is no longer robust to the parameter estimation error uncertainty when one estimates μ , σ and ν . We use Proposition 4.3 to construct a robust moment $m_{\alpha,\beta}^*(x, \nu)$ (see Appendix C for the numerical details).

We first study the size properties of our tests. We also compute the Kolmogorov-Smirnov test (denoted KS) and the test developed by Bai (2003) denoted S_{Bai} . We display here the

results when ν equals 5.¹⁰ In Table 2, we assume that μ , σ and ν are first known (in the first block of columns) and then estimated (second and third blocks of columns). For the last two blocks, we provide the tests' performances when the variances of the moments are computed theoretically and estimated in the sample.

The size performances are globally good and the rejection frequencies are close to the nominal level. Unsurprisingly, the non-robust moments $m_{\alpha,\beta}(x, \nu)$ are very bad when the parameters are estimated. This highlights that the parameter estimation error uncertainty can have a severe impact on the tests performances. When the variance of the moments are computed empirically (third block of columns), size properties are, in this case, comparable. The KS test has good properties when the parameters are estimated but a high distortion toward under-rejection when they are estimated.¹¹ Bai's test presents severe size distortions due to the estimation of both the location parameter (as in the KS test) and the degrees of freedom of the Student distribution. These distortions could be severe even when the sample size is large.¹²

The last panel of the third block provides normality tests implemented as follows. We first transform the variable x_i into a standard Normal variable and then test the normality using three tests based on the third and fourth Hermite polynomials H_3 , H_4 and H_{3-4} (see Bontemps and Meddahi, 2005). These moments are no longer robust and therefore we have to transform them to obtain robust ones, i.e. having $P_g = 0$ in Eq. (4.9). It is also no longer possible to compute the correction analytically, implying that expectations of interest are estimated in the sample. As a result, the size properties are very poor.¹³ We recover the nominal 5% rejection rate for very large sample sizes (at least 5,000). This result is in line with similar ones found in the context of the Information Matrix test for probit models (see Orme, 1990).

In Table 3, we study the power properties of our tests against an asymmetric distribution and against the mixture of two normals. We first consider asymmetric distributions: $\chi^2(p)$ distributions with $p = 5, 15, 30$. When p increases, the percentage of rejections decreases because the $\chi^2(p)$ variable converges to a location-scale transformation of a standard normal variable, which is the limit of a $T(\nu)$ when $\nu \rightarrow +\infty$. Our joint test performs fairly well also. The results clearly highlight that our tests based on even moments have very good power, much better than the power of Bai's test which are displayed here for comparison. It seems that this test suffers from a lack of power against asymmetric alternatives.

We then consider some mixtures of two centered normals. The weights $(p, 1 - p)$ of the

¹⁰The cases $\nu = 10$ and $\nu = 20$ lead to similar results.

¹¹It is worth noting that this distortion vanishes when one does not estimate the location parameter. The KS test is therefore more sensitive to the estimation of the location than the estimation of the variance. Further simulations in the Normal and Student cases without location confirm this result. See also Boldin (1982). We are grateful to a referee for pointing out this result.

¹²Some additional simulations, not provided here, show that one recovers the nominal rejection rate when the sample size reaches 5,000 for this particular example.

¹³Consequently, we do not use these tests for the power properties.

normal distributions are respectively set to the values $(0.7, 0.3)$, $(0.8, 0.2)$, $(0.9, 0.1)$. The variances of the two distributions are chosen to fit the second and fourth moments of a $T(5)$ and a $T(20)$. When p increases, the sixth moment of the mixture distribution increases; we report in Appendix C.3 the corresponding moments as well as the theoretical variance of each component of the mixture. Of course, even moments perform better than odd ones since the expectations of odd moments are zero under the null and under the alternative. The results suggest that our tests have a very good power whatever the sample size when the variance of the true distribution fits the one of a $T(5)$, more than the one of Bai (2003). The joint test is less powerful, as it combines a symmetric and an asymmetric moment but still performs well. The power is somewhat lower when p equals 0.8.

In contrast, the power decreases significantly when the true distribution has the same variance as a $T(20)$ one. We also perform the likelihood ratio test where the critical values are computed by simulations for each sample size. We know by the Neyman-Pearson Lemma that this test is the optimal one. The simulated rejection frequencies are 6.3%, 9.6%, and 12.7% when the sample size equals 100, 500, and 1,000 respectively. It means that any test has a low power against these particular alternatives example for such sample sizes.

5.1.2 The GARCH(1,1) model with Student innovations

We now implement our test procedure for the T-GARCH(1,1) model of Bollerslev (1987). This is a popular model in empirical finance where the implied kurtosis fits empirically better the observed one than the Normal-GARCH(1,1) model. Using the results derived previously, we can implement moment-based tests quite simply while controlling the parameter estimation error uncertainty. We consider the following model:

$$y_t = \mu + \sqrt{v_t}u_t, \quad v_t = \omega + \alpha(y_{t-1} - \mu)^2 + \beta v_{t-1}, \quad u_t = \sqrt{\frac{\nu - 2}{\nu}}x_t, \quad x_t \sim T(\nu), \quad (5.2)$$

where $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$ and $\beta = 0.8$. We only present here the case $\nu = 5$; other values leading to similar results as previously.

The parameters μ, ω, α and β are estimated by Gaussian QMLE 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. We estimate ν using the fourth moment of u_t , i.e. $Eu_t^4 = 3(\nu - 2)/(\nu - 4)$. Therefore:

$$\hat{x}_t = \sqrt{\frac{\hat{\nu}}{\hat{\nu} - 2}} \frac{(y_t - \hat{\mu})}{\sqrt{\hat{v}_t}}.$$

We know from Section 4.3 that the moments used in Tables 2 and 3, $m_{\alpha, \beta}^*(x, \nu)$, are also robust in the GARCH context. The results are reported in Table 4. The size properties are quite comparable to those of Table 2. For the power analysis, we use the same distributions as in

Table 3. We observe qualitatively the same results as in the location-scale case with a slight lack of power due to the estimation of five parameters instead of three.

5.1.3 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 combined with the HAC method to estimate the variances of the moments. The HAC method is developed by using the quadratic kernel with an automatic lag selection procedure as in Andrews (1991). However, we do not perform Bai's test given that it is not valid for serial correlation cases.

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 $(\nu/2, 2/\nu, 0)$ distribution, where ν equals 5 or 20 as in our previous simulations. There is a dependence in u_t while s_t is i.i.d. We assume 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. Consequently, the unconditional distribution of x_t is $T(\nu)$ but there is serial correlation. When studying the power of the tests, we simulate an AR(1) process x_t , $x_t = \rho x_{t-1} + \varepsilon_t$, whose innovation process ε_t is, as in the previous simulations, a mixture of two normals where p equals 0.7.

The results are reported in Table 5. There exists some size distortion for $\rho = 0.9$, a case for which it is known that the HAC performs worse. Otherwise the size and power properties are similar to the ones in the previous tables.

5.2 The Inverse Gaussian Distribution

This subsection considers testing Inverse Gaussian (IG) distributions. It is common to model positive variables by lognormal distributions. Unfortunately, the lognormal distribution is not robust to temporal aggregation, i.e. the sum of independent lognormal random variables is not lognormal. The robustness to temporal aggregation could be an important property when one models time series like volatility. It turns out that this is the case for the IG distribution. Another advantage of IG distributions is in modeling conditional variance models. Specifically, we assume that the conditional distribution of a return r given the variance σ^2 is $\mathcal{N}(0, \sigma^2)$, while the unconditional distribution of σ^2 is IG. Then, the return's unconditional distribution is Normal Inverse Gaussian. Forsberg and Bollerslev (2000) used the two properties of IG distribution in order to model realized variance and daily returns. We will consider the same empirical example in the next section.

The Inverse Gaussian distribution with parameters μ, λ is defined by its p.d.f. $q(\cdot)$ on $[0, +\infty[$:

$$q(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right). \quad (5.3)$$

We can therefore construct moments using Eq. (3.2) from the test functions $\psi_k(x) = x^{k+1}$, $k \in \mathbb{Z}$:

$$m_k(x) = x^{k+1} - \frac{2\mu^2}{\lambda} \left(k - \frac{1}{2}\right) x^k - \mu^2 x^{k-1}.$$

We assume here that we observe x and test the Inverse Gaussian distributional assumption. We first estimate μ and λ by Maximum Likelihood and construct robust moments $m_k^\perp(x)$ following Section 4. Additional details are provided in Appendix D for the analytical expression of the variance matrix. As for the Student case, our procedure allows us to derive simple test statistics which perform well in terms of size and power.

5.2.1 The i.i.d. case

We first study the size and power properties for i.i.d. observations in Table 6, the alternative p.d.f. being the standard lognormal distribution.

We consider four tests based on a single moment: $m_{-1}^\perp(x)$, $m_1^\perp(x)$, $m_2^\perp(x)$ and $m_3^\perp(x)$. We do not include $m_0(x)$ because it is generated by the score, and therefore not useful for testing purposes. We also consider three joint moment tests $m_{j,g}^\perp(x)$ which combine the first g single moments of the previous list ($g = 2, 3$ or 4). The size properties are good and the power is the highest with the single moment $m_{-1}^\perp(x)$. The tests behave better in terms of size and power when the variance of the moments is computed theoretically.

5.2.2 The serial correlation case

We also generate samples which are serially correlated as in the case of realized variance studied in the empirical section. For this purpose, we simulate the stationary diffusion process

$$dx_t = \left(-\frac{\lambda}{2} x_t^2 + \frac{\mu^2}{2} x_t + \frac{\lambda\mu^2}{2} \right) dt + \sqrt{2}\mu x_t dW_t,$$

where W_t is a standard Brownian process. The marginal distribution of x_t is $\text{IG}(\lambda, \mu)$ (see Appendix D).

Table 7 displays the size and power properties. The size properties are similar to the ones found in the previous tables. For the power properties, we generate a series by taking the exponential of a Gaussian AR(1) with a standard normal marginal distribution and correlation equal to $\rho = 0.4$ and 0.9 respectively. The power performances decrease when ρ increases. It is worth noting that, like in the i.i.d. case, we can improve the small sample properties of the tests in terms of power if we use the theoretical variance to replace the estimated one.¹⁴

¹⁴The rejection frequencies for the test related to m_{-1}^\perp equal respectively 41.0%, 92.6%, 99.6% when $\rho = 0.4$ and 9.8%, 22.5%, 42.7% when $\rho = 0.9$. It may however cause the estimated matrix to be not positive definite in small sample.

6 Empirical examples

6.1 GARCH(1,1) model with Student innovations for exchange rates

As mentioned earlier, the GARCH(1,1) model with Student innovations seems to fit financial returns well (for a survey on GARCH models, see Bollerslev, Engle and Nelson, 1994). Using a Bayesian likelihood criterion, Kim, Shephard and Chib (1998) show that a T-GARCH(1,1) outperforms the lognormal stochastic volatility model of Taylor (1986), popularized by Harvey, Ruiz and Shephard (1994) and Jacquier, Polson and Rossi (1994).

Using the same data¹⁵ - observations of weekday close exchange rates from 1/10/81 to 28/6/85 (U.K. Pound, French Franc, Swiss Franc and Japanese Yen, all versus the U.S. Dollar) - we now test the Student distributional assumption for the innovations in a T-GARCH(1,1) model (see Equation 5.2). The results are provided in Table 8.

The model is estimated by Gaussian QML. We find that the degrees of freedom of the returns of FF-US\$, UK-US\$, SF-US\$, and Yen-US\$, equal respectively 9.61, 9.56, 6.64, and 5.54. Except for the SF-US\$ case, none of our tests reject the Student distributional assumption. For the SF-US\$, the rejection is due to even values for β , i.e. odd moments, which would infer that the fitted residuals are not symmetric.¹⁶ Bai's test does not reject the assumption but the simulations highlight that this test is not very good for detecting asymmetric distributions. However, Bai's test rejects the Student assumption for the Yen-US\$ rate, which conflicts with the results of our tests. It could be due to the size distortions of Bai's test suggested by the Monte Carlo experiments. Except for the SF-US\$ series, our results corroborate the findings of Kim, Shephard and Chib (1998).

6.2 Distribution of Realized Variance

The recent literature on volatility highlights the advantage of using high-frequency data to measure volatility of financial returns (Andersen and Bollerslev, 1998, Andersen, Bollerslev, Diebold and Labys, 2001, and Barndorff-Nielsen and Shephard, 2001). The realized variance is the sum of squared intra-day returns. Andersen, Bollerslev, Diebold and Labys (ABDL, 2003) suggests that it is lognormally distributed; an assumption formally rejected by Bontemps and Meddahi (2005). In contrast, Forsberg and Bollerslev (2000) assume that the distribution is an Inverse Gaussian one. We now test this distributional assumption.

We consider the same data as in ABDL (2003), i.e. returns of three exchange rates, DM-US\$, Yen-US\$ and Yen-DM, from December 1, 1986 through June 30, 1999. The realized variances are based on observations at five and thirty minutes. We therefore have six series.

¹⁵We are grateful to Neil Shephard for providing us the data. These data are used in Harvey, Ruiz and Shephard (1994) and Kim, Shephard and Chib (1998).

¹⁶It is common however to assume that foreign exchange rates have symmetric distributions.

Table 9 provides the empirical results. The second row of the table displays the estimates of the parameters μ and λ of the Inverse Gaussian distribution defined in Eq. (5.3). The Inverse Gaussian assumption is rejected. In the lognormal case, the skewness was one of the reasons for rejecting the distributional assumption. Here, except for the DM-US series, this is no longer the case. The rejection comes mostly from the moment labeled m_1 which is related to the expectations of x and x^2 . This is a constraint of the Inverse Gaussian distribution that is not satisfied by the data.

As pointed out in Bontemps and Meddahi (2005), a potential limitation of the analysis done above is the presence of long memory in realized variance. We assume that the long-run variance matrix of the moments is well defined, excluding long memory. Such analysis is devoted for future work.

Another possible explanation might be the presence of jumps. We therefore use the data from Huang and Tauchen (2007), i.e. the five-minute returns of the S&P index cash 1997-2002. We compute two different measures of the within-day price variance, the realized variance and the realized bipower variation. The latter is known to be a consistent estimator of the integrated variance with or without jumps. Table 10 displays the results of the same tests than in Table 9. The distributional assumption for the two measures is still strongly rejected. However, the test statistics are lower when one uses the realized bipower variation. The presence of jumps has an impact on the results but it does not seem to be the main reason for the rejection of the inverse Gaussian assumption.

7 Conclusion

We develop in this paper generic moment-based tests for testing parametric, continuous and univariate distributional assumptions. Our approach is simple. We consider the problem of parameter estimation error uncertainty and show how to construct robust moments with respect to this problem. Importantly, we derive moment tests in location-scale models that are robust, whatever the form of the conditional mean and variance, as for the ARMA/GARCH models. We also use the HAC method to handle some potential serial correlation in the variable of interest. An extensive simulation exercise for the Student and the Inverse Gaussian distributions assesses that the finite sample properties of our tests are very good, in particular in terms of power.

Our solution to take into account the problem of parameter estimation error uncertainty is to project the moment of interest on the space orthogonal to the one spanned by the score function. It could be done in population or in sample. Like for Wooldridge's (1990) approach or others, ours consists of modifying the moment of interest. It should be noted that any robust moment is orthogonal to the score function, but, of course, there are many ways to transform a moment into a robust one. Our transformation differs from Wooldridge's one.

It is important to stress the attractiveness of our moment-based tests. Using our framework, the choice of the moment is left to the researcher who can adapt the strategy to the problem or the alternative that is being considered. More importantly, this approach can be adapted to various cases such as discrete distributions (Bontemps, 2009) or multivariate ones (Bontemps, Feunou and Meddahi, 2011).

There are still some open questions. Optimality is one of them. We propose here a solution which consists of picking some particular moments which are appealing for their tractability. The question of optimality is a difficult task which is devoted to a separate paper (Bontemps and Meddahi, 2011). There is a trade-off between optimality and simplicity and, in this paper, we focus on providing simple tests, yet powerful for the Student and Inverse Gaussian distributions.

Appendix

A Orthogonal Polynomial and Pearson family

Let $q(\cdot)$ be the p.d.f. of a Pearson distribution:

$$\frac{q'(x)}{q(x)} = \frac{A(x)}{B(x)} = \frac{-(x+a)}{c_0 + c_1x + c_2x^2}. \quad (\text{A.1})$$

Let P_n be the polynomial of degree n which generates the orthonormal family¹⁷ with positive coefficient on the highest degree term. It is defined using some adaptation (to ensure the unit variance) of the Rodrigues Formula:

$$P_n = \alpha_n \frac{1}{q(x)} [B^n(x)q(x)]^{(n)}, \quad (\text{A.2})$$

where

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

The sequence of polynomials satisfies

$$n \geq 1, \quad P_{n+1}(x) = -\frac{1}{a_n} ((b_n - x)P_n(x) + a_{n-1}P_{n-1}(x)), \quad P_0(x) = 1, \quad P_{-1}(x) = 0, \quad (\text{A.3})$$

where

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

Table 1 reports the expression of the coefficients a_n and b_n , and the first polynomial for the well-known Pearson distributions.

B Proof of the propositions

B.1 Proof of Proposition 3.1

An integration by parts leads to:

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

where $E[\cdot]$ denotes the expectation with respect to the distribution of x . Hence, we get that

$$E[\psi'(x) + \psi(x)(\log q)'(x)] = 0, \quad (\text{B.1})$$

¹⁷This family is infinite and also dense in L^2 for all the distributions considered in Table 1, except for the Student case.

Table 1: Pearson distributions and orthonormal polynomials

Distribution	$q(x)$ $\frac{q'(x)}{q(x)}$	a_n b_n	First Polynomial
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $-\frac{x-\mu}{\sigma^2}$	$\sigma\sqrt{n}$ μ	$\frac{x-\mu}{\sigma}$
Student(ν)	$\frac{1}{\sqrt{\nu}B(\frac{\nu}{2}, \frac{1}{2})} \left[1 + \frac{x^2}{\nu}\right]^{-\frac{(\nu+1)}{2}}$ $-\frac{x}{\nu+x^2}$	$\sqrt{\frac{(n+1)\nu(\nu-n)}{(\nu-2n)(\nu-2n-2)}}$ 0	$\sqrt{\frac{\nu-2}{\nu}}x$
Gamma(α, β)	$\frac{x^{\alpha-1} \exp(-\frac{x}{\beta})}{\beta^\alpha \Gamma(\alpha)}$ $\frac{\beta(\alpha-1)-x}{\beta x}$	$\beta\sqrt{(n+1)(\alpha+n)}$ $\beta(\alpha+2n)$	$\frac{1}{\sqrt{\alpha}}\left(\frac{x}{\beta}-\alpha\right)$
Beta(α, β)	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$ $\frac{(-\alpha-\beta+2)x+\alpha-1}{x(1-x)}$	$\sqrt{\frac{(n+1)(\alpha+\beta+n-1)(\alpha+n)(\beta+n)}{(\alpha+\beta+2n)^4 - (\alpha+\beta+2n)^2}}$ $\frac{\alpha^2 + \alpha\beta + 2(\alpha+\beta)n + 2n^2 - 2\alpha - 2n}{(\alpha+\beta+2n)(\alpha+\beta+2n-2)}$	$\sqrt{\frac{\alpha+\beta+1}{\alpha\beta}}((\alpha+\beta)x - \alpha)$
Uniform $[0, 1]$	1 0	$\frac{n+1}{2\sqrt{(2n+1)(2n+3)}}$ $\frac{1}{2}$	$\sqrt{3}(2x-1)$

under the assumption L.

Conversely, if the moment m is given, it is not difficult to find the expression of the test-function ψ such that:

$$m(x) = \psi'(x) + \psi(x)(\log q)'(x).$$

This is a linear differential equation and the solution is

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

Observe that assumption L is equivalent to $E[m(x)] = 0$. It is important to keep in mind that the connection in Eq. (3.1) holds without the expectation operator. Consequently, the statistical properties (size, power) of a test based on (3.2) coincide with those of a test based on (3.3).

As pointed out in Section 2.1.2, Hansen and Scheinkman (1995) provide test functions of marginal distributions of continuous time processes. These test functions coincide with Eq. (3.2). More precisely, from Eq. (2.4), one gets

$$\frac{q'(x)}{q(x)} = \frac{2\mu(x) - (\sigma^2)'(x)}{\sigma^2(x)}. \quad (\text{B.2})$$

As a consequence, by plugging $\mu(x)$ from Eq. (B.2) in Eq. (2.5), one gets after some manipulations

$$E[(g'\sigma^2)'(x) + (\log q)'(x)(g'\sigma^2)(x)] = 0, \quad (\text{B.3})$$

which is exactly Eq. (3.2) using the test function $\psi = (g'\sigma^2)$. Hansen and Scheinkman (1995) assume that the variable x_t is Markovian to derive it, while we do not make this assumption to derive Eq. (3.2).

B.2 Proof of Proposition 4.1

Let $h_t^{-1}(\cdot, \phi)$ be the inverse function of h_t ($y_t = h_t^{-1}(x_t, \phi^0)$) and let $g_t(y, \phi) = m(h_t(y, \phi), \beta)$, $g_t^0(y) = g_t(y, \phi^0)$, $\frac{\partial g_t^0}{\partial \phi^\top}(y) = \left(\frac{\partial}{\partial \phi^\top} g_t(y, \phi) \right)_{\phi=\phi^0}$ and $m^0(x) = m(x, \beta^0)$.

$\phi^0 = (\theta^0, \beta^0)^\top$ is estimated using a procedure which provides a square-root consistent estimator $\hat{\phi}$ (like a ML or a GMM estimator). A Taylor expansion can be used to derive the asymptotic distribution of $\frac{1}{\sqrt{T}} \sum_{t=1}^T g_t(y_t, \phi)$ at the estimated value, $\hat{\phi}$:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g_t(y_t, \hat{\phi}) = \frac{1}{\sqrt{T}} \sum_{t=1}^T g_t^0(y_t) + E \left[\frac{\partial g_t^0}{\partial \phi^\top}(y_t) \right] \left[\sqrt{T}(\hat{\phi} - \phi^0) \right] + o_P(1). \quad (\text{B.4})$$

It is a function of the asymptotic deviation of $\hat{\phi}$ and its covariance with $\frac{1}{\sqrt{T}} \sum_{t=1}^T g_t^0(y_t)$. However, it is clear that a sufficient condition for the robustness of g_t against the parameter estimation

error uncertainty is

$$P_g = E \left[\frac{\partial g_t^0}{\partial \phi^\top}(y_t) \right] = 0. \quad (\text{B.5})$$

As the expectation of $g_t(\cdot, \phi)$ w.r.t. the conditional distribution of y_t is zero for any ϕ in the parameter space,

$$\int g_t(y_t, \phi) q_t(y_t, \phi) dy_t = 0,$$

we can derive the previous expression w.r.t. ϕ to obtain:

$$E_t \left[\frac{\partial g_t^0}{\partial \phi^\top}(y_t) \right] + E_t [g_t^0(y_t) s_t^\top(y_t, \phi^0)] = 0, \quad (\text{B.6})$$

where $s_t(y_t, \phi^0)$ is the conditional score function of the variable y_t given z_t and E_t the expectation with respect to the conditional distribution of y_t given z_t . Equation (B.6) is the generalization of the information matrix equality, which has been used, for instance, in Newey and McFadden (1994). This equation generalizes to an unconditional expectation by the law of iterated expectations. Consequently, the condition $P_g = 0$ holds if and only if $g_t^0(\cdot)$ is orthogonal to the score $s_t(\cdot, \phi^0)$, i.e.

$$0 = E[g_t^0(y_t) s_t^\top(y_t, \phi^0)] = E[m(x_t, \beta^0) s_t^\top(h_t^{-1}(x_t, \phi^0), \phi^0)]. \quad (\text{B.7})$$

B.3 Proof of Proposition 4.2

Let $m^\perp(x, \phi)$ be defined in (4.5):

$$m^\perp(x_t, \phi) = m(x_t, \beta) - E_t[m(x_t, \beta) s_t^\top(h_t^{-1}(x_t, \phi), \phi)] (V_t s_t^\top(h_t^{-1}(x_t, \phi), \phi))^{-1} s_t(h_t^{-1}(x_t, \phi), \phi).$$

First, it is a linear combination of $m(x, \beta)$ and $s_t(h_t^{-1}(x_t, \phi), \phi)$ which are both of expectation equal to zero under the null at the true value. The expectation of $m^\perp(x, \phi^0)$ is therefore also equal to zero.

Second,

$$E_t [m^\perp(x_t, \phi^0) s_t^\top(h_t^{-1}(x_t, \phi^0), \phi^0)] = E_t [g_t^\perp(y_t, \phi^0) s_t^\top(y_t, \phi^0)] = 0,$$

with $g_t^\perp(y, \phi) = m^\perp(h_t(y, \phi), \phi)$. This equality extends to unconditional expectation. Consequently, using the general information matrix equality (B.6), $m^\perp(\cdot)$ is a robust moment as it is, by construction, orthogonal to the score function.

B.4 The case without conditioning variables

When the link between y_t and x_t reduces to a transformation h_t which does not involve conditioning variables z_t , the conditional expectations and joint expectation which appear in Proposition 4.1 simplify to marginal expectations (h_t reduces to h and g_t to g). Proposition

4.1 can simply be rewritten by considering marginal distributions instead of conditional ones. Therefore, $g(\cdot)$ is robust if

$$P_g = E \left[\frac{\partial g^0}{\partial \phi^\top}(y_t) \right] = 0. \quad (\text{B.8})$$

A robust moment $m^\perp(\cdot)$ can be built by the projection of $m(\cdot)$ onto the orthogonal space to the one spanned by the marginal score, $s(\cdot)$:

$$m^\perp(x_t, \phi) = m(x_t, \beta) - E[m(x_t, \beta) s^\top(h_t^{-1}(x_t, \phi), \phi)] (V s^\top(h_t^{-1}(x_t, \phi), \phi))^{-1} s(h_t^{-1}(x_t, \phi), \phi).$$

B.5 Orthogonalization when the parameters are estimated by Maximum Likelihood

Assume now that we are in the marginal case and that the parameters are estimated by MLE. Like before, the functions s_t , m_t and g_t reduce to s , m and g . Under standard regularity assumptions, the ML estimator $\hat{\phi}_T$ satisfies the First Order Condition:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T s(y_t, \hat{\phi}_T) = 0.$$

So, using the definition of $g^\perp(\cdot)$, $g^\perp(y_t, \phi) = m^\perp(h(y_t, \phi), \phi)$ from Eq. (4.5), we obtain that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g(y_t, \hat{\phi}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^T g^\perp(y_t, \hat{\phi}_T).$$

Furthermore the variance of $g^\perp(y_t, \phi^0)$ is equal to:

$$V(g^\perp(y_t, \phi^0)) = V(g(y_t, \phi^0)) - E[g(y_t, \phi^0) s^\top(y_t, \phi^0)] V(s(y_t, \phi^0))^{-1} E[s(y_t, \phi^0) g^\top(y_t, \phi^0)],$$

because $g^\perp(y_t, \phi^0) = g(y_t, \phi^0) - E[g(y_t, \phi^0) s^\top(y_t, \phi^0)] V(s(y_t, \phi^0))^{-1} s(y_t, \phi^0)$.

From Newey (1985b)¹⁸, we know that:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g(y_t, \hat{\phi}_T) \xrightarrow[T \rightarrow \infty]{d} N(0, V(g^\perp(y_t, \phi^0))).$$

Consequently, the statistics built from our moment $g^\perp(\cdot)$ and the one derived from the non-robust moment $g(\cdot)$ - after having corrected for the parameter estimation uncertainty error - are the same.

B.6 Parameter uncertainty in a location-scale model

Assume $y = \mu + \sigma x$, where $x \sim P_\theta$, a parametric distribution indexed by θ and with p.d.f. $q(\cdot, \theta)$.

¹⁸See the limit of the quantity in Eq (2.11), p. 1052, with $L_T = [I_s, 0]$ using the notations of the paper.

With the notations of Section 4,

$$x = h(y, \mu, \sigma) = \frac{y - \mu}{\sigma}.$$

The score function can therefore be expressed directly:

$$s(y, (\mu, \sigma, \theta)) = \frac{\partial}{\partial(\mu, \sigma, \theta)^\top} \left[-\log \sigma + \log q \left(\frac{y - \mu}{\sigma}, \theta \right) \right] = \begin{bmatrix} s_\mu(y) = -\frac{1}{\sigma} \frac{\partial \log q}{\partial x}(x, \theta) \\ s_\sigma(y) = -\frac{1}{\sigma} - \frac{1}{\sigma} x \frac{\partial \log q}{\partial x}(x, \theta) \\ s_\theta(y) = \frac{\partial \log q}{\partial \theta}(x, \theta) \end{bmatrix}. \quad (\text{B.9})$$

If $m(x, \theta)$, a moment such as

$$E(m(x, \theta)) = 0,$$

is orthogonal to $\frac{\partial \log q}{\partial x}(x, \theta)$, $x \frac{\partial \log q}{\partial x}(x, \theta)$ and $\frac{\partial \log q}{\partial \theta}(x, \theta)$, it is orthogonal to the score function $s(y, (\mu, \sigma, \theta))$.

Assume now that $y = \mu(\phi) + \sigma(\phi)x$, where $x \sim P_\theta$ and where θ is part of ϕ . Under differentiability assumptions on $m(\phi)$ and $\sigma(\phi)$, we can derive similarly the score function $s(y, \phi)$, with $x = \frac{y - \mu(\phi)}{\sigma(\phi)}$:

$$s(y, \phi) = -\frac{\frac{\partial \mu}{\partial \phi}(\phi)}{\sigma(\phi)} \frac{\partial \log q}{\partial x}(x, \theta) - \frac{\partial \log \sigma}{\partial \phi}(\phi) \left(1 + x \frac{\partial \log q}{\partial x}(x, \theta) \right) + \frac{\partial \log q}{\partial \theta}(x, \theta). \quad (\text{B.10})$$

This score function is a linear combination (up to the constant term $-\frac{\partial \log \sigma}{\partial \phi}(\phi)$) of the three functions used in the previous location-scale model. A moment orthogonal to these three functions is therefore orthogonal to this new score function.

Assume now that $y_t = \mu_t(\phi) + \sigma_t(\phi)x_t$, where $\mu_t(\phi) = E(y_t | y_\tau, \tau \leq t-1)$, $\sigma_t^2(\phi) = V(y_t | y_\tau, \tau \leq t-1)$ and $x_t \sim P_\theta$, i.i.d. Under differentiability assumptions on $m_t(\phi)$ and $\sigma_t(\phi)$, we can derive similarly the score function $s_t(y_t, \phi)$, with $x_t = \frac{y_t - \mu_t(\phi)}{\sigma_t(\phi)}$:

$$s_t(y_t, \phi) = -\frac{\frac{\partial \mu_t}{\partial \phi}(\phi)}{\sigma_t(\phi)} \frac{\partial \log q}{\partial x}(x_t, \theta) - \frac{\partial \log \sigma_t}{\partial \phi}(\phi) \left(1 + x_t \frac{\partial \log q}{\partial x}(x_t, \theta) \right) + \frac{\partial \log q}{\partial \theta}(x_t, \theta). \quad (\text{B.11})$$

Any moment of x_t orthogonal to the three functions in (B.9) is orthogonal to this score function as x_t is independent from $y_\tau, \tau \leq t-1$.

B.7 Proof of Proposition 4.3

We define ψ^* in Eq. (4.8). For the sake of simplicity, we replace $q(x, \beta)$ by q , $s_t(h_t^{-1}(x, \phi), \phi)$ by s_t , $\psi^*(x, \beta)$ by ψ^* , $\psi(x, \beta)$ by ψ , $\frac{\partial}{\partial x}$ by $'$ and $\frac{\partial^2}{\partial x^2}$ by $''$. We have:

$$\psi^* = \psi - E_t[\psi s_t'^\top][E_t s_t' s_t'^\top]^{-1} s_t'.$$

The moment constructed from the robust test function ψ^* is:

$$\begin{aligned} m^* &= \psi'^* + \psi^* \frac{q'}{q} \\ &= \psi' + \psi \frac{q'}{q} - E_t[\psi s_t'^\top][E_t s_t' s_t'^\top]^{-1} \left(s_t'' + s_t' \frac{q'}{q} \right). \end{aligned}$$

Using the same integration by parts as in Eq. (3.2), we prove that m^* is orthogonal to the score function:

$$\begin{aligned} E_t(m^* s_t^\top) &= E_t \left[\left(\psi' + \psi \frac{q'}{q} \right) s_t^\top \right] - E_t[\psi s_t'^\top][E_t s_t' s_t'^\top]^{-1} E_t \left[\left(s_t'' + s_t' \frac{q'}{q} \right) s_t^\top \right] \\ &= \int ((\psi' q + \psi q') s_t^\top) dx - E_t[\psi s_t'^\top][E_t s_t' s_t'^\top]^{-1} \int ((s_t'' q + s_t' q') s_t^\top) dx \\ &= - \int \psi q s_t'^\top dx + E_t[\psi s_t'^\top][E_t s_t' s_t'^\top]^{-1} \int s_t' q s_t'^\top dx \quad (\text{by integration by parts}) \\ &= -E_t[\psi s_t'^\top] + E_t[\psi s_t'^\top][E_t s_t' s_t'^\top]^{-1} E_t[s_t' s_t'^\top] = 0. \end{aligned}$$

B.8 Transformation and parameter uncertainty

Let us denote by Q_t (resp. q_t), the c.d.f. (resp. the p.d.f.) of y_t (eventually conditional on z_t if there are some instruments involved). We transform y_t in some i.i.d. standard normal variable x_t using

$$x_t = \Phi^{-1} \circ Q_t(y_t, \theta^0),$$

where Φ is the c.d.f. of the standard normal distribution. Under the null, x_t is i.i.d., normally distributed. The score function w.r.t. the observable y_t is equal to:

$$s_t(y_t, \theta) = \frac{\partial \log q_t}{\partial \theta}(y_t, \theta).$$

Consequently, a moment $m(x_t)$ is robust if it is orthogonal to the last function taken at $y_t = Q_t^{-1}(\Phi(x_t), \theta^0)$ and at the true value $\theta = \theta^0$:

$$E[m(x_t) s_t^\top(Q_t^{-1}(\Phi(x_t), \theta^0), \theta^0)] = 0.$$

C Computations for the Student distribution

C.1 Preliminaries

Let

$$\psi_{\alpha, \beta}(x) = \frac{x^\beta}{(x^2 + \nu)^\alpha}.$$

We use in our Monte Carlo exercise values for β such as $\beta \leq 2\alpha + 2$. This ensures that the moment constructed is $O(x)$ and therefore of finite variance provided that $\nu > 2$.

Following (3.2), the (non-robust) moment derived from $\psi_{\alpha,\beta}$ is equal to

$$m_{\alpha,\beta}(x, \nu) = \frac{\beta \nu x^{\beta-1} - (2\alpha - \beta + \nu + 1)x^{\beta+1}}{(\nu + x^2)^{\alpha+1}}.$$

For the computations of the robust moments, we use the following expectations w.r.t. the $T(\nu)$ distribution. For any positive value α ,

$$A_\alpha^\nu = B_{\alpha,0}^\nu = E \left[\frac{1}{(x^2 + \nu)^\alpha} \right] = \frac{1}{\nu^\alpha} \frac{\Gamma(\alpha + \frac{\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\alpha + \frac{\nu+1}{2})}. \quad (\text{C.1})$$

For any even β such as $\beta \leq 2\alpha + 1 + \nu/2$,

$$B_{\alpha,\beta}^\nu = E \left[\frac{x^\beta}{(x^2 + \nu)^\alpha} \right] = B_{\alpha-1,\beta-2}^\nu - \nu B_{\alpha,\beta-2}^\nu. \quad (\text{C.2})$$

For any odd β , $B_{\alpha,\beta}^\nu = 0$ by symmetry.

We can, also, derive from (C.2) the following quantity:

$$\begin{aligned} \text{Cov}(m_{\alpha,\beta}(x, \nu), m_{\alpha',\beta'}(x, \nu)) &= (2\alpha - \beta + \nu + 1)(2\alpha' - \beta' + \nu + 1) B_{\alpha+\alpha'+2,\beta+\beta'+2}^\nu \\ &\quad - \nu (\beta(2\alpha' - \beta' + \nu + 1) + \beta'(2\alpha - \beta + \nu + 1)) B_{\alpha+\alpha'+2,\beta+\beta'}^\nu \\ &\quad + \beta\beta'\nu^2 B_{\alpha+\alpha'+2,\beta+\beta'-2}^\nu. \end{aligned} \quad (\text{C.3})$$

C.2 Location-scale model

In the location-scale model:

$$y = \mu + \sigma x, \quad x \sim T(\nu).$$

We denote by $q(\cdot, \nu)$ the p.d.f. of the Student distribution with ν degrees of freedom. The score function for this location-scale model has been derived before in (B.9). In the particular case of the Student distribution, it is equal to:

$$s(y) = \begin{bmatrix} \frac{\nu+1}{\sigma} \frac{x}{\nu+x^2} \\ -\frac{1}{\sigma} \left(1 - \frac{(\nu+1)x^2}{\nu+x^2} \right) \\ \frac{\partial}{\partial \nu} \log q(x, \nu) \end{bmatrix}. \quad (\text{C.4})$$

In consequence, its derivative with respect to y is:

$$\frac{\partial}{\partial y} s(y) = \begin{bmatrix} -\frac{\nu+1}{\sigma^2} \frac{(\nu-x^2)}{(\nu+x^2)^2} = -\frac{\nu+1}{\sigma^2} (2\nu\psi_{2,0}(x, \nu) - \psi_{1,0}(x, \nu)) \\ -2\frac{\nu(\nu+1)}{\sigma^2} \frac{x}{(\nu+x^2)^2} = -2\frac{\nu(\nu+1)}{\sigma^2} \psi_{2,1}(x, \nu) \\ \frac{1}{\sigma} \frac{x-x^3}{(\nu+x^2)^2} = \frac{1}{\sigma} ((\nu+1)\psi_{2,1}(x, \nu) - \psi_{1,1}(x, \nu)) \end{bmatrix}. \quad (\text{C.5})$$

C.2.1 β is even

When β is even, $\psi_{\alpha,\beta}(x)$ is symmetric and orthogonal to any asymmetric function, in particular the last two components of the score function. If we want to compute ψ^* , we only need to make it orthogonal to $2\nu\psi_{2,0}(x, \nu) - \psi_{1,0}(x, \nu)$. Following (4.8) and (C.2),

$$\begin{aligned} E[\psi_{\alpha,\beta}(x, \nu)(2\nu\psi_{2,0}(x, \nu) - \psi_{1,0}(x, \nu))] &= 2\nu B_{\alpha+2,\beta}^\nu - B_{\alpha+1,\beta}^\nu, \\ E(2\nu\psi_{2,0}(x, \nu) - \psi_{1,0}(x, \nu))^2 &= 4\nu^2 A_4^\nu - 4\nu A_3^\nu + A_2^\nu. \end{aligned}$$

Then we have an exact expression for the robust test function:

$$\psi_{\alpha,\beta}^*(x, \nu) = \psi_{\alpha,\beta}(x, \nu) - \underbrace{\frac{2\nu B_{\alpha+2,\beta}^\nu - B_{\alpha+1,\beta}^\nu}{4\nu^2 A_4^\nu - 4\nu A_3^\nu + A_2^\nu}}_{k_3(\alpha,\beta,\nu)} (2\nu\psi_{2,0}(x, \nu) - \psi_{1,0}(x, \nu)).$$

The robust moment $m_{\alpha,\beta}^*(x, \nu)$ can be written in a closed form:

$$\begin{aligned} m_{\alpha,\beta}^*(x, \nu) &= m_{\alpha,\beta}(x, \nu) - \frac{2\nu B_{\alpha+2,\beta}^\nu - B_{\alpha+1,\beta}^\nu}{4\nu^2 A_4^\nu - 4\nu A_3^\nu + A_2^\nu} (2\nu m_{2,0}(x, \nu) - m_{1,0}(x, \nu)) \\ &= \frac{\beta\nu x^{\beta-1} - (2\alpha - \beta + \nu + 1)x^{\beta+1}}{(\nu + x^2)^{\alpha+1}} - \frac{2\nu B_{\alpha+2,\beta}^\nu - B_{\alpha+1,\beta}^\nu}{4\nu^2 A_4^\nu - 4\nu A_3^\nu + A_2^\nu} \left(\frac{(\nu + 3)x^3 - \nu x(\nu + 7)}{(x^2 + \nu)^3} \right). \end{aligned} \quad (\text{C.6})$$

The coefficients A_α^ν and $B_{\alpha,\beta}^\nu$ are given in (C.1) and (C.2), the variances and covariances can be deduced from (C.3).

C.2.2 β is odd

When β is odd, $\psi_{\alpha,\beta}(x)$ is orthogonal to the first component of the score function. To be robust, it has to be orthogonal to the last two components which is equivalent to being orthogonal to $\psi_{2,1}(x, \nu)$ and $\psi_{1,1}(x, \nu)$.

The coefficients can be derived from (4.8):

$$\begin{aligned} \begin{bmatrix} k_1(\alpha, \beta, \nu) \\ k_2(\alpha, \beta, \nu) \end{bmatrix} &= \begin{bmatrix} E\psi_{1,1}^2(x, \nu) & E(\psi_{1,1}(x, \nu)\psi_{2,1}(x, \nu)) \\ E(\psi_{1,1}(x, \nu)\psi_{2,1}(x, \nu)) & E\psi_{2,1}^2(x, \nu) \end{bmatrix}^{-1} \begin{bmatrix} E(\psi_{\alpha,\beta}(x, \nu)\psi_{1,1}(x, \nu)) \\ E(\psi_{\alpha,\beta}(x, \nu)\psi_{2,1}(x, \nu)) \end{bmatrix} \\ &= \frac{1}{B_{2,2}^\nu B_{4,2}^\nu - (B_{3,2}^\nu)^2} \begin{bmatrix} B_{4,2}^\nu & -B_{3,2}^\nu \\ -B_{3,2}^\nu & B_{2,2}^\nu \end{bmatrix} \begin{bmatrix} B_{\alpha+1,\beta+1}^\nu \\ B_{\alpha+2,\beta+1}^\nu \end{bmatrix}. \end{aligned} \quad (\text{C.7})$$

We derive the robust-moment and the expression of its variance in a similar way as before:

$$\begin{aligned}
m_{\alpha,\beta}^*(x, \nu) &= m_{\alpha,\beta}(x, \nu) - k_1(\alpha, \beta, \nu)m_{1,1}(x, \nu) - k_2(\alpha, \beta, \nu)m_{2,1}(x, \nu) \\
&= \frac{\beta\nu x^{\beta-1} - (2\alpha - \beta + \nu + 1)x^{\beta+1}}{(\nu + x^2)^{\alpha+1}} \\
&\quad - \frac{B_{4,2}^\nu B_{\alpha+1,\beta+1}^\nu - B_{3,2}^\nu B_{\alpha+2,\beta+1}^\nu}{B_{2,2}^\nu B_{4,2}^\nu - (B_{3,2}^\nu)^2} \frac{\nu - (2 + \nu)x^2}{(x^2 + \nu)^2} \\
&\quad - \frac{B_{2,2}^\nu B_{\alpha+2,\beta+1}^\nu - B_{3,2}^\nu B_{\alpha+1,\beta+1}^\nu}{B_{2,2}^\nu B_{4,2}^\nu - (B_{3,2}^\nu)^2} \frac{\nu - (4 + \nu)x^2}{(x^2 + \nu)^3}.
\end{aligned} \tag{C.8}$$

C.2.3 Particular test-statistic for testing the Student Distribution

Following the previous computations, we can derive a family of test statistics associated with $m_{\alpha,\beta}^*(x, \nu)$ for any positive value α and any integer β (provided that $\beta + 1 - 2(\alpha + 1) < \frac{\nu}{2}$).

$$\xi_{\alpha,\beta} = \frac{T \left(\frac{1}{T} \sum_{t=1}^T m_{\alpha,\beta}^*(x_t, \nu) \right)^2}{Vm_{\alpha,\beta}^*(x_t, \nu)} \xrightarrow[T \rightarrow \infty]{d} \chi^2(1). \tag{C.9}$$

The variance $Vm_{\alpha,\beta}^*(x_t, \nu)$ can be derived analytically from Eq. (C.3). These statistics have power against symmetric alternatives when β is odd, and power against asymmetric alternatives when β is even. They are valid for any specification of μ and σ , in particular for any T-GARCH model.

In the simulations, it appears that one joint moment combining one odd moment and one even moment has good power against a wide range of alternatives. We therefore propose the following statistic using $m_{5/2,1}^*$ and $m_{1,2}^*$ (the two individual statistics are asymptotically independent):

$$\begin{aligned}
BM_T &= \xi_{5/2,1} + \xi_{1,2} \\
&= T \left(\frac{1}{Vm_{5/2,1}^*(x_t, \nu)} \left(\frac{1}{T} \sum_{t=1}^T m_{5/2,1}^*(x_t, \nu) \right)^2 + \frac{1}{Vm_{1,2}^*(x_t, \nu)} \left(\frac{1}{T} \sum_{t=1}^T m_{1,2}^*(x_t, \nu) \right)^2 \right).
\end{aligned} \tag{C.10}$$

which is asymptotically $\chi^2(2)$ distributed under the null. The expression of the moments are given in (C.6) and (C.8).

C.3 Mixtures of normals used in the power analysis

In the Monte Carlo section, we consider a mixture of two random normal variables as alternative for our power analysis. Let p be the weight associated to the first normal distribution. We compute the variances of the two normal distributions, denoted respectively $\sigma_1^2(p, \nu)$ and

$\sigma_2^2(p, \nu)$, to fit the first five moments of a T-distribution with ν degrees of freedom. We have

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

The following tabular displays the first three even moments of these mixtures for various values of p and the corresponding moments of the T-distribution they are supposed to match.

Moments of the mixtures and of the t-distribution

Panel A: $\nu = 5$				
		EX^2	EX^4	EX^6
$T(\nu)$		1.66	25	—
Mixture	$p = 0.7$	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
Mixture	$p = 0.7$	1.11	4.16	29.09
	$p = 0.8$	1.11	4.16	29.66
	$p = 0.9$	1.11	4.16	30.72

D Computations for the Inverse Gaussian Distribution

The p.d.f. of the Inverse Gaussian distribution (IG) with parameters μ, λ is equal to:

$$q(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp \left[-\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right],$$

and therefore:

$$\frac{q'(x)}{q(x)} = - \left(\frac{3}{2x} + \frac{\lambda(x^2 - \mu^2)}{2\mu^2 x^2} \right).$$

We must first note that, except for some degenerate case where λ is equal to 0 (and the variance is infinite), this distribution does not belong to the Pearson family of distributions.

Taking $\psi_k(x) = x^{k+1}$, Assumption L is satisfied for this choice for positive and negative values of k . The moment m_k constructed from this test function using Eq. (3.2) is given, up to some scale value, by the following expression:

$$m_k(x) = x^{k+1} - \frac{2\mu^2}{\lambda} \left(k - \frac{1}{2} \right) x^k - \mu^2 x^{k-1}. \quad (\text{D.1})$$

We derive consequently:

$$a_k = Ex^k = \frac{2\mu^2}{\lambda} \left(k - \frac{3}{2}\right) a_{k-1} + \mu^2 a_{k-2}, \quad (\text{D.2})$$

using the initial conditions $a_0=1$ and $a_1 = \mu$. The a_k are used to derive the exact expression of $V_k = Vm_k(x)$.

For example,

$$Em_k(x)x^j = \frac{2\mu^2}{\lambda} j a_{k+j}, \quad (\text{D.3})$$

and

$$Em_k(x)m_j(x) = \frac{4\mu^2}{\lambda^2} [(k+1)(j+1) - 3/2] a_{k+j} + \lambda a_{k+j-1}. \quad (\text{D.4})$$

If we observe x , the parameter estimation error uncertainty concerns only the parameters of the distribution, i.e. λ and μ . The score function $s_{\lambda,\mu}$ is equal to:

$$s_{\lambda,\mu}(x) = \begin{bmatrix} s_\lambda(x) \\ s_\mu(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial \log q}{\partial \lambda}(x) \\ \frac{\partial \log q}{\partial \mu}(x) \end{bmatrix} = \begin{bmatrix} \frac{-1}{2\mu^2} m_0(x) + \left(\frac{1}{\lambda} + \frac{1}{\mu} - x^{-1}\right) \\ \frac{\lambda}{\mu^3} (x - \mu) \end{bmatrix}. \quad (\text{D.5})$$

The variance of the score is:

$$V s_{\lambda,\mu} = \begin{bmatrix} \frac{1}{2\lambda^2} & 0 \\ 0 & \frac{\lambda}{\mu^3} \end{bmatrix}, \quad (\text{D.6})$$

and the robust moments are derived using (D.3) and (D.4):

$$\begin{aligned} m_k^\perp(x) &= m_k(x) - 2\lambda^2 E \left[m_k(x) \frac{\partial \log q}{\partial \lambda}(x) \right] \frac{\partial \log q}{\partial \lambda}(x) - \frac{\mu^3}{\lambda} E \left[m_k(x) \frac{\partial \log q}{\partial \mu}(x) \right] \frac{\partial \log q}{\partial \mu}(x) \\ &= m_k(x) + 2\mu^2(2k-1)a_k s_\lambda(x) - \frac{2}{\mu} a_{k+1}(x - \mu). \end{aligned} \quad (\text{D.7})$$

Diffusion process with Inverse Gaussian marginal distribution

Let the diffusion process y_t be defined by the stochastic differential equation:

$$dy_t = m(y_t)dt + \sqrt{2\mu}dW_t,$$

where $m(y) = -\frac{\mu^2}{2} - \frac{\lambda}{2} \exp(y) + \frac{\lambda\mu^2}{2} \exp(-y)$.

Let $x_t = \exp(y_t)$. Using Ito's lemma, x_t satisfies the stochastic differential equation:

$$dx_t = \underbrace{(m(\log(x_t))x_t + \mu^2 x_t)}_{\mu_x(x_t)} dt + \underbrace{\sqrt{2\mu}x_t}_{\sigma_x(x_t)} dW_t. \quad (\text{D.8})$$

The marginal p.d.f. $q(\cdot)$ of x_t satisfies the differential equation:

$$\begin{aligned}
\frac{q'(x)}{q(x)} &= \frac{2\mu_x(x) - (\sigma_x^2)'(x)}{\sigma_x^2(x)} \\
&= \frac{-3\mu^2x - \lambda x^2 + \lambda\mu^2}{2\mu^2x^2} \\
&= \frac{-3}{2x} - \frac{\lambda}{2\mu^2} \frac{x^2 - \mu^2}{x^2},
\end{aligned} \tag{D.9}$$

which is the one of the Inverse Gaussian distribution with parameters λ and μ .

Table 2: Size of the tests, $\nu = 5$

λ, μ and ν known.	λ, μ and ν est. Theoretical Var.	λ, μ and ν est. In-sample Var.
Non-robust moments		
T 100 500 1000	T 100 500 1000	T 100 500 1000
$m_{0,1}$ 5.2 4.9 5.0	$m_{0,1}$ 0.0 0.0 0.0	$m_{0,1}$ 0.1 0.0 0.0
$m_{\frac{3}{2},1}$ 5.4 5.0 5.0	$m_{\frac{3}{2},1}$ 0.7 0.7 0.7	$m_{\frac{3}{2},1}$ 0.8 0.7 0.7
Robust moments		
T 100 500 1000	T 100 500 1000	T 100 500 1000
$m_{0,1}^*$ 4.5 5.5 4.9	$m_{0,1}^*$ 4.5 6.6 5.8	$m_{0,1}^*$ 6.4 7.1 5.8
$m_{1/2,1}^*$ 4.5 5.6 4.8	$m_{1/2,1}^*$ 4.6 6.3 5.6	$m_{1/2,1}^*$ 6.2 6.6 5.7
$m_{5/2,1}^*$ 4.9 4.9 4.6	$m_{5/2,1}^*$ 4.3 5.9 5.6	$m_{5/2,1}^*$ 5.4 6.1 5.5
$m_{1/2,0}^*$ 5.1 5.0 5.0	$m_{1/2,0}^*$ 6.9 5.6 5.2	$m_{1/2,0}^*$ 7.0 5.6 5.3
$m_{1,0}^*$ 4.9 5.1 5.0	$m_{1,0}^*$ 6.6 5.7 5.3	$m_{1,0}^*$ 7.0 5.7 5.4
$m_{1,2}^*$ 5.5 5.1 4.9	$m_{1,2}^*$ 7.4 5.8 5.4	$m_{1,2}^*$ 7.6 5.6 5.4
m_j^* 5.1 4.9 4.8	m_j^* 6.0 6.3 5.8	m_j^* 6.6 6.8 5.8
KS 4.8 4.5 4.9	KS 0.7 0.7 1.0	H_3 23.1 10.5 7.6
	S_{Bai} 7.0 12.1 10.8	H_4 36.6 20.6 15.6
		H_{34} 52.3 26.1 18.1

Note: for each sample size T (100, 500 and 1000), we report the rejection frequencies for a 5% significance level test of the Student distributional assumption in a constant location-scale model. The data are i.i.d. from a $T(5)$ distribution. KS is the Kolmogorov-Smirnov test, S_{Bai} the test from Bai (2003) and H_i the Hermite polynomial test for normality (implemented after having transformed the variable into a normal one).

Table 3: Power of the Student tests

Asymmetric alternatives:												
$X \sim \chi^2(5)$				$X \sim \chi^2(15)$				$X \sim \chi^2(30)$				
T	100	500	1000	T	100	500	1000	T	100	500	1000	
$m_{0,1}^*$	26.1	94.6	99.9	$m_{0,1}^*$	5.6	32.9	63.5	$m_{0,1}^*$	1.9	11.8	25.6	
$m_{1/2,1}^*$	28.9	96.0	100.0	$m_{1/2,1}^*$	6.1	34.4	65.3	$m_{1/2,1}^*$	2.2	12.5	26.6	
$m_{5/2,1}^*$	32.2	97.5	100.0	$m_{5/2,1}^*$	6.8	36.1	67.9	$m_{5/2,1}^*$	2.7	13.1	27.7	
$m_{1/2,0}^*$	95.0	100.0	100.0	$m_{1/2,0}^*$	59.4	99.9	100.0	$m_{1/2,0}^*$	33.4	98.4	100.0	
$m_{1,0}^*$	94.2	100.0	100.0	$m_{1,0}^*$	58.9	99.8	100.0	$m_{1,0}^*$	33.0	98.2	100.0	
$m_{1,2}^*$	97.5	100.0	100.0	$m_{1,2}^*$	62.2	100.0	100.0	$m_{1,2}^*$	34.1	99.1	100.0	
m_j^*	97.2	100.0	100.0	m_j^*	49.9	100.0	100.0	m_j^*	20.8	98.1	100.0	
S_{Bai}	3.2	70.7	100.0	S_{Bai}	3.7	17.4	65.4	S_{Bai}	2.9	10.1	27.0	

Mixtures of Normals: $VX = \frac{5}{3}$.

$p = 0.7$				$p = 0.8$				$p = 0.9$				
T	100	500	1000	T	100	500	1000	T	100	500	1000	
$m_{0,1}^*$	95.9	100.0	100.0	$m_{0,1}^*$	12.6	18.6	25.7	$m_{0,1}^*$	8.3	38.8	69.9	
$m_{1/2,1}^*$	94.9	100.0	100.0	$m_{1/2,1}^*$	11.2	13.1	15.6	$m_{1/2,1}^*$	9.0	39.6	70.3	
$m_{5/2,1}^*$	91.8	100.0	100.0	$m_{5/2,1}^*$	8.3	7.9	7.6	$m_{5/2,1}^*$	8.4	29.1	54.3	
$m_{1/2,0}^*$	37.0	37.7	38.7	$m_{1/2,0}^*$	12.2	11.8	12.1	$m_{1/2,0}^*$	5.6	4.4	4.4	
$m_{1,0}^*$	37.1	38.1	38.9	$m_{1,0}^*$	11.8	11.4	11.5	$m_{1,0}^*$	5.4	4.2	4.4	
$m_{1,2}^*$	34.2	34.9	35.0	$m_{1,2}^*$	14.8	14.4	14.9	$m_{1,2}^*$	6.9	5.8	5.2	
m_j^*	96.7	100.0	100.0	m_j^*	17.0	17.9	20.4	m_j^*	8.6	31.1	59.8	
S_{Bai}	16.1	97.8	100.0	S_{Bai}	4.4	3.4	11.2	S_{Bai}	14.0	31.8	37.9	

Mixtures of Normals: $VX = \frac{20}{18}$.

$p = 0.7$				$p = 0.8$				$p = 0.9$				
T	100	500	1000	T	100	500	1000	T	100	500	1000	
$m_{0,1}^*$	1.6	2.2	3.2	$m_{0,1}^*$	1.4	2.8	3.0	$m_{0,1}^*$	1.3	3.3	5.2	
$m_{1/2,1}^*$	1.6	2.3	3.3	$m_{1/2,1}^*$	1.6	2.9	3.3	$m_{1/2,1}^*$	1.6	3.6	5.3	
$m_{5/2,1}^*$	1.9	2.9	3.8	$m_{5/2,1}^*$	2.0	3.4	3.8	$m_{5/2,1}^*$	1.9	4.0	5.1	
$m_{1/2,0}^*$	3.6	5.7	5.5	$m_{1/2,0}^*$	4.1	4.4	4.7	$m_{1/2,0}^*$	3.4	4.4	5.0	
$m_{1,0}^*$	3.6	5.7	5.5	$m_{1,0}^*$	4.2	4.4	4.8	$m_{1,0}^*$	3.4	4.3	5.0	
$m_{1,2}^*$	3.4	5.6	5.3	$m_{1,2}^*$	3.9	4.3	5.0	$m_{1,2}^*$	3.4	4.2	5.2	
m_j^*	2.1	3.8	3.9	m_j^*	2.0	3.1	4.0	m_j^*	1.9	3.7	5.3	
S_{Bai}	3.5	8.9	7.7	S_{Bai}	3.6	10.3	10.1	S_{Bai}	3.6	12.5	13.9	

Note: the data are i.i.d. and we test the Student distributional assumption in a constant location-scale model. The true DGP is either some asymmetric distribution (χ^2) or a symmetric one (mixture of normals). We report the rejection frequencies for a 5% significance level test. See Table 2 for notations.

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

Size				Power against asym. distribution ($\chi^2(30)$)				Power against mixture of normals ($p = 0.7$)			
T	100	500	1000	T	100	500	1000	T	100	500	1000
$m_{0,1}^*$	4.4	6.8	6.0	$m_{0,1}^*$	1.9	11.1	24.6	$m_{0,1}^*$	93.3	100.0	100.0
$m_{1/2,1}^*$	4.7	6.7	5.9	$m_{1/2,1}^*$	2.1	11.6	25.8	$m_{1/2,1}^*$	91.4	100.0	100.0
$m_{5/2,1}^*$	4.9	6.5	5.1	$m_{5/2,1}^*$	2.4	12.6	27.7	$m_{5/2,1}^*$	87.1	100.0	100.0
$m_{1/2,0}^*$	6.0	5.7	5.1	$m_{1/2,0}^*$	28.8	98.1	100.0	$m_{1/2,0}^*$	34.1	38.4	37.3
$m_{1,0}^*$	6.0	5.8	5.1	$m_{1,0}^*$	28.6	97.9	100.0	$m_{1,0}^*$	34.4	38.7	37.8
$m_{1,2}^*$	6.7	6.4	5.5	$m_{1,2}^*$	29.8	98.7	100.0	$m_{1,2}^*$	31.8	35.1	34.3
m_j^*	6.0	7.2	5.7	m_j^*	17.7	97.2	100.0	m_j^*	94.0	100.0	100.0
S_{Bai}	5.7	10.4	9.4	S_{Bai}	2.9	10.8	25.9	S_{Bai}	14.9	97.2	100.0

Note: the data are generated from a GARCH model with i.i.d. innovations. The true DGP for the innovation is either a Student distribution ($\nu = 5$), some asymmetric distribution ($\chi^2(30)$) or some mixture of normals. We test the Student distributional assumption on the fitted residuals. We report the rejection frequencies for a 5% significance level test.

Table 6: Size and Power - IG (i.i.d. case)

Variance matrix computed theoretically				
$X \sim IG(0.5, 0.5)$				
	T	100	500	1000
m_{-1}^\perp		3.6	4.6	4.8
m_1^\perp		3.1	4.1	4.3
m_2^\perp		2.6	3.4	3.6
m_3^\perp		1.4	2.6	3.1
$m_{j,2}^\perp$		4.1	4.5	4.9
$m_{j,3}^\perp$		3.3	4.5	4.9
$m_{j,4}^\perp$		3.2	3.9	4.0
$X \sim \text{lognormal}$				
	T	100	500	1000
m_{-1}^\perp		44.2	94.5	99.8
m_1^\perp		4.1	14.6	21.2
m_2^\perp		3.2	11.4	17.9
m_3^\perp		1.7	9.2	15.7
$m_{j,2}^\perp$		40.9	94.7	99.8
$m_{j,3}^\perp$		37.2	93.0	99.7
$m_{j,4}^\perp$		34.4	91.1	99.6

Variance matrix computed in the sample

$X \sim IG(0.5, 0.5)$				
	T	100	500	1000
m_{-1}^\perp		6.1	6.9	6.6
m_1^\perp		4.8	7.4	7.2
m_2^\perp		2.0	11.8	12.5
m_3^\perp		0.1	18.4	22.1
$m_{j,2}^\perp$		5.3	9.7	9.7
$m_{j,3}^\perp$		3.7	8.6	15.2
$m_{j,4}^\perp$		11.0	7.7	15.2
$X \sim \text{lognormal}$				
	T	100	500	1000
m_{-1}^\perp		7.2	55.2	84.4
m_1^\perp		8.3	20.6	19.8
m_2^\perp		0.2	13.4	12.8
m_3^\perp		0.1	0.3	10.4
$m_{j,2}^\perp$		4.8	43.7	75.4
$m_{j,3}^\perp$		9.8	35.8	72.6
$m_{j,4}^\perp$		28.9	35.2	81.4

Note: we test the Inverse Gaussian distributional assumption. We report the rejection frequencies for a 5% significance level test. m_{-1}^\perp , m_1^\perp , etc. denote a single moment, $m_{j,g}^\perp$ the joint moment which takes the first g single moments. The IG (0.5,0.5) is used for assessing the size performances, the standard lognormal is used for the power study.

Table 7: Size and Power - IG (serial correlation case)

Size				Power							
				$\rho = 0.4$				$\rho = 0.9$			
T	100	500	1000	T	100	500	1000	T	100	500	1000
m_{-1}^\perp	2.9	5.6	6.0	m_{-1}^\perp	6.2	50.2	82.2	m_{-1}^\perp	3.3	11.1	25.7
m_1^\perp	2.3	5.0	5.7	m_1^\perp	4.7	17.2	18.0	m_1^\perp	1.1	2.4	5.8
m_2^\perp	0.6	4.0	6.5	m_2^\perp	0.2	7.3	9.5	m_2^\perp	0.4	0.3	0.4
m_3^\perp	0.2	1.2	6.9	m_3^\perp	0.1	0.1	2.5	m_3^\perp	0.1	0.0	0.1
$m_{j,2}^\perp$	1.2	4.9	6.7	$m_{j,2}^\perp$	2.8	36.4	71.4	$m_{j,2}^\perp$	0.9	5.6	16.6
$m_{j,3}^\perp$	1.6	3.4	5.1	$m_{j,3}^\perp$	7.2	29.5	67.6	$m_{j,3}^\perp$	2.3	9.2	17.8
$m_{j,4}^\perp$	6.1	3.6	4.8	$m_{j,4}^\perp$	24.9	27.7	73.6	$m_{j,4}^\perp$	3.6	20.3	21.6

Note: we test the Inverse Gaussian distributional assumption. We report the rejection frequencies at a 5% significance level test. See Table 6 for details. The variance estimator is a HAC estimator *à la* Andrews.

Table 8: 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
$m_{0,1}^*$	0.101 (0.75)	1.474 (0.22)	0.002 (0.97)	0.003 (0.95)
$m_{1/2,1}^*$	0.128 (0.72)	1.308 (0.25)	0.014 (0.91)	0.013 (0.91)
$m_{5/2,1}^*$	0.225 (0.64)	0.829 (0.36)	0.319 (0.57)	0.186 (0.67)
$m_{1/2,0}^*$	2.925 (0.09)	0.795 (0.37)	6.050 (0.01)	0.233 (0.63)
$m_{1,0}^*$	2.871 (0.09)	0.935 (0.33)	5.687 (0.02)	0.192 (0.66)
$m_{1,2}^*$	2.970 (0.08)	0.246 (0.62)	7.747 (0.01)	0.423 (0.52)
m_j^*	3.097 (0.21)	1.554 (0.46)	7.761 (0.02)	0.437 (0.80)
S_{Bai}	2.732 (≥ 0.05)	1.776 (≥ 0.05)	1.476 (≥ 0.05)	3.834 (≤ 0.05)

Note: we test the standardized Student distributional assumption for the innovation term of a GARCH(1,1) model estimated by the Gaussian QML method. The test statistics and their corresponding p-values (in brackets) are reported. The notations are defined in Table 2.

Table 9: Testing the Inverse Gaussian distributional assumption of realized variance

	DM-US\$-5	DM-US\$-30	Yen-US\$-5	Yen-US\$-30	Yen-DM-5	Yen-DM-30
$(\hat{\mu}, \hat{\lambda})$	(0.55,0.95)	(0.44,0.46)	(0.55,0.96)	(0.44,0.46)	(0.55,0.94)	(0.44,0.46)
m_{-1}^{\perp}	0.27 (0.60)	3.24 (0.07)	0.87 (0.35)	2.63 (0.10)	0.57 (0.45)	2.94 (0.09)
m_1^{\perp}	13.12 (0.00)	7.31 (0.01)	4.57 (0.03)	4.80 (0.03)	4.51 (0.03)	3.97 (0.05)
m_2^{\perp}	10.41 (0.00)	5.52 (0.02)	2.47 (0.12)	3.71 (0.05)	2.99 (0.08)	3.55 (0.06)
m_3^{\perp}	7.11 (0.01)	3.93 (0.05)	1.57 (0.21)	2.77 (0.10)	2.31 (0.13)	2.59 (0.11)
$m_{j,2}^{\perp}$	16.60 (0.00)	16.18 (0.00)	4.64 (0.10)	14.02 (0.00)	6.43 (0.04)	8.93 (0.01)
$m_{j,3}^{\perp}$	22.65 (0.00)	24.44 (0.00)	15.48 (0.00)	20.50 (0.00)	19.12 (0.00)	10.58 (0.01)
$m_{j,4}^{\perp}$	23.68 (0.00)	24.94 (0.00)	21.25 (0.00)	24.03 (0.00)	19.72 (0.00)	11.73 (0.02)

Note: we test the Inverse Gaussian assumption for the realized variance of exchange rates. The test statistics and their corresponding p-values (in brackets) are reported. The notations are defined in Table 6.

Table 10: Testing the Inverse Gaussian distributional assumption of realized variance, S&P 500, 1997-2002

	Realized variance	Realized bipower variation
m_{-1}^{\perp}	11.30 (0.00)	6.91 (0.01)
m_1^{\perp}	9.35 (0.00)	6.22 (0.01)
m_2^{\perp}	6.61 (0.01)	4.49 (0.03)
m_3^{\perp}	4.90 (0.03)	3.47 (0.06)
$m_{j,2}^{\perp}$	12.51 (0.00)	7.79 (0.02)
$m_{j,2}^{\perp}$	14.39 (0.00)	10.98 (0.01)
$m_{j,2}^{\perp}$	19.41 (0.00)	11.33 (0.02)

Note: we test the Inverse Gaussian assumption for the realized variance of the S&P index. The test statistics and their corresponding p-values (in brackets) are reported. The notations are defined in Table 6.

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 (2010), "Operator Methods for Continuous-Time Markov Processes", in: Y. Aït-Sahalia and L. P. Hansen (Eds), *Handbook of Financial Econometrics*, 1-66.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42-55.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.
- Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 60, 953-966.
- 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*, 23, 49-60.
- Barndorff-Nielsen, O.E. and N. Shephard (2001), "Non-Gaussian OU based Models and some of their uses in Financial Economics," with discussion, *Journal of the Royal Statistical Society, B*, 63, 167-241.
- Bera, A. and Y. Biliias (2002), "The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis," *Journal of Econometrics*, 107, 51-86.
- Boldin, M.V. (1982), "Estimation of the distribution of noise in an autoregressive scheme", *Theory Probab. Appl.*, 27, 866-871.
- 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," in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol IV, 2959-3038, Elsevier, North-Holland, Amsterdam.
- 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. (2009), "Moment Based Tests for Discrete Distributions", working paper, Université de Toulouse.
- Bontemps, C., B. Feunou, and N. Meddahi (2011), "Moment Based Tests of Multivariate Distributions," work in progress, Université de Toulouse.
- Bontemps, C. and N. Meddahi (2005), "Testing Normality: A GMM Approach," *Journal of Econometrics*, 124, 149-186.
- Bontemps, C. and N. Meddahi (2011), "Optimal Moment Based Tests of Distributional Assumptions," work in progress, Université de Toulouse.
- Chen, X., L.P. Hansen and J. Scheinkman (2009), "Nonlinear Principal Components and Long-Run Implications of Multivariate Diffusions", *Annals of Statistics*, 37, 4279-4312.
- 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 Tests", *Econometrica*, 65, 627-646.
- Chihara, T.S. (1978), *An Introduction to Orthogonal Polynomials*, Gordon and Breach Science Publishers.
- Christoffersen, P. F. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, pp. 841-862.
- 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 (2005), "Bootstrap Specification Test for Diffusion Processes," *Journal of Econometrics*, 124, 117-148
- 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 Exchange," *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. White, 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.
- Durbin, J. (1973), "Weak Convergence of Sample Distribution Functions When Parameters Are Estimated," *Annals of Statistics*, 1, 279-290.
- Fiorentini, G., E. Sentana, and G. Calzolari (2004), "On the Validity of the Jarque-Bera Normality Test in Conditionally Heteroskedastic Dynamic Regression Models," *Economics Letters*, 83, 307-312.
- Forsberg L. and T. Bollerslev (2002), "Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the EURO): the GARCH-NIG model", *Journal of Econometrics*, 17, 535-548.
- Gallant, A.R. (1980), "Explicit Estimators of Parametric Functions in Nonlinear Regressions," *Journal of the American Statistical Association*, 75, 182-193.
- Gourieroux, C. and J. Jasiak (2006), "Multivariate Jacobi Process with Application to Smooth Transitions," *Journal of Econometrics*, 131, 475-505.
- 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.
- Huang, X. and G. Tauchen (2005), "The Relative Contribution of Jumps to Total Price Variance", *Journal of Financial Econometrics*, 3, 456-499.
- 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.
- Kalliovirta L. K. (2006), "Misspecification Tests Based on Quantile Residuals", working paper.
- Karlin, S. and H.M. Taylor (1981), *A Second Course in Stochastic Processes*, Academic Press.
- Khmaladze, E.V. (1981), "Martingale Approach in the Theory of Goodness-of-Tests," *Theory of Probability and its Application*, 26, 240-257.
- Khmaladze, E.V. and H. L. Koul (1984), "Martingale transforms goodness-of-fit tests in regression models," *Annals of Statistics*, 32, 995-1034.
- Kim, S., N. Shephard and S. Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

- 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.
- Lobato, I. N. and C. Velasco (2004), "A Simple Test Of Normality For Time Series," *Econometric Theory*, 20, 671-689.
- Michael, J.R., W.R. Schucany and R.W. Haas (1976), "Generating Random Variates Using Transformations with Multiple Roots", *The American Statistician*, 30, 88-90.
- Newey W.K. (1985a), "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229-256.
- Newey W.K. (1985b), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047-1070.
- Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol IV, 2111-2245, Elsevier, North-Holland, Amsterdam.
- Newey W.K. and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
- Orme, C.D. (1990), "The Small Sample Performance of the Information Matrix Test", *Journal of Econometrics*, 46, 309-331.
- 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.
- White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25.
- 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.