Testing Distributional Assumptions: A GMM Approach*

Christian Bontemps† and **Nour Meddahi**‡

April 18, 2006

INCOMPLETE

Abstract

In this paper, we consider testing marginal distributional assumptions. Special cases that we consider are the Pearson's family like the Normal, Student, Gamma, Beta and uniform distributions. The test statistics we consider are based on the first moment conditions derived by Hansen and Scheinkman (1995) when one considers a continuous time model. These moment conditions are valid even if the observations are not a sample of a continuous time model. We treat in detail the parameter uncertainty problem when the considered process is not observed but depends on estimators of unknown parameters. We also consider the time series case and adopt a HAC approach for this purpose. This is a generalization of Bontemps and Meddahi (2005) who considered this approach for the Normal case.

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

JEL codes: C12, C15.

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

†University of Toulouse, GREMAQ & IDEI, 21 alle de Brienne, 31000 Toulouse, France. E-mail: bontemps@cict.fr.

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

1 Introduction

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

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

where the function $\psi(\cdot)$ follows some regularity conditions and constraints on the boundary support of x discussed later on. The Equation (1.1) is clearly important for modeling, estimation and specification testing purposes. The main goal of the paper is the use of Eq. (1.1) and the generalized method of moments (GMM) of Hansen (1982) for testing distributional assumptions. The paper extends Bontemps and Meddahi (2005) who used the same approach for testing normality. In this case, when one wants to test that x is a standard normal random variable, one has $\log(q)'(x) = -x$, and therefore Eq. (1.1) becomes

$$E[\psi'(x) - \psi(x)x] = 0,$$

which is known as the Stein equation (Stein, 1972).

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

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

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

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

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

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

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

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

2 Test functions

2.1 Moment conditions

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

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

Hence, we get that

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

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

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

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

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

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

while feasible test statistics will be based on

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

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

In many cases, one has moment restrictions like

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

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

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

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

is given by

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

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

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

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

2.2 Transformed distributions

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

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

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

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

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

In addition, we have

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

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

2.3 Pearson's distributions and their generalizations

At the end of the nineteenth century, Karl Pearson introduced his famous family distribution that extends the classical normal distribution (see, e.g., Bera and Bilias (2002) for a historical review). If a distribution with a density function $q(\cdot)$ on (l, r) belongs to the Pearson family, then the ratio $q'(\cdot)/q(\cdot)$ equals the ratio of two polynomials $A(\cdot)$ and $B(\cdot)$, where $A(\cdot)$ is affine and $B(\cdot)$ is quadratic and positive on (l, r), i.e.,

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

The Pearson's class of distributions include as special examples the Normal, Student, Gamma, Beta, and Uniform distributions. We will study in detail this class of distributions in the next section. A major motivation for introducing this family is their simple estimation. By using (2.1) for $\psi_i(x) = x^i B(x)$, $i = 1, 2, ...,$, one gets this recursive equations

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

Pearson solved this system for $j = 1, ..., 4$, i.e., he derived $\theta = (a, c_0, c_1, c_2)^\top$ in terms of $E[X^j]$, $j = 1, 2, 3, 4$, and then provided estimator for θ by using the empirical counterpart of $E[X^j]$ (under the assumption that these moments exist). This was the introduction of the method of moments.

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

2.4 Marginal distribution of scalar diffusions

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

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

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t, \tag{2.8}$$

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

$$q(x) = K\sigma^{-2}(x)\exp\left(\int_z^x \frac{2\mu(u)}{\sigma^2(u)}du\right), \tag{2.9}$$

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

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

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

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

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

¿From (2.9), one gets easily

$$\frac{q'(x)}{q(x)} = \frac{2\mu(x) - (\sigma^2)'(x)}{\sigma^2(x)}. \tag{2.12}$$

As a consequence, by using (2.12) in (2.10), one gets after some manipulations

$$E[(g\sigma^2)'(x) + (\log q)'(x)(g\sigma^2)(x)] = 0, \tag{2.13}$$

which is exactly the general test function (2.1) applied to the function $\psi = (g\sigma^2)'$. Again, Hansen and Scheinkman (1995) assumed that the variable x_t is Markovian to derive (2.10) (and (2.13)) while we did not for deriving (2.1).

2.5 Asymptotic distribution of the test statistics

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

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

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

Throughout the paper, we assume the matrix Σ defined by

$$\Sigma \equiv \lim_{T\to+\infty} Var\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right] = \sum_{h=-\infty}^{+\infty} E[m(x_t)m(x_{t-h})^\top], \tag{2.14}$$

is finite and positive definite. In the context of time series, this assumption ruled out some long memory processes; see Bontemps and Meddahi (2005). Under some regularity conditions, we know since Hansen (1982) that

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

while

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right)^\top \Sigma^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right) \sim \chi^2(p). \tag{2.15}$$

For the feasibility of the test procedure, one needs the matrix Σ or a consistent estimator of it.

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

$$\Sigma = Var[m(x)] = E[m(x)m(x)^\top]. \tag{2.16}$$

Two cases may arise. One can explicitly compute the matrix Σ and, hence, one can use the test statistic (2.15). We will see later for the Pearson's distributions that this is the case for some functions $\psi_i(\cdot)$, i.e., one explicitly knows the matrix Σ. In particular, we will show that (2.1) implies that $E[P_i(x)] = 0$ where $P_i(\cdot)$ is a sequence of orthonormal polynomials, i.e., $E[P_i(x)P_j(x)] = \delta_{i,j}$ where $\delta_{\cdot,\cdot}$ is the Kronecker symbol. Consequently, the matrix Σ will be the identity matrix, implying that the univariate test statistics based on $E[P_i(x)] = 0$ are asymptotically independent.

In the second case, computing Σ explicitly is not possible (or difficult), then one can use any consistent estimator of Σ and denoted by $\hat{\Sigma}_T$, like

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

In this case, one can use the following test statistic

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

Assume now that the observations are correlated. Then without additional assumptions on the dependence, one can not explicitly compute the matrix Σ. For instance, knowing that the marginal distribution of a process is normal does not imply that its conditional distribution is normal and therefore one has not information about $E[m(x_t)m(x_{t-h})]$ in (2.14) for $h \neq 0$. When one does not have information about the dependence in the process x_t, one has to estimate Σ. A traditional solution is to estimate this matrix by using a Heteroskedastic-Autocorrelation-Consistent (HAC) method like Newey and West (1987) or Andrews (1991). This is one of the motivations of using a GMM approach for testing normality. We will follow this approach as did Richardson and Smith (1993), Bai and Ng (2005) and Bontemps and Meddahi (2005) for testing normality, and by Aït-Sahalia (1996), Conley, Hansen, Luttmer and Scheinkman (1997), and Corradi and Swanson (2002) for testing marginal distributions of nonlinear scalar diffusion processes.

3 Optimality

In this section, we are interested by the characterization of the function $\psi(\cdot)$ or $m(\cdot)$ that maximizes the power of the specification tests based on empirical counterpart of (2.15). As usual, optimal tests deal with fully parametric models. Consequently, we will focus on testing these models.

3.1 Point Optimal Tests

In this section, we are interested in point optimal tests, i.e., we want to test a distribution $q_0(\cdot)$ against an alternative one $q_a(\cdot)$. We allow the two distribution functions to be in the same class of distributions (e.g., test a $T(5)$ distribution against a $T(20)$ distribution) or in different classes (e.g., test a $T(5)$ distribution against a finite mixture of normal distributions). However, we assume that the support of the two distributions are the same.

We will study the case of consistent tests. Therefore, given that the test statistic (2.15) becomes in the univariate and i.i.d. case

$$T\ \frac{\left(\frac{1}{T}\sum_{t=1}^{T} m(x_t)\right)^2}{Var[m(x_t)]},$$

we will consider tests where the denominator is finite. Two cases may hold. The first one holds when one knows $Var[m(x)]$ under the null (analytically or by simulation). Then, a test statistic that one can consider is

$$T\ \frac{\left(\frac{1}{T}\sum_{t=1}^{T} m(x_t)\right)^2}{\underset{0}{V}[m(x_t)]}, \tag{3.1}$$

where $\underset{0}{V}[m(x_t)]$ equals the variance of $m(x)$ under the null.

A second test statistic that one can uses is the one that corresponds to (3.1) when one uses the empirical counterpart of $\underset{0}{V}(m(x_t) = E[m^2(x_t)])$, i.e.,

$$T\ \frac{\left(\frac{1}{T}\sum_{t=1}^{T} m(x_t)\right)^2}{\left(\frac{1}{T}\sum_{t=1}^{T} m^2(x_t)\right)}. \tag{3.2}$$

The asymptotic limit of the denominator in (3.2) is $\underset{a}{E}[m^2(x_t)]$, where $\underset{a}{E}[\cdot]$ denotes the expectation operator under the alternative case. Again, in order to get consistent tests, we will consider the cases where $\underset{a}{E}[m(x_t)]$ is finite.

Hall (2000) showed that one gains power by centering the empirical moment of $m(\cdot)$ in the denominator of (3.2), i.e., by considering

$$T\ \frac{\left(\frac{1}{T}\sum_{t=1}^{T} m(x_t)\right)^2}{\left(\frac{1}{T}\sum_{t=1}^{T}\left(m(x_t) - \frac{1}{T}\sum_{s=1}^{T} m(x_s)\right)^2\right)}. \tag{3.3}$$

Consequently, the asymptotic limit of the denominator in (3.3) is $\underset{a}{V}[m(x_t)]$, where $\underset{a}{V}[\cdot]$ denotes the variance operator under the alternative case.

The three test statistics will have power as soon as $\underset{a}{E}[m(x_t)] \neq 0$. In these cases, the statistics will diverge to infinity whatever $m(\cdot)$ with $\underset{a}{E}[m(x_t)] \neq 0$. Therefore one needs a criteria to rank the tests based on (3.1), or (3.2), or (3.3), or a test based on a combination of the three statistics. Geweke (1981) studied this problem in a general context. By using that theory, one can show that one has to maximize the approximate slope of the test statistic (Bahadur, 1967) which in turn is

$$\frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{0}{V}[m(x_t)]} = \frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{0}{E}[m^2(x_t)]} \tag{3.4}$$

when one considers the class of test statistics defined by (3.1) and

$$\frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{a}{E}[m^2(x_t)]} \tag{3.5}$$

when one considers the class of tests defined by (3.2).

An implicit assumption here is that $\underset{a}{E}[\| \ m(x_t) \ \|] < +\infty$; otherwise, the test statistics go to infinity at a faster speed than T. The assumption $\underset{a}{E}[\| \ m(x_t) \ \|] < +\infty$ automatically holds when $q_a(x) = O(q_0(x))$ when $x \to l$ and $x \to r$, i.e., the tails of the distribution under the null are fatter than the alternative ones (when $l = -\infty$ or $r = +\infty$). We define the two classes of test functions that we will study below:

$$\mathcal{C}_1 = \left\{ m(\cdot), \text{such that } \underset{0}{E}[m(x_t)] = 0, \ \underset{0}{E}[m^2(x_t)] < +\infty, \text{ and } \underset{a}{E}[|m(x_t)|] < +\infty \right\}. \tag{3.6}$$

$$\mathcal{C}_2 = \left\{ m(\cdot), \text{such that } \underset{0}{E}[m(x_t)] = 0, \text{ and } \underset{a}{E}[m^2(x_t)] < +\infty \right\}. \tag{3.7}$$

In the sequel, we will also make two assumptions:

Assumption A2: $\underset{0}{E}\left[\left(\frac{q_a(x)}{q_0(x)}\right)^2\right] < +\infty.$

Assumption A3: $\underset{0}{E}\left[\frac{q_0(x)}{q_a(x)}\right] < +\infty.$

Assumption A2 allows the function $m_1^*(\cdot)$ defined by

$$m_1^*(x) = \frac{q_a(x)}{q_0(x)} - 1, \tag{3.8}$$

to be square-integrable under the null. In addition, Assumption A2 implies that $\underset{a}{E}[|m_1^*(x_t)|] < +\infty$. Observe that $\underset{0}{E}[m_1^*(x)] = 0$ and therefore $m_1^*(\cdot) \in \mathcal{C}_1$. Assumption A3 allows the function $m_2^*(\cdot)$ given by

$$m_2^*(x) = \frac{\frac{q_0(x)}{q_a(x)}}{\underset{0}{E}\left[\frac{q_0(x_t)}{q_a(x_t)}\right]} - 1, \tag{3.9}$$

to be well defined. Observe that $\underset{0}{E}[m_2^*(x_t)] = 0$. In addition, we have $\underset{a}{E}[(m_2^*(x_t))^2] < +\infty$ given that

$$\underset{a}{E}\left[\left(\frac{q_0(x)}{q_a(x)}\right)^2\right] = \underset{0}{E}\left[\frac{q_0(x)}{q_a(x)}\right] < +\infty.$$

Consequently, $m_2^*(\cdot) \in \mathcal{C}_2$.

Proposition 3.1

1) Consider a function $m(\cdot) \in \mathcal{C}_1$ and assume that Assumption A2 holds. Then,

$$\frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{0}{V}[m(x_t)]} \leq \frac{\left(\underset{a}{E}[m_1^*(x_t)]\right)^2}{\underset{0}{V}[m_1^*(x_t)]} = \underset{0}{E}\left[\left(\frac{q_a(x_t)}{q_0(x_t)} - 1\right)^2\right] = \underset{0}{E}\left[\left(\frac{q_a(x_t)}{q_0(x_t)}\right)^2\right] - 1. \tag{3.10}$$

In addition, the inequality in (3.10) is an equality if and only if $m(\cdot)$ is proportional to $m_1^*(\cdot)$. Consequently, an optimal test within the class of test functions $\mathcal{C}_1$ is the test that corresponds to $m_1^*(\cdot)$.

2) Consider a function $m(\cdot) \in \mathcal{C}_2$ and assume that Assumption A3 holds. Then,

$$\frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{a}{E}[m^2(x_t)]} \leq \frac{\left(\underset{a}{E}[m_2^*(x_t)]\right)^2}{\underset{a}{E}[(m_2^*(x_t))^2]} = 1 - \frac{1}{\underset{0}{E}\left[\frac{q_0(x_t)}{q_a(x_t)}\right]}. \tag{3.11}$$

In addition, the inequality in (3.11) is an equality if and only if $m(\cdot)$ is proportional to $m_2^*(\cdot)$. Consequently, an optimal test within the class of test functions $\mathcal{C}_2$ is the test that corresponds to $m_2^*(\cdot)$.

3) Consider a function $m(\cdot) \in \mathcal{C}_2$ and assume that Assumption A3 holds. Then,

$$\frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{a}{V}[m(x_t)]} \leq \frac{\left(\underset{a}{E}[m_2^*(x_t)]\right)^2}{\underset{a}{V}[(m_2^*(x_t))^2]} = \underset{0}{E}\left[\frac{q_0(x_t)}{q_a(x_t)}\right] - 1. \tag{3.12}$$

In addition, the inequality in (3.12) is an equality if and only if $m(\cdot)$ is proportional to $m_2^*(\cdot)$. Consequently, an optimal test within the class of test functions $\mathcal{C}_2$ is the test that corresponds to $m_2^*(\cdot)$.

3.2 Local Optimality

Chesher and Smith (1997) studied the relationship between moment based tests, i.e., (2.15), and the likelihood ratio tests (LR) when on considers i.i.d. data. In particular, they derived a class of models where the moment based test statistic (2.15) is asymptotically equivalent to LR tests. Meanwhile, we know that LR tests are optimal; consequently it is of interest to characterize the class of models where (2.15) are optimal. We follow in the sequel Chesher and Smith (1997) for deriving this class of models.

Let $h(\cdot)$ be a positive valued real function, $h : R \to R^+$, with finite derivatives of all orders, such that $h(0) = h'(0) = 1$; an obvious example is the exponential function. Assume that the density function $q(\cdot)$ depends on an unknown parameter β; we therefore adopt the notation $q(x, \beta)$. Let $f(x, \theta)$ be the function defined by

$$f(x, \theta) = C^{-1}(\theta) q(x, \beta) h(\lambda^\top m(x)), \tag{3.13}$$

where $\lambda \in R^p$, $\theta = (\beta^\top, \lambda^\top)^\top$, and $C(\theta)$ is the normalized constant

$$C(\theta) = \int_l^r q(x, \beta) h(\lambda^\top m(x)) dx.$$

Here, we assume that $h(\cdot)$ is chosen such that $C(\theta)$ exists. Clearly, under our assumptions, $f(x, \theta)$ is the density function of an augmented model. The function $h(\cdot)$ is called a "carrier function" because it caries the argument of the moment condition into the augmented density. Under some regularity conditions, Chesher and Smith (1997) showed that

$$E_\lambda[m(x)] = 0 \Longleftrightarrow \lambda = 0, \tag{3.14}$$

where $E_\lambda[\cdot]$ denotes the expectation operator with respect to density function $f(x, \theta)$. Consequently, the test statistic (2.15) is optimal with the class of models defined by (3.13). This results is a generalization of Jarque and Bera (1980) and Kiefer and Salmon (1983) who studied optimal test in the case of linear homoskedastic regression model with Gaussian errors.

Also, (3.13) has an entropy interpretation; in particular, it is connected with the study of Duan and So (2001). It is also connected with the empirical likelihood literature (e.g., Kitamura and Stutzer, 1997).

4 Parameter uncertainty

In general, the density function involved in (2.1) depends on unknown parameters. Moreover, the variable x may be not observable but can depend on unknown parameters like, e.g., residuals in a regression model. Therefore, one has to first estimate these parameters before implementing any distributional test procedure. However, it is well known that the asymptotic distribution of the feasible test statistic based on (2.15) is, in general, different from the unfeasible one that uses the true (unknown) parameter. The main purpose of this section is to derive sufficient conditions in order to avoid the parameter uncertainty problem, i.e., making the asymptotic distribution of the feasible and unfeasible test statistics coincide.

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

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

where the function $h(\cdot)$ is a known function and θ^0 is an unknown parameter different from β^0. The aim of the test is to assess if the model satisfies:

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

For some reasons that will appear shortly, we consider four examples for the models.

Example 1: x_t is observable; it is either an i.i.d. sample or a serially correlated process. This is the example where an econometrician has some observable data and he wonders if these data follow some particular distribution. For example, a popular model of the short term interest rate is the square-root process (Cox, Ingersoll and Ross, 1984) whose marginal distribution is gamma. In this case, $z_t = x_t$, there is no parameter θ^0, and $h(\cdot)$ is the identity function, i.e., $h(z_t, \beta) = z_t = x_t$.

Example 2: x_t is unobservable and

$$x_t = \frac{y_t - w_t^\top \gamma^0}{\sigma^0},$$

with $z_t \equiv (y_t, w_t)$, $\theta^0 = (\gamma^0, \sigma^0)$, where y_t is the dependent variable and w_t are the regressors. Note that in this example the function $h(\cdot)$ defined in (4.1) does not depend on β^0. The classical example of linear regression with normal errors is a special case where $\theta^0 = (\gamma^0, \sigma^0)$ and $\beta^0 = (0, 1)$. This case was studied in detail by Bontemps and Meddahi (2005).

Example 3: x_t is unobservable and equals the standardized innovation in a T-GARCH model (Bollerslev 1986, 1987):

$$y_t = \mu^0 + \varepsilon_t,\ \varepsilon_t = \sqrt{v_t(\theta^0)}u_t,\ v_t = \omega^0 + \alpha^0 \varepsilon_{t-1}^2 + \eta^0 v_{t-1}(\theta^0),\ \ x_t \equiv \sqrt{\frac{\nu^0}{\nu^0 - 2}}\, u_t,\ x_t \text{ i.i.d.} \sim \mathrm{T}(\nu^0),$$

where $\theta^0 = (\mu^0, \omega^0, \alpha^0, \eta^0)$. In this case, z_t is y_t and its past values, i.e., $z_t = (y_t, y_{t-1}, ...)$, $\beta^0 = \nu^0$, and

$$x_t = h(z_t, \beta^0, \theta^0) = \sqrt{\frac{\nu^0}{\nu^0 - 2}}\, \frac{y_t - \mu^0}{\sqrt{v_t(\theta^0)}}.$$

Example 4: x_t is unobservable and its distribution does not depend on an unknown parameter, e.g., $\mathcal{N}(0, 1)$ or uniform distribution on $(0, 1)$. For instance, in their evaluating density forecasts analysis, Diebold, Gunther and Tay (1998) transformed a variable y_t into a uniform distribution by applying $x_t = Q(y_t, \theta^0)$ where $Q(., \theta^0)$ is the cumulative distribution function of y_t. This is also the case of Bai (2003) or Duan (2003) who tested distributional assumptions for dynamic models. In this case $h(y, \beta^0, \theta^0) = Q(y, \theta^0)$ and does not depend on β^0.

We will turn back to these examples later after having derived the results on the parameter uncertainty problem. In the sequel, we allow the test-function $\psi(\cdot)$ in (2.1) to depend on both β^0 and θ^0. Thus, (2.1) becomes

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

while we will use the notation

$$m(x, \beta, \theta) \equiv \frac{\partial \psi}{\partial x}(x, \beta, \theta) + \psi(x, \beta, \theta)\frac{\partial \log q}{\partial x}(x, \beta), \quad \psi(\cdot) = (\psi_1(\cdot), ..., \psi_p(\cdot))^\top, \tag{4.4}$$

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

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

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

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

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

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

i.e.,

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

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

$$P_{\psi\beta} = E\left[\frac{\partial m^0}{\partial \beta^\top}(x_t) + \frac{\partial m^0}{\partial x}(x_t) \frac{\partial h^0}{\partial \beta^\top}(h^{-1}(x_t, \beta^0, \theta^0))\right], \tag{4.6}$$

$$P_{\psi\theta} = E\left[\frac{\partial m^0}{\partial \theta^\top}(x_t) + \frac{\partial m^0}{\partial x}(x_t) \frac{\partial h^0}{\partial \theta^\top}(h^{-1}(x_t, \beta^0, \theta^0))\right], \tag{4.7}$$

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

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

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

$$P_m = [P_{\psi\beta} \ P_{\psi\theta}] = 0. \tag{4.8}$$

In the sequel, we will propose three approaches that ensure (4.8). The first one will be the characterization of the functions $m(\cdot)$ such that (4.8) holds. The main idea of the second and third approaches is the transform of the vector $m(\cdot)$ onto a new vector denoted $\tilde{m}(\cdot)$ such that $P_{\tilde{m}} = 0$. In the second approach, we will transform each function $\psi_i(\cdot)$ in (4.3) onto a function $\tilde{\psi}_i(\cdot)$ to get $P_{\tilde{m}} = 0$. In contrast, in the third approach we will get (4.8) by transforming jointly the functions $\psi_1(\cdot), ..., \psi_p(\cdot)$. For a systematic analysis of specification tests under parameter uncertainty, see Bontemps, Dufour, Gonçalves, and Meddahi (2005).

4.1 First approach: orthogonality to the score function

In this section, we will first provide a result of general interest for the analysis of specification tests under parameter uncertainty. We will then specify this result in the context of testing distributional assumption.

Proposition 4.1 Bontemps, Dufour, Gonçalves, and Meddahi (2005). *Let u be a random variable with a density function $f(u,\gamma^0)$ and assume that a vectorial function $n(u,\gamma^0)$ is such that $E[n(u,\gamma^0)]=0$. Then we have*

$$E\left[\frac{\partial n}{\partial \gamma^\top}(u,\gamma^0)\right]=0 \Longleftrightarrow E[n(u,\gamma^0)\ s(u,\gamma^0)^\top]=0, \tag{4.9}$$

where $s(u,\gamma)$ is the score function, i.e.,

$$s(u,\gamma)=\frac{\partial \log f}{\partial \gamma}(u,\gamma).$$

This proposition is of general interest and its implications are studied in Bontemps, Dufour, Gonçalves, and Meddahi (2005) (the proof is also provided in the Appendix). It shows that moment test functions are robust against parameter uncertainty when they are orthogonal to the score function. This explains the result of Bontemps and Meddahi (2005) who showed that Hermite polynomials, $H_i(\cdot)$, $i\geq 3$, are robust against parameter uncertainty when one tests that an unobservable variable follows a $\mathcal{N}(\mu^0,(\sigma^0)^2)$ distribution (as in Example 1). In this case, $\gamma^0=(\mu^0,(\sigma^0)^2)^\top$ and the score function is given by

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

where $H_1(\cdot)$ and $H_2(\cdot)$ are the first and second Hermite polynomials. However, the distribution of $(x-\mu^0)/\sigma^0$ is $\mathcal{N}(0,1)$; $\forall i\geq 1$, $E[H_i((x-\mu^0)/\sigma^0)]=0$; and $\forall\ i,j,\ i\neq j$, $E[H_i((x-\mu^0)/\sigma^0)H_j((x-\mu^0)/\sigma^0)]=0$. Hence, $\forall i\geq 3$, the test statistics based on $E[H_i((x-\mu^0)/\sigma^0)]=0$ are robust against parameter uncertainty.

Actually, Bontemps and Meddahi (2005) also showed this robustness result when x is not observable and is, for instance, the residual of a heteroskedastic and nonlinear regression model (as in Examples 2, 3, and 4). In this case, one has to take into account the uncertainty in $h(z_t,\hat\beta_T,\hat\theta_T)$. However, the general result of Proposition 4.1 holds. In the remaining of this subsection, we will characterize the functions $\psi(\cdot)$ such that (4.8) holds.

By using (4.4), one easily shows

$$\begin{aligned}
P_{\psi\beta}&=E\left[\frac{\partial^2\psi^0}{\partial x\partial\beta^\top}(x)+\frac{\partial\log q^0}{\partial x}\frac{\partial\psi^0}{\partial\beta^\top}(x)+\psi^0(x)\frac{\partial^2\log q^0}{\partial x\partial\beta^\top}(x)\right]\\
&\quad+E\left[\left(\frac{\partial^2\psi^0}{\partial^2 x}(x)+\frac{\partial\log q^0}{\partial x}\frac{\partial\psi^0}{\partial x}(x)+\psi^0(x)\frac{\partial^2\log q^0}{\partial^2 x}(x)\right)\frac{\partial h^0}{\partial\beta^\top}(h^{-1}(z,\beta^0,\theta^0))\right],\\
P_{\psi\theta}&=E\left[\frac{\partial^2\psi^0}{\partial x\partial\theta^\top}(x)+\frac{\partial\log q^0}{\partial x}\frac{\partial\psi^0}{\partial\theta^\top}(x)\right]\\
&\quad+E\left[\left(\frac{\partial^2\psi^0}{\partial^2 x}(x)+\frac{\partial\log q^0}{\partial x}\frac{\partial\psi^0}{\partial x}(x)+\psi^0(x)\frac{\partial^2\log q^0}{\partial^2 x}(x)\right)\frac{\partial h^0}{\partial\theta^\top}(h^{-1}(x,\beta^0,\theta^0))\right].
\end{aligned}$$

Therefore, when $\psi(\cdot)$ is such that both $P_{\psi\beta}$ and $P_{\psi\theta}$ equal zero, the test statistic (4.3) is robust against parameter uncertainty. The form of $P_{\psi\beta}$ and $P_{\psi\theta}$ involves the derivative of the $\psi(\cdot)$ with respect to β, θ, and x. Therefore, their form are not easy to interpret. For this reason, we will use (4.3) to write $P_{\psi\beta}$ and $P_{\psi\theta}$ without these derivatives.

Proposition 4.2 *Let $\psi(x,\beta,\theta)$ be a test-function such that Assumption A1 holds for $\psi(x,\beta^0,\theta^0)$, $\frac{\partial\psi}{\partial\beta^\top}(x,\beta^0,\theta^0)$ and $\frac{\partial\psi}{\partial\theta^\top}(x,\beta^0,\theta^0)$. Then*

$$\begin{aligned} P_{\psi\beta} &= E\left[\psi^0(x)\frac{\partial^2\log q^0}{\partial x\partial\beta^\top}(x)\right] \\ &\quad + E\left[\left(\frac{\partial^2\psi^0}{\partial^2 x}(x)+\frac{\partial\log q^0}{\partial x}(x)\frac{\partial\psi^0}{\partial x}(x)+\psi^0(x)\frac{\partial^2\log q^0}{\partial^2 x}(x)\right)\frac{\partial h^0}{\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right], \end{aligned} \tag{4.10}$$

$$P_{\psi\theta} = E\left[\left(\frac{\partial^2\psi^0}{\partial^2 x}(x)+\frac{\partial\log q^0}{\partial x}(x)\frac{\partial\psi^0}{\partial x}(x)+\psi^0(x)\frac{\partial^2\log q^0}{\partial^2 x}(x)\right)\frac{\partial h^0}{\partial\theta^\top}(h^{-1}(x,\beta^0,\theta^0))\right]. \tag{4.11}$$

Observe that here we made the additional assumption that Assumption A1 holds for the derivative function $\psi(\cdot)$ with respect to β and θ at the true values β^0 and θ^0. This is not however a restrictive assumption and it holds in all our examples. The most interesting results is that the dependence of $P_{\psi\beta}$ and $P_{\psi\theta}$ on these derivatives does not appear. In other words, the uncertainty in $\psi(x,\hat{\beta}_T,\hat{\theta}_T)$ does not matter for the robustness of (4.3) against parameter uncertainty. This result is similar to the theory of optimal instruments where the optimal instrument depends in nonlinear models on the unknown parameters; however, the feasible optimal instrument achieves the optimality (asymptotically).

There is still in (4.10) and (4.11) the dependence of the test statistic (4.3) on the uncertainty in $h(\hat{x},\beta^0,\theta^0)$ through the derivative of $\psi(\cdot)$ with respect to x. The goal of the following proposition is to remove this dependence:

Proposition 4.3 *Let $\psi(x,\beta,\theta)$ be a test function under the assumptions of Proposition 4.2. In addition, assume that Assumption A1 holds for $\frac{\partial\psi^0}{\partial x}(x)\frac{\partial h^0}{\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))$, $\psi^0(x)\frac{\partial^2 h^0}{\partial x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))$, $\frac{\partial\psi^0}{\partial x}(x)\frac{\partial h^0}{\partial\theta^\top}(h^{-1}(x,\beta^0,\theta^0))$, and $\psi^0(x)\frac{\partial^2 h^0}{\partial x\partial\theta^\top}(h^{-1}(x,\beta^0,\theta^0))$. Then*

$$P_{\psi\beta} = E\left[\psi^0(x)\frac{\partial^2\log q^0}{\partial x\partial\beta^\top}(x)\right] + E\left[\psi^0(x)b^0_\beta(x)\right], \tag{4.12}$$

$$P_{\psi\theta} = E\left[\psi^0(x)b^0_\theta(x)\right], \tag{4.13}$$

where

$$\begin{aligned} b_\beta(x,\beta,\theta) &= \frac{\partial^3 h}{\partial^2 x\partial\beta^\top}(h^{-1}(x,\beta,\theta),\beta,\theta)+\frac{\partial\log q}{\partial x}(x,\beta,\theta)\frac{\partial^2 h}{\partial x\partial\beta^\top}(h^{-1}(x,\beta,\theta),\beta,\theta) \\ &\quad + \frac{\partial^2\log q}{\partial^2 x}(x,\beta,\theta)\frac{\partial h}{\partial\beta^\top}(h^{-1}(x,\beta,\theta),\beta,\theta), \end{aligned} \tag{4.14}$$

$$\begin{aligned} b_\theta(x,\beta,\theta) &= \frac{\partial^3 h}{\partial^2 x\partial\theta^\top}(h^{-1}(x,\beta,\theta),\beta,\theta)+\frac{\partial\log q}{\partial x}(x,\beta,\theta)\frac{\partial^2 h}{\partial x\partial\theta^\top}(h^{-1}(x,\beta,\theta),\beta,\theta) \\ &\quad + \frac{\partial^2\log q}{\partial^2 x}(x,\beta,\theta)\frac{\partial h}{\partial\theta^\top}(h^{-1}(x,\beta,\theta),\beta,\theta). \end{aligned} \tag{4.15}$$

Consequently, (4.3) is robust against parameter uncertainty when $\psi^0(x)$ is orthogonal to $\frac{\partial^2 \log q^0}{\partial x \partial \beta^\top}(x)$, $b^0_\beta(x)$, and $b^0_\theta(x)$, i.e.,

$$E[\psi^0(x)\frac{\partial^2 \log q^0}{\partial x \partial \beta^\top}(x)] = 0,\ E[\psi^0(x) b^0_\beta(x)] = 0,\ \text{and } E[\psi^0(x) b^0_\theta(x)] = 0. \tag{4.16}$$

Again, we made some additional assumptions that are not restrictive and hold in all our examples. Of course, (4.16) is only a sufficient condition for having $P_{\psi\beta} = 0$ and $P_{\psi\theta} = 0$ which ensure (4.8). However, the interpretation is clear: the uncertainty of β in the density function leads to the first condition in (4.16) while the non observability of x leads to the second and third conditions in (4.16).

The natural question is the derivation of functions $\psi(\cdot)$ such that (4.16) holds. In the case of testing normality, Bontemps and Meddahi (2005) showed that (4.16) holds for Hermite polynomials $H_i(\cdot)$, $i \geq 3$, when one considers a regression-type model like in Examples 1, 2, 3. However, it is not easy to derive such general results for any distribution.

In contrast, by considering ad hoc functions $\psi(\cdot)$, one can always transform them in order to get (4.16), either analytically or by regression. The main goal of the following subsection is to propose the regression approach.

4.2 Univariate transform: robustness by regression

The main idea of the approach is to regress (in population) $\psi^0(\cdot)$ onto the variables $\frac{\partial^2 \log q^0}{\partial x \partial \beta^\top}(\cdot)$, $b^0_\beta(\cdot)$, and $b^0_\theta(\cdot)$. Then by construction, the residual function, denoted $\psi^\perp(\cdot)$, leads to test-functions in (4.3) that are robust against parameter uncertainty:

Proposition 4.4 *Let $\psi(x, \beta, \theta)$ be a test function under the assumptions of Proposition 4.3. Define the function $\psi^\perp(x, \beta, \theta)$ by*

$$\psi^\perp(x,\beta,\theta) = \psi(x,\beta,\theta) - E[\psi(x,\beta,\theta)\zeta^\top(x,\beta,\theta)]\left(E[\zeta(x,\beta,\theta)\zeta^\top(x,\beta,\theta)]\right)^{-1}\zeta(x,\beta,\theta), \tag{4.17}$$

where

$$\zeta(x,\beta,\theta) = \left(\frac{\partial^2 \log q}{\partial x \partial \beta^\top}(x,\beta,\theta), b_\beta(x,\beta,\theta), b_\theta(x,\beta,\theta)\right)^\top. \tag{4.18}$$

Assume that Assumption A1 holds for $\zeta(x, \beta^0, \theta^0)$. Then the test-function (4.3) based on $\psi^\perp(x, \beta, \theta)$ is robust against parameter uncertainty.

This regression approach is also due to Wooldridge (1990) in the context of conditional moment restrictions. Here, we do not have (necessarily) conditional moment restrictions. However, we have a large class of test functions which play the role of the instruments in Wooldridge (1990)'s approach.

4.3 Joint transform: robustness by moments combination

An alternative transform method is the multiplication of the moment condition $m(\cdot)$ in (4.3) by a matrix S such that

$$SP_m = 0.$$

Observe that multiplying $m(\cdot)$ by S is tantamount to multiply $\psi(\cdot)$ by S. Hence, one gets moment conditions that are robust against parameter uncertainty by combining them. As shown in Proposition 4.1, this means that the new moment conditions are orthogonal to the score functions; for more details, see Bontemps, Dufour, Gonçalves and Meddahi (2005).

This approach is not always possible. In particular, one needs that the dimension of m, i.e., p, exceeds the dimension of (β^0, θ^0), denoted k $(p > k)$. In this case, when one assume that P_m has a full rank, a simple choice of S is

$$S = I_p - P_m[P_m^\top P_m]^{-1} P_m^\top. \tag{4.19}$$

This general approach is due to Wooldridge (1990). Note that transforming $m(\cdot)$ by S or by a consistent estimator lead to the same asymptotic distribution.

Of course, the solution (4.19) is not unique. However, one needs to know the form of the matrix P_m. For instance, this is the case of Duan (2003). Also, the method adopted by by Bai (2003) and based on Khmaladze (1981)'s transform is the infinite dimension version of this approach.

4.4 The four examples revisited

We will know turn to our four examples to see how the sufficient conditions of the previous propositions can be simplified.

Example 1: The variable x is observable. Therefore, (4.10) as well as the condition (4.16) become

$$P_{\psi\beta} = E\left[\psi^0(x)\frac{\partial^2 \log q^0}{\partial x \partial \beta^\top}(x)\right] = 0.$$

Example 2: Here, x_t is unobservable and

$$h(z_t, \beta, \theta) = \frac{y_t - w_t^\top \gamma}{\sigma}, \ \theta = (\gamma, \sigma).$$

Therefore,

$$\frac{\partial h^0}{\partial \theta^\top} = [-w_t^\top, -\frac{x_t}{\sigma^0}] \ \text{ and } \ b_\theta^0(x) = \frac{\partial \log q^0}{\partial x}(x)[0, -\frac{1}{\sigma^0}] + \frac{\partial^2 \log q^0}{\partial^2 x}(x)[-w_t^\top, -\frac{x_t}{\sigma^0}].$$

Consequently, (4.10) as well as the first and second conditions in (4.16) become

$$P_{\psi\beta} = E\left[\psi^0(x)\frac{\partial^2 \log q^0}{\partial x \partial \beta^\top}(x)\right] = 0.$$

while (4.11) becomes

$$P_{\psi\theta} = E\left[\left(\frac{\partial^2 \psi^0}{\partial^2 x}(x) + \frac{\partial \log q^0}{\partial x}(x)\frac{\partial \psi^0}{\partial x}(x) + \psi^0(x)\frac{\partial^2 \log q^0}{\partial^2 x}(x)\right)[-w^\top, -\frac{x}{\sigma^0}]\right]$$

and the third condition in (4.16) becomes

$$E\left[\psi^0(x)\left(\frac{\partial \log q^0}{\partial x}(x)[0, -\frac{1}{\sigma^0}] + \frac{\partial^2 \log q^0}{\partial^2 x}(x)[-w_t^\top, -\frac{x_t}{\sigma^0}]\right)\right] = 0.$$

This case was studied by Bontemps and Meddahi (2005) in the case of testing normality. In the particular, they showed that these conditions mean that $\psi(\cdot)$ should be orthogonal to the functions 1 and x (i.e, $H_0(x)$ and $H_1(x)$). When $\psi(\cdot)$ is a Hermite polynomial, $H_i(x)$, i should be greater or equal to 2. The corresponding $m(\cdot)$ function is proportional to $H_{i+1}(\cdot)$.

Example 3: x_t is the residual of a GARCH model with Student error. There is no simplification of the expression of the two matrices.

Example 4: z_t is observable but x_t follows a known distribution (e.g., uniform or standard normal distribution). The parameter β^0 is known: $x_t = Q^{-1} \circ F(z_t, \theta^0) = h(z_t, \theta^0)$ where $F(\cdot)$ is the c.d.f. of z_t and $Q(\cdot)$ is the c.d.f. of a known distribution (e.g., uniform or standard normal distribution). Then $P_{\psi\theta}$ has the general expression (4.11) while $P_{\psi\beta} = 0$.

Proposition 4.5 *We have*

$$P_{\psi\theta} = \frac{\partial(\log f)'}{\partial\theta}(F_{\theta^0}^{-1} \circ Q(x), \theta^0)\frac{q(x)}{f(F_{\theta^0}^{-1} \circ Q(x), \theta^0)}, \tag{4.20}$$

where $f(x, \theta) = \dfrac{\partial F}{\partial x}(x, \theta)$.

This last case is empirically important. Without loss of generality, assume that we transform some observable variable in a normal one. We know from Bontemps and Meddahi (2005) that the Hermite polynomials $H_i(\cdot)$, $i \geq 3$, are robust in the case of Examples 1, 2, and 3. However, this is not the case of Example 4 in general. For instance, assume that $F_\nu(\cdot)$ is the c.d.f. of a T(ν) random variable z_t (while $f_\nu(\cdot)$ is the p.d.f). Then the third condition in (4.16) becomes

$$E\left(\psi(x)\frac{\phi(x)}{f_\nu \circ F_\nu^{-1} \circ \Phi(x)}\left(\frac{x}{\nu + x^2} + \frac{(\nu+1)x}{(\nu+x^2)^2}\right)\right) = 0,$$

which does not hold for Hermite polynomials. In other words, transforming a random variable in a simple one (e.g., uniform or standard normal distribution) does not simple the parameter uncertainty analysis.

5 Pearson's distributions and orthonormal polynomials

In this section, we will follow Johnson, Kotz and Balakrishnan (1994), Wong (1964), and Schoutens (2000) to present the Pearson's class of distributions and their connections with (2.1).

5.1 The general case

The class of Pearson's distribution given by (2.7) can be expanded onto seven types. These types depend crucially on the polynomial $B(\cdot)$ and in particular on its roots (if they exist). In other words, the degree of the polynomial $B(\cdot)$ as well as the roots of $B(\cdot)$ (real or complex; same sign or not) when the degree of $B(\cdot)$ is two. This lead to five cases.

1) $\deg B = 0$ ($c_1 = c_2 = 0$, $c_0 > 0$): The random variable x follows a normal distribution, $\mathcal{N}(-a, c_0)$, and the solution of (2.7) is

$$q(x) = \frac{1}{\sqrt{2\pi c_0}} \exp\left(-\frac{(x+a)^2}{2c_0}\right), \; x \in \mathbb{R}. \tag{5.1}$$

While the normal distribution is with in the Pearson's class, it is not among one of the seven types that we will consider shortly. However, the normal distribution is a limit of all the types.

2) $\deg B = 1$ ($c_2 = 0$, $c_1 \neq 0$): The solution of (2.7) is given by

$$q(x) = K(c_0 + c_1 x)^m \exp\left(-\frac{x}{c_1}\right),\ m = c_0 c_1^{-1} - a,\ x > -\frac{c_0}{c_1} \text{ if } c_1 > 0 \text{ and } x < -\frac{c_0}{c_1} \text{ if } c_1 < 0, \tag{5.2}$$

while K is a constant such that the integral of $q(\cdot)$ over the domain of x equals one. This case corresponds to Type III of Pearson's distributions. When $c_1 > 0$, x follows a gamma distribution, gamma $(m+1, c_1, -c_0/c_1)$. The density function $q(\cdot)$ given in (5.2) may be written as (Johnson, Kotz and Balakrishnan (1994))

$$q(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)}(x-\gamma)^{\alpha-1}\exp(-(x-\gamma)/\beta),\ x > \gamma,\ \alpha = m+1,\ \beta = c_1,\ \gamma = -c_0 c_1^{-1}, \tag{5.3}$$

where $\Gamma(\cdot)$ denotes the Gamma function, i.e.,

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

When $\alpha = 1$, one gets the exponential distribution while one gets Erlang distributions when α is an integer. Finally, When $c_1 < 0$, the random variable $y \equiv -x$ follows a gamma distribution, gamma $(m+1, -c_1, c_0/c_1)$.

3) $\deg B = 2$ and $B(x) = 0$ has two different real roots ($c_2 \neq 0$ and $c_1^2 - 4c_0c_2 > 0$): Denote the roots by a_1 and a_2 and without loss of generality assume $a_1 < a_2$. Then (2.7) becomes

$$\frac{q'(x)}{q(x)} = -\frac{x+a}{c_2(x-a_1)(x-a_2)} = \frac{m_1}{x-a_1} + \frac{m_2}{x-a_2},\ \text{where } m_1 = \frac{a_1+a}{c_2(a_2-a_1)},\ m_2 = -\frac{a_2+a}{c_2(a_2-a_1)}.$$

Consequently, the solution of (2.7) is given by

$$q(x) = K|x-a_1|^{m_1}|x-a_2|^{m_2},\ x \in \{u \in \mathbb{R}, c_2(u-a_1)(u-a_2) > 0\}. \tag{5.4}$$

3-a) If $c_2 > 0$, $\{u \in \mathbb{R}, c_2(u-a_1)(u-a_2) > 0\} = (a_1, a_2)$. Therefore (5.4) becomes

$$q(x) = K(x-a_1)^{m_1}(a_2-x)^{m_2},\ x \in (a_1, a_2), \tag{5.5}$$

which is the density function of a beta distribution, beta (m_1+1, m_2+2), over the interval (a_1, a_2). The density function $q(\cdot)$ in (5.5) may be written as (Johnson, Kotz and Balakrishnan (1994))

$$q(x) = \frac{1}{B(p,q)}\frac{(x-a_1)^{p-1}(a_2-x)^{q-1}}{(a_2-a_1)^{p+q-1}},\ x \in (a_1, a_2),\ p = m_1+1,\ q = m_2+1, \tag{5.6}$$

where $B(\cdot,\cdot)$ denotes the Beta function, i.e.,

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

This case corresponds to Type I while it becomes Type II when $m_1 = m_2$.

3-b) If $c_2 > 0$, $\{u \in \mathbb{R}, c_2(u-a_1)(u-a_2) > 0\} = (-\infty, a_1) \cup (a_2, +\infty)$. However, under the assumption that x is a continuous distribution, x is either in $(-\infty, a_1)$ or in $(a_2, +\infty)$. Therefore, (5.4) becomes

$$q(x) = K(x-a_1)^{m_1}(x-a_2)^{m_2},\ x > a_2 > a_1,$$

or

$$q(x) = K(a_1 - x)^{m_1}(a_2 - x)^{m_2}, \ x < a_1 < a_2.$$

This case corresponds to Type VI and is not a popular example.

4) $\deg B = 2$ and $B(x) = 0$ has a double real root ($c_2 \neq 0$ and $c_1^2 - 4c_0c_2 = 0$): Denote the double root $-C_1$, $C_1 = c_1/2c_2$. Then we have

$$\frac{q'(x)}{q(x)} = -\frac{x+a}{c_2(x+C_1)^2} = -\frac{1}{c_2(x+C_1)} - \frac{a-C_1}{c_2(x+C_1)^2}.$$

Consequently, the solution of (2.7) is given by

$$q(x) = K|x+C_1|^{-1/c_2} \exp\left(\frac{a-C_1}{c_2(x+C_1)}\right), \ x \neq -C_1. \tag{5.7}$$

4-a) If $(a - C_1)/c_2 \neq 0$. For instance, if $(a - C_1)/c_2 < 0$, the integrability of $q(x)$ given in (5.7) when $x \to -C_1$ needs $x + C_1 > 0$. One gets a similar result when $(a - C_1)/c_2 > 0$. Hence, (5.7) becomes

$$\begin{aligned} & q(x) = K(x+C_1)^{-1/c_2} \exp\left(\frac{a-C_1}{c_2(x+C_1)}\right), \ x > -C_1, \\ \text{or} \quad & q(x) = K(-x-C_1)^{-1/c_2} \exp\left(\frac{a-C_1}{c_2(x+C_1)}\right), \ x < -C_1. \end{aligned} \tag{5.8}$$

This case corresponds to Type V. Some Inverse Gaussian distributions are special examples.[2]

4-b) If $(a - C_1)/c_2 = 0$. We start by assuming that $x > -C_1$. Then, (5.7) becomes

$$q(x) = K(x+C_1)^{-1/c_2}.$$

The integrability of $q(x)$ when $x \to -C_1$ needs $1/c_2 \leq 1$ while that integrability needs $1/c_2 > 1$ when $x \to +\infty$. Hence, we have to exclude either $-C_1$ or ∞ from the possible boundaries of the random variable x. If one consider the support (C_1, r) with $r < \infty$, one needs $c_2 \geq 1$ or $c_2 < 0$ while one needs $0 < c_2 < 1$ if one consider the support (l, ∞) with $-C_1 < l$. Similar cases hold when one assumes $x < -C_1$. Consequently, (5.7) becomes

$$\begin{aligned} & q(x) = K(x+C_1)^{-1/c_2}, \ x \in (-C_1, r), \ r < \infty, \ c_2 < 0 \text{ or } 1 \leq c_2, \\ \text{or} \quad & q(x) = K(x+C_1)^{-1/c_2}, \ x \in (l, \infty), \ -C_1 < l, \ 0 < c_2 < 1, \\ \text{or} \quad & q(x) = K(-x-C_1)^{-1/c_2}, \ x \in (l, -C_1), \ -\infty < l, \ c_2 < 0 \text{ or } 1 \leq c_2, \\ \text{or} \quad & q(x) = K(-x-C_1)^{-1/c_2}, \ x \in (-\infty, r), \ r < -C_1, \ 0 < c_2 < 1. \end{aligned} \tag{5.9}$$

These cases are sometimes caller Type VIII (when $c_2 > 0$) and Type IX (when $c_2 < 0$).

5) $\deg B = 2$ and $B(x) = 0$ has no real root ($c_2 \neq 0$ and $c_1^2 - 4c_0c_2 < 0$): We can write

$$c_0 + c - 1x + c_2x^2 = C_0 + c_2(x+C_1)^2, \quad C_0 = c_0 - c_1^2/c_2 \ \text{ and } \ C_1 = c_1/2c_2.$$

[2] The density function of a general Inverse Gaussian distribution is

$$q(x) = \frac{1}{\sqrt{2\pi\beta x^3}} d \exp\left(-\frac{(d-vx)^2}{2\beta x}\right), \ x > 0.$$

Then, we have

$$\frac{q'(x)}{q(x)} = -\frac{x+a}{C_0 + c_2(x+C_1)^2} = -\frac{(x+C_1)}{C_0 + c_2(x+C_1)^2} - \frac{a-C_1}{C_0 + c_2(x+C_1)^2}.$$

Observe that $c_2 C_0 = c_0 c_2 - c_1^2/4 > 0$. Consequently, the solution of (2.7) is given by

$$q(x) = K(C_0 + c_2(x+C_1)^2)^{-1/2c_2} \exp\left(\frac{a-C_1}{\sqrt{c_2 C_0}} \tan^{-1}\left(\frac{x+C_1}{\sqrt{C_0/c_2}}\right)\right). \quad (5.10)$$

5-a) $a - C_1 \neq 0$. This case corresponds to Type IV and is not a popular example and had been recently used by Premaratne and Bera (2001) for modeling stock return data.

5-b) $a - C_1 = 0$. (5.10) becomes

$$q(x) = K(c_0 + c_2 x^2)^{-1/2c_2}.$$

The integrability of $q(\cdot)$ when $|x| \to \infty$ needs $0 < c_2 < 1$. Consequently, the solution of (2.7) is given by

$$q(x) = K(C_0 + c_2(x+C_1)^2)^{-1/2c_2},\ x \in \mathbb{R},\ 0 < c_2 < 1. \quad (5.11)$$

This case corresponds to the Type VII. An important example among this family is the Student T(ν) which corresponds to the case $a = C_1 = 0$ and $c_0 = 1 - c_2$. Therefore, (5.11) becomes

$$q(x) = K\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2},\ x \in \mathbb{R},\ \nu = \frac{1}{c_2} - 1,\ 0 < c_2 < 1. \quad (5.12)$$

Finally, when $\nu = 1$, the distribution T(1) is called the Cauchy distribution.

5.1.1 Orthonormal polynomials

We know that we can easily define an orthonormal polynomial family P_n with respect to the Pearson p.d.f. q. This family can be infinite and dense in $L^2(]l, r[)$ (as in the normal, gamma or uniform case) or finite (student case). This family can be determined by the so-called Rodrigue's formula. The following Proposition 5.1 reviews how P_n can be build from the Pearson distribution $q(\cdot)$. Throughout the paper, we denote by $f^{(n)}(\cdot)$ the n-th derivative function of $f(\cdot)$.

Proposition 5.1 *Let $q(\cdot)$ be the density function of a Pearson's random variable given by (2.7). Assume that there exists an integer N_0 such that $\int_l^r B^{N_0}(x)q(x)dx < +\infty$. For any integer n, define the function $\tilde{P}_n(x)$ by*

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

Then, for any integer n, $\tilde{P}_n(x)$ is a polynomial, whose degree is exactly n, and $\tilde{P}_n(x)$ follows the differential equation:

$$B(x)\tilde{P}_n''(x) + (A(x) + B'(x))\tilde{P}_n'(x) = \lambda_n \tilde{P}_n(x)$$

where

$$\lambda_n = n(B' + A)'(x) + \frac{n(n-1)}{2} B''(x) = n(1 - c_2(n+1)).$$

For $n \le N_0$, define the polynomial sequence $P_n(x)$ by

$$P_n = \frac{\alpha_n}{q(x)} \left[B^n(x)q(x)\right]^{(n)} = \alpha_n \tilde{P}_n(x) \tag{5.13}$$

where

$$\alpha_n = \frac{(-1)^n}{\sqrt{(-1)^n n! d_n \int_l^r B^n(x)q(x)dx}} \text{ and } d_n = \prod_{k=0}^{n-1}\left(A'(x) + \frac{n+k+1}{2}B''(x)\right) = \prod_{k=0}^{n-1}(-1+(n+k+1)c_2)$$

Then

$$P_0(x) = 1,\ \forall n, 1 \le n \le N_0,\ E[P_n(x)] = 0,\ \forall n,m, 0 \le n,m \le N_0,\ E[P_n(x)P_m(x)] = \delta_{n,m};$$

$P_n(x)$ satisfies the recurrence relation

$$xP_n(x) = a_n P_{n+1}(x) + b_n P_n(x) + a_{n-1}P_{n-1}(x) \tag{5.14}$$

where

$$a_n = \frac{\alpha_n d_n}{\alpha_{n+1}d_{n+1}},\ b_n = n\mu_n - (n+1)\mu_{n+1},\ \mu_n = \frac{A(0)+nB'(0)}{A'(0)+nB''(0)} = \frac{-a+nc_1}{-1+2nc_2},\ P_{-1}(x) = 0,\ P_0(x) = 1,$$

as well as the differential equation

$$B(x)P_n''(x) + (A(x)+B'(x))P_n'(x) = \lambda_n P_n(x). \tag{5.15}$$

The proof can be found in Chihara (1978); Eq. (5.13) is known as the Rodrigues' formula.

5.1.2 Special test functions

The orthonormal family can be used as special test functions (as mentioned above, this family can be finite or infinite). Applying (2.1) to $B(x)P_n'(x)$ with (5.15) yields to

$$E[P_n(x)] = 0 \tag{5.16}$$

This property is particularly important in the context of i.i.d. data because we can construct an orthonormal family of moment conditions which yield to simple test statistics. More precisely, assume that when want to test the moment conditions

$$E[m(x)] = 0,\ m(x) = (P_{i_1}(x), P_{i_2}(x), ..., P_{i_p}(x))^\top,$$

where i_1, i_2,...,i_p are p different integers. In this case, the matrix (2.16) equals the identity matrix and the test statistic (2.15) becomes

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right)^\top \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m(x_t)\right) \sim \chi^2(p),$$

implying that the test statistics based on $E[P_{i_j}(x)] = 0$, $j = 1, 2, ..., p$, are asymptotically independent.

In the case of serial correlation, we still have $E[P_n(x_t)P_m(x_t)] = 0$. However, without additional assumptions, one does not have

$$E[P_n(x_t)P_m(x_{t-h})] = 0, \; n \neq m, \; h \neq 0. \tag{5.17}$$

Several scalar diffusion processes have as the stationary distribution the normal $\mathcal{N}(0,1)$ distribution but (5.17) does not hold because it is related to the conditional distribution of the process $\{x_t\}$. In contrast, by assuming that the conditional distribution of x_t given its past values is Gaussian, one gets (5.17). For instance, when one assumes that the process x_t is a normal autoregressive process of order one, AR(1), that is

$$x_t = \gamma x_{t-1} + \sqrt{1-\gamma^2}\,\varepsilon_t, \quad \varepsilon_t \text{ is i.i.d. and } \sim \mathcal{N}(0,1), \quad \text{and} \quad |\gamma| < 1. \tag{5.18}$$

In this case, the Hermite polynomial $H_i(x_t)$ given by

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

are the orthonormal polynomials associated with the $\mathcal{N}(0,1)$ distribution. In addition, each Hermite polynomial $H_i(x_t)$ is an AR(1) process whose autoregressive coefficient equals γ^i, that is

$$E[H_i(x_{t+1})|x_\tau, \tau \leq t] = \gamma^i H_i(x_t). \tag{5.19}$$

In this case, one can show that

$$\Sigma_{ij} = \sum_{h=-\infty}^{+\infty} E[H_i(x_t)H_j(x_{t-h})] = \frac{1+\gamma^i}{1-\gamma^i}\delta_{ij}. \tag{5.20}$$

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

$$\frac{1-\gamma^i}{1+\gamma^i} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_i(x_t)\right)^2 \sim \chi^2(1). \tag{5.21}$$

Thus, by ignoring the dependence of the Hermite polynomial $H_i(x_t)$, one overrejects the normality when $\gamma \geq 0$ or i is even and underrejects otherwise. Monte Carlo simulations in Bontemps and Meddahi (2005) assessed this. This is important in practice since many economic time series are positively autocorrelated.

It is worth noting that Σ is also diagonal for other time series processes, in particular for scalar diffusions whose marginal distribution is among the Pearson's class and the drift is affine. This is the case of the square-root process.

Also, considering only the orthonormal polynomials for testing purposes is in many important cases necessary and sufficient:

Proposition 5.2 *Let $q(\cdot)$ be the density function of a Pearson's random variable with the corresponding orthonormal polynomial sequence $P_n(.)$ and the integer N_0 defined in Proposition 5.1. Assume that $N_0 = \infty$. Then, the density function of a random variable x equals $q(\cdot)$ is and only if*

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

The main argument of the proof is that the sequence of polynomials $P_n(x)$, $n \in \mathbf{N}$, is dense in the set of square-integrable functions with respect to the density function $q(\cdot)$; for a formal proof, see Gallant (1980, Theorem 3, page 192). This proposition means that for statistical inference purposes, in particular testing, one could use orthonormal polynomials only when $N_0 = \infty$. This is the case of normal, gamma, beta and uniform distributions. Unfortunately, this is not the case of the Student distributions.

5.2 Examples of Pearson's distributions

This subsection details the polynomial family that can be used for the most popular Pearson's distribution, i.e., the normal, student, gamma, beta, and uniform distributions.

5.2.1 The Normal distribution

When $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$\frac{\partial(\log q)}{\partial x} = -\frac{x-\mu}{\sigma^2}$$

The test equation (2.1) is the Stein equation:

$$E[\psi'(X) - \frac{x-\mu}{\sigma^2}\psi(X)] = 0$$

Applying (5.13) gives:

$$P_n(x, \mu, \sigma) = \frac{(-1)^n}{\sqrt{n!}} \frac{1}{\phi\left(\frac{x-\mu}{\sigma}\right)} \phi^{(n)}\left(\frac{x-\mu}{\sigma}\right) = H_n\left(\frac{x-\mu}{\sigma}\right)$$

where ϕ is the standard normal p.d.f. and H_n is the normalized Hermite polynomial of degree n. This case was treated in details in Bontemps and Meddahi (2005).

5.2.2 The Student distribution

The probability density function of a $T(\nu)$ is

$$q(x, \nu) = \frac{1}{\sqrt{\nu} B(\frac{\nu}{2}, \frac{1}{2})} \left[1 + \frac{x^2}{\nu}\right]^{-(\nu+1)/2}, \ \nu > 0; x \in I\!R.$$

Hence, we get

$$\frac{\partial(\log q)}{\partial x}(x, \nu) = -(\nu+1)\frac{x}{\nu + x^2}.$$

$A(x) = -(\nu+1)x$ and $B(x) = \nu + x^2$. The specificity is that the family of orthogonal polynomials is not infinite because moments of order greater or equal than ν are not defined. However, using (5.14), we can construct the finite polynomial family (the Romanovski polynomials) that are only defined for $n < \frac{\nu}{2}$:

$$R_n(x, \nu+1) = \sqrt{\frac{(\nu-2n)(\nu-2n-2)}{(n+1)\nu(\nu-n)}} x R_n(x, \nu) - \sqrt{\frac{n(\nu-n+1)(\nu-2n-2)}{(n+1)(\nu-n)(\nu-2n+2)}} R_n(x, \nu-1) \tag{5.22}$$

The first ones are:

$$R_1(x, \nu) = \sqrt{\frac{\nu-2}{\nu}} x, \ R_2(x, \nu) = \sqrt{\frac{\nu-4}{2(\nu-1)}} \left(\frac{\nu-2}{\nu} x^2 - 1\right), \ R_3(x, \nu) = \sqrt{\frac{(\nu-2)(\nu-6)}{6\nu(\nu-1)}} \left(\frac{\nu-4}{\nu} x^3 -\right.$$

5.2.3 The Gamma distribution

The p.d.f. of a gamma (α, β, γ) is:

$$q(x, \alpha, \beta, \gamma) = \frac{(x-\gamma)^{\alpha-1} \exp[-\frac{(x-\gamma)}{\beta}]}{\beta^\alpha \Gamma(\alpha)}, \ \alpha > 0, \ \beta > 0; \ x > \gamma.$$

We have

$$\frac{\partial(\log q)}{\partial x}(x, \alpha, \beta, \gamma) = \frac{\alpha - 1 - \frac{(x-\gamma)}{\beta}}{x-\gamma}$$

The polynomials P_n are related to the generalized Laguerre polynomials $L_n(x, \alpha)$ (orthogonal with respect to the gamma$(\alpha+1, 1, 0)$ distribution):

$$P_n(x, \alpha, \beta, \gamma) = \sqrt{\frac{\Gamma(\alpha)}{n!\Gamma(n+\alpha)}} \frac{[(x-\gamma)^n(x) q(x, \alpha, \beta, \gamma)]^{(n)}}{q(x; \alpha, \beta, \gamma)} = \frac{(-1)^n \sqrt{n!}}{\sqrt{\frac{\Gamma(n+\alpha)}{\Gamma(\alpha)}}} L_n(\frac{x-\gamma}{\beta}, \alpha - 1)$$

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

The first polynomials are :

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

As this family is always defined, it is necessary and sufficient to test our moment equations on this particular polynomials.

5.2.4 The Beta distribution

The p.d.f. of the standard beta distribution $B(\alpha, \beta)$ is:

$$q(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Thus:

$$\frac{\partial(\log q)}{\partial x}(x, \alpha, \beta) = \frac{(-\alpha - \beta + 2)z + \alpha - 1}{x(1-x)}$$

The polynomials are the orthonormalized Jacobi polynomials:

$$P_{n+1}(x, \alpha, \beta) = \frac{1}{a_n} \left((x - b_n) P_n(x, \alpha, \beta) - a_{n-1} P_{n-1} \right)(x, \alpha, \beta)$$

with

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

The first polynomials are:

$$P_1(x,\alpha,\beta) = \sqrt{\frac{\alpha+\beta+1}{\alpha\beta}}\left((\alpha+\beta)x-\alpha\right)$$

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

5.2.5 The Uniform distribution

The main differences with the preceding standard distribution is that the interval is closed and bounded: the p.d.f. is constant equal to one on $[0,1]$. The uniform distribution is a limit of the previous case with α and β equal to zero and with a affine transformation to ensure that $x \in [0,1]$.

The polynomials associated to the uniform distribution are related to the Legendre polynomials L_n. A strict application of (5.13) yields to:

$$\begin{aligned}
P_n(x) &= \frac{1}{n!\sqrt{2n+1}}\left(x^n(x-1)^n\right)^{(n)} = \sqrt{2n+1}L_n(2x-1) \\
&= \sqrt{2n+1}\sum_{k=0}^{n}\begin{pmatrix} n \\ k \end{pmatrix}\begin{pmatrix} n \\ n-k \end{pmatrix}x^k(x-1)^{n-k} \\
&= \frac{\sqrt{2n+1}}{n}\left(\sqrt{2n-1}(2x-1)P_{n-1}(x)-\frac{n-1}{\sqrt{2n-3}}P_{n-2}(x)\right)
\end{aligned}$$

The first polynomials are:

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

6 A Monte Carlo Study

In this section, we provide Monte Carlo simulations to assess the finite sample properties of our test procedures. All the simulations are based on 10 000 replications of samples. Three sample sizes are considered: 100, 500 and 1000. We focus on testing the Student distributional assumption. Various cases are studied: the variable is observable, not observable, in a cross-section case or a time series case. With no *a priori* on the degrees of freedom ν of the Student distribution, it seems very delicate to use polynomial for testing our distributional assumption. We indeed assumed in 2.14 that the variance of the moment used is finite. The order of the polynomial can then not exceed the half of the (unknown) degrees of freedom of the Student distribution we want to test. We will therefore focus on a subset of test-functions $\psi_\alpha(x,\nu) = \frac{x}{(x^2+\nu)^\alpha}$ for various values of α. Following (2.1), the associated moment m_α based on ψ_α is equal to:

$$\begin{aligned}
m_\alpha(x,\nu) &= \psi'_\alpha(x,\nu)+\psi_\alpha(x,\nu)(\log q_\nu)'(x) \\
&= \frac{\nu-(2\alpha+\nu)x^2}{(\nu+x^2)^{\alpha+1}}
\end{aligned} \tag{6.23}$$

whose expectation is equal to zero under the null (q_ν is the p.d.f. of the Student $\mathcal{T}(\nu)$ distribution). The variance can be theoretically derived with simple functions of ν and α. The analytic expression of $\int m_\alpha^2(x,\nu)q_\nu(x)dx$ is given in the appendix.

The variance of m_α is finite for $\alpha > -\frac{\nu}{4}$ which is satisfied for any positive value of α. We thus consider individual tests m_α based on a particular set of positive values $\{0,\frac{1}{2},1,2,3,4\}$ to test our Student distributional assumption. It can be shown that most of these moments are highly correlated. This is the reason why we only perform one joined test (denoted m_j) by testing simultaneously m_0 and m_1.

In the case of observable variables x_t, we know from (4.3) that a sufficient condition for having robustness against parameter uncertainty is to choose a moment $m(\cdot)$ constructed from a test-function $\psi(\cdot)$ such that:

$$E\left[\psi(x,\nu)\frac{x-x^3}{(\nu+x^2)^2}\right]=0$$

First, one can notice that the last equation holds for any even test-function ψ. The moment associated to this function is odd (see (4.3)). Any odd moment condition is also robust for testing the Student distribution. Such moments could be used to have power against asymmetric distribution. However, if the alternative distribution is symmetric, it seems preferable to use even moments *i.e.*, constructed from odd test functions. This is the case for our test-function series $\psi_\alpha(x,\nu)$.

We follow the proposition (4.4) to transform m_α in a moment which is robust against parameter uncertainty. After some calculations provided in the Appendix, we find that the moment $m_\alpha^\perp(x,\nu)$ constructed from the projection of $m_\alpha^\perp(x,\nu)$ on the orthogonal of the score is equal to:

$$m_\alpha^\perp(x,\nu)=\frac{\nu-(2\alpha+\nu)x^2}{(\nu+x^2)^{\alpha+1}}-k_\alpha(\nu)\left(\frac{x^4(\nu+2)-4x^2(\nu+1)+\nu}{(\nu+x^2)^3}\right) \tag{6.24}$$

where $k_\alpha(\nu)$ is given in the Appendix[3]. The variance of $m_\alpha^\perp(x,\nu)$ is also given in the Appendix.

We also compare our tests to the tests developed by Bai (2003) and Duan (2004). These tests handle the problem of serial correlation as well as the problem of parameter uncertainty. The variable of interest is transformed into a Uniform i.i.d variable (Bai's Test) or a Standard Normal i.i.d variable (Duan's Test) by composing first with the conditional c.d.f. . We first notice that this techniques exclude the case of discrete variables and of serial correlation of unknown form. The problem of parameter uncertainty is solved by regression or projection techniques (see section 4).

Bai Test is a Kolmogorov type test where the problem of parameter uncertainty is solved by a continuous projection on the score function.:

$$\xi_{Bai}=\max_{j\in[1..n]}\sqrt{n}\left|\frac{j}{n}-\frac{1}{n}\sum_{k=1}^{j}A_k\right|$$

where n is the sample size and A_k is a function of the score function of the $n-k$ highest values (see Bai (2003) for more details). In the table, we use two statistics. The first one called S_{Bai} is the statistics defined above. The second one S_{Bai}^T is the same expression but maximized

---

[3] $$k_\alpha(\nu)=\frac{\Gamma(\alpha+\nu/2)\Gamma((\nu+1)/2)}{\Gamma(\alpha+(\nu+1)/2)\Gamma(\nu/2)\nu^\alpha}\left(\frac{\frac{(-1+(2\nu+1)/\nu(\alpha+\nu/2)}{(\alpha+(\nu+1)/2)}-\frac{(1+1/\nu)(\alpha+1+\nu/2)(\alpha+\nu/2))}{((\alpha+1+(\nu+1)/2)(\alpha+(\nu+1)/2)}}{\frac{1}{(\nu+1)}-\frac{(3\nu+2)(\nu+2)}{\nu(\nu+1)(\nu+3)}+\frac{(3\nu+1)(\nu+4)(\nu+2)}{\nu^2(\nu+5)(\nu+3)}-\frac{(\nu+1)(\nu+6)(\nu+4)(\nu+2)}{\nu^2(\nu+7)(\nu+5)(\nu+3)}}\right)$$

over the sample trimmed from the 10% highest values. Of course, these two statistics are asymptotically equivalent.

Duan's test first transforms the variable of interest into a standard normal distribution and then splits the sample in blocks of size m for various values of m (1 to 4 in our simulations). It then tests distributional assumptions based on treatment of these blocks. For example the first statistics developed by Duan (called D_1) is based on the fact that the sum of all the elements of a size-m block is a $\mathcal{N}(0, m)$. The second statistics D_2 on the fact that the sum of all the squares of the elements of a size-m block is a $\chi^2(m)$, the third statistics D_3 on the fact that the square of the sum of all the elements of a size-m block divided by m is a $\chi^2(1)$. After a recombination with the c.d.f. of this new statistics we obtain for each of the three statistics $\frac{n}{m}$ (n is the sample size) Uniform variables $Y_{m,1}(\hat{\nu}), .., Y_{m,\frac{n}{m}}(\hat{\nu})$ and $D = \begin{bmatrix} Z_1 \\ \vdots \\ Z_4 \end{bmatrix}$ is asymptotically normal ($Z_i = \frac{1}{\sqrt{i}}[\frac{n}{i}]\sum_{j=1}^{[\frac{n}{i}]} Y_{m,1}(\hat{\nu})$) with a covariance matrix that can be simulated once for all. Duan treats the problem of the parameter uncertainty by multiplying D by some matrix P such that $PE[\frac{\partial D}{\partial \nu}] = 0$ and $PP' = Id$. P depends on the model tested. In the simulations we will choose for commodity a matrix P of size 2×4. There are two ways to compute P. The first strategy consist in first estimating ν and then compute P by simulating $E[\frac{\partial D}{\partial \nu}]$ under the null (i.e. a student with $\hat{\nu}$ degrees of freedom because we are supposed to know the true value of ν). The second strategy which is less time intensive consists in computing the derivative of D with respect to ν in the samples and use these values to compute P. In the simulations we show the two results for P of size 2×4 for the first three statistics of Duan (see Duan (2004) for more details). When we estimate P using the simulation procedure we note the statistics D_1^P, D_2^P, D_3^P and D_1^S, D_2^S, D_3^S otherwise.

6.1 Independent sample, observable variables

We first study the size properties of our tests in Table 1 and Table 2: we start by simulating i.i.d. samples of Student distributions for two values of ν: 5 (Panel A) and 20 (Panel B).

In Table 1, we first assume that ν is known and then unknown. In this case we estimate ν by the second moment. We first focus on the case where ν is assumed to be known (first column of Table 1). Two sets of moments are considered : the moment $m_\alpha(x, \nu)$ and the moment $m_\alpha^\perp(x, \nu)$ for various values of α. We recall that the second ones are robusts against the problem of parameter uncertainty when we estimate ν whereas the first ones are not. In the case where ν is assumed to be known, these moments have certainly the same performances. The variances used to derive the test statistics (2.15) are the theoretical ones (the results are given in the appendix). By looking at the results for Table 1, the finite sample properties of these tests are clearly good for all the values of α and do not depend on the true value of ν. The percentage of rejection are very close to 5% even for very small sample sizes (100 observations). There are also very small differences between $m_\alpha(x, \nu)$ and $m_\alpha^\perp(x, \nu)$ as we are in the case where ν is known. These results will be used as a benchmark for the second case where ν is estimated. Let us now look at the results when ν is estimated by the second moment. Under H_0, $Vx = \frac{\nu}{\nu - 2}$. When the empirical variance is less than 1 (which can happens in very small sample or when the value of ν increases), we decide to put $\hat{\nu}$ equal to 500[4]. Two sub-cases are considered. In the first one (the second column labelled 'in population'), this is the same test-statistics as in the first column except that ν is now estimated. In the second

[4]The performance properties seem insensitive to the choice of this high value.

one (the third column labelled 'in sample') the variances of the test-statistic based on $m_\alpha^\perp(x,\nu)$ are computed within the sample as the average of the square of the moment values.

When ν is estimated, one can not use directly a test based on $m_\alpha(x,\nu)$ by ignoring this first step estimation of ν. By looking at the results on Table 1, we can observe that the percentages of rejections fall dramatically in most of the cases. This fall is however not uniform in α as it is related to the degree of correlation of $m_\alpha(x,\nu)$ with the score function. The less $m_\alpha(x,\nu)$ is correlated to the score, the closer $m_\alpha(x,\nu)$ and $m_\alpha^\perp(x,\nu)$ are. If we use the tests $m_\alpha^\perp(x,\nu)$ which are constructed to be robusts against the estimation of ν, we observe that the results are similar to the results found when ν was assumed to be known. The difference in size performance between knowing ν or estimating ν decreases when the sample size increases or when ν decreases. However, even in the "worst" cases, the results are very good when the sample size is greater or equal to 500. We then look at the performance properties when we estimate the variance within the sample. We can notice that the small distortions which were observed when the variance was computed with the theoretic formula for small sample sizes vanish here.

The size of the Bai Test S_{Bai} is good for low values of ν but the performances are worst when ν increases. We notice than the small sample properties are of the same kind than for our procedures when we estimate the variance and the projection. This is the contrary for S_{Bai}^T which was derived for having good size properties for the standard normal case.

As for the Duan's test, it appears that estimating the partial derivative of the statistics with respect to the parameters in the sample does not deteriorate the properties except for D_1^S. The size properties are quite good. In the following tables concerning the power properties we will use D_2^S and D_3^S which can be computed faster than D_2^P and D_3^P.

In Table 2, we draw the same samples than in Table 1. However, ν is estimated by MLE to measure the potential differences with the case where ν is estimated by the second moment. It happens that, in some cases, especially in very small sample sizes or for high values of ν^0, the MLE failed to converge. It is the same cases where the empirical variance is less than 1. While tracing the likelihood as a function of ν, it appears that it is very flat for high values of ν and that the maximum is clearly in this area. In these cases, we decide to set the estimator value to 500 like for the case when ν was estimated by the second moment. We focus on the performance properties of the same moments used in Table 1. We naturally only consider now moments which are robust against the problem of the parameter uncertainty. In this table, the variance and the projection of $m_\alpha(x,\nu)$ on the orthogonal of the score are computed either theoretically[5] (first 7 rows) or in the sample (second 7 rows). By comparing the results of Table 1 to those of Table 2, we must first notice that the results are insensitive to the fact that ν is estimated by MLE or only by the second moment. Estimating the projection and the variance within the sample seems to improve the small sample properties of our tests. The small sample properties are worse than in the case where everything is known and this difference is also greater when ν increases or the sample size T decreases (because in these cases the estimator is less precise). However, the percentage of over-rejection is quite small as we observe a range 5.0-5.8 for most of the cases. The joined moment $m_j^\perp$ (joined test of $m_0^\perp$ and $m_1^\perp$) is the worst test. This comes from the fact that the correlation between $m_0^\perp$ and $m_1^\perp$ is more difficult to estimate than the unconditional variance of $m_0^\perp$ or of $m_1^\perp$.

In Table 3 we study the power properties of our tests. We focus on two cases: an asymmetric distribution and mixtures of two standard normals. ν is estimated by the second moment assuming that the data are iid Student. All the expectations (for the projection and variance) are computed in the sample. We compare the power properties with those of the tests developed

[5]see the results in the appendix

by Bai (2003) (S_{Bai} and S_{Bai}^T) and Duan (2004) (D_2^S and D_3^S). The reader can see the comments before which explains the construction of the statistics.

When x_t follows a χ^2 with 7 degrees of freedom up to some affine transformation which ensures that the expectation of x is equal to zero and that the variance is equal to the variance of a Student distribution with 5 or 20 degrees of freedom. We can first notice that all the tests have very high power against these alternatives. Bai and Duan Tests have also very good power properties as well as the joint moment.

When x_t is a mixture of two normal distributions, we choose three different values for the weights (p, $1-p$):0.7, 0.8 and 0.9. The standard deviations of the normal distributions are computed in order to obtain the first sixth moments of the mixture equal to the first sixth moments of a Student with 5 degrees of freedom or 20 degrees of freedom. As p increases the distribution of x_t has lower tails and when p tends to one, x_t tends to a standard normal distribution. We perform the same tests than before.

Let us first notice that, as foreseen, the percentage of rejection decreases as p increases in the case $\nu = 5$. In this case, we observe that the Duan and Bai Test have less power in the case $p = 0.9$. When the moments of the mixture are equal to the ones of the Student distribution with 20 degrees of freedom, the power is clearly not good. It is quite impossible to distinguish between a mixture of two normals and a Student distribution when we estimate the degrees of freedom.

6.2 Observable variables with serial correlation of unknown form.

In Tables 4, we consider the case where the variables of interest are observable and serially correlated.

We consider autoregressive process of order one, AR(1), i.e., we assume that the conditional distribution of the variable of interest denoted by x_t given its past is a given function of x_{t-1} with correlation ρ. We consider ρ equal to 0.4 and 0.9.

ν is estimated with the second moment of x_t. We used the same moments than for the independent case. We assume that we have no information on the serial correlation but we take into account this dependence by estimating the variance of the empirical moments (joined or not) by a HAC method. The HAC method is developed by using the quadratic kernel with an automatic lag selection procedure of Andrews (1991). The Bai and Duan tests are not developed to be implementable in such cases and are therefore not in the tables.

For the size properties, x_t is AR(1) with marginal distribution $\mathcal{T}(\nu)$ for ν equal to 5 or 20. x_t is generated by simulating a normal one Gaussian AR(1) process and by dividing each x_t by the square root of an iid gamma process:

$$\begin{aligned}
&u_t|u_{t-1} \sim \mathcal{N}(\rho u_{t-1}; 1-\rho^2) \\
&s_t \sim iid\ \Gamma(\frac{\nu}{2}, \frac{2}{\nu}) \\
&x_t = \frac{u_t}{\sqrt{s_t}} \sim \mathcal{T}(\nu)
\end{aligned}$$

We can observe that the size properties are quite good as soon as ρ is not to close to 1. For $\rho = 0.9$, the percentage of over-rejection increases as it is well known that HAC procedures are getting worse as ρ gets closer to 1 (see Andrews (1991)). When ν increases, the size decrease a little because of the fact that the estimator of ν is less precise.

For the power properties, $x_t = \rho x_{t-1} + \varepsilon_t$. The distribution considered for ε_t is a mixture of normals. The mixtures are the same than in the previous table except that we focus on the two cases $p = 0.7$ and $p = 0.9$. As p increases, the innovation gets closer to a standard normal and the process to a standard normal too. The results show that, in serial correlation cases, we can recover the same size and power properties than in i.i.d. cases even if we have no information on the serial correlation.

6.3 Student GARCH models

At last, we consider a GARCH(1,1) model:

$$\begin{aligned} y_t = \mu + \varepsilon_t,\ \varepsilon_t = \sqrt{v_t}u_t, v_t = \omega + \alpha y_{t-1}^2 + \beta v_{t-1} \\ E[u_t \mid I_{t-1}] = 0,\ E[u_t^2 \mid I_{t-1}] = 1,\ I_{t-1} \equiv \sigma(y_\tau, \tau \leq t-1) \end{aligned} \tag{6.25}$$

where $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$ and $\beta = 0.8$. Here u_t is a linear transformation of a Student distribution (as the marginal variance is equal to one). $u_t = \sqrt{\frac{\nu-2}{\nu}}x_t$ where $x_t \sim \mathcal{T}(\nu)$.

We use different distributions for u_t (up to some affine transformation to ensure that the mean and variance of u_t are respectively equal to 0 and 1): $\mathcal{T}(5)$, $\mathcal{T}(20)$ for the size properties, $\chi^2(7)$ and mixture of normals for the power properties. We consider the same distribution than for the observable cross-section case.

The parameters $\theta \equiv (\mu, \omega, \alpha, \beta)^\top$ are estimated with a Gaussian-QMLE procedure which is known to be consistent provided that the conditional mean and variance process of y_t are correctly specified (Bollerslev and Wooldridge (1992)). We then construct an estimator of u_t by using $\hat{u}_t = \frac{y_t - \hat{\mu}}{\sqrt{\hat{h}_t}}$. Under H_0, u_t is a linear transformation of a Student distribution. We estimate ν by using the fourth moment of u_t as : $Eu_t^4 = \frac{3(\nu-2)}{\nu-4}$.

Following (4.3) and using the fact that $x_t = h(y_t, \mu, \omega, \alpha, \beta, \nu) = \sqrt{\frac{\nu}{\nu-2}}\frac{y_t-\mu}{\sqrt{v_t}}$, we can show that in this particular context the moment $m(.)$ used to test the Student distribution assumption is robust to this first step estimation of $(\mu, \omega, \alpha, \beta, \nu)^\top$ if the test-function $\dot{\psi}()$ is orthogonal to three functions: $\frac{x}{(x^2+\nu)^2}$, $\frac{x^2-\nu}{(x^2+\nu)^2}$ and $\frac{x^3}{(x^2+\nu)^2}$. With the set of test-function $\psi_\alpha(x, \nu)$ used until now, the orthogonality of $\psi_\alpha(x, \nu)$ with $\frac{x^2-\nu}{(x^2+\nu)^2}$ is always satisfied (ψ is odd).

In Table 5 and 6, $m_\alpha^\perp(x, \nu)$ is a moment test constructed from $\psi_\alpha(x, \nu)$ projected on the orthogonal of the space spanned by $\frac{x}{(x^2+\nu)^2}$ and $\frac{x^3}{(x^2+\nu)^2}$. One must first notice that these function are linear combinations of $\psi_1(x, \nu)$ and $\psi_2(x, \nu)$. We can not used any more these two test-functions as their projection on the orthogonal space is zero. The calculations are given in the appendix. The projections are implemented theoretically and not in the sample. In these tables $m_j^\perp$ is the joined test combining $m_0^\perp$ and $m_3^\perp$.

One can notice that the size properties are quite good though we have over-rejection in some cases. The performances are quite similar to Bai test. For the power case (Table 6 for the asymmetric distribution and Table 5 for the mixture of normal distributions), we observe qualitatively the same results than in the observable case with less power due to the fact that we estimate 5 parameters instead of 1.

7 Empirical examples

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

$$y_t = \sqrt{h_t}u_t \quad \text{with } h_t = \omega + \alpha y_{t-1}^2 + \psi h_{t-1}, \text{ where } \omega \geq 0, \alpha \geq 0, \psi \geq 0, \alpha + \psi < 1, \tag{7.1}$$

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

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

The example we consider in our empirical study is testing the Student distribution of the standardized residual u_t up to a scale parameter. The scale parameter ensures that $E[u_t^2] = 1$ which implies that h_t is the conditional variance of y_t given its past observations. Given that the variance of a T(ν) random variable equals $\nu/(\nu-2)$ (with $\nu > 2$), the T-GARCH model corresponds to the case where $\sqrt{(\nu-2)/\nu}\; u_t$ follows a $T(\nu)$ distribution.

We consider the same data as Harvey, Ruiz and Shephard (1994) and Kim, Shephard and Chib (1998),[6] and Bontemps and Meddahi (2005), i.e., observations of weekday close exchange rates from 1/10/81 to 28/6/85. The exchange rates are the U.K. Pound, French Franc, Swiss Franc and Japanese Yen, all versus the U.S. Dollar. After the QML estimation, we get the fitted residuals $\hat{u}_t$ and test their distribution.

We already tested in Bontemps and Meddahi (2005) the normality of u_t and we strongly rejected it, corroborating the results of Kim, Shephard and Chib (1998). These authors estimated the T-GARCH model by the Maximum Likelihood method and find that the degree of freedom of the returns of FF-US$, UK-US$, SF-US$, and Yen-US$, equals 12.82, 9.71, 7.57, and 6.86 respectively.

TO BE COMPLETED.

8 Conclusion

TO BE COMPLETED.

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

Appendix

Table 1: Size of the tests

ν **known.**				ν **estimated by the 2nd moment.**							
				in population				in sample			
Panel A: $\nu = 5$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
m_0	4.6	5.0	5.0	m_0	0.6	0.7	0.9				
$m_{\frac{1}{2}}$	4.9	5.0	5.2	$m_{\frac{1}{2}}$	1.4	1.7	1.8				
m_0	5.1	5.1	5.2	m_1	2.3	2.5	2.6				
m_2	5.2	5.0	5.0	m_2	3.4	3.8	4.0				
m_3	5.2	4.9	5.2	m_3	3.8	4.3	4.7				
m_4	5.0	4.9	5.2	m_4	4.2	4.5	5.0				
m_j	4.6	5.1	4.9	m_j	2.3	2.4	2.4				
KS	4.9	5.0	4.8	KS	4.8	4.8	4.6				
$m_0^{\perp}$	5.1	5.0	5.1	$m_0^{\perp}$	5.0	5.0	5.1	$m_0^{\perp}$	5.1	5.0	5.1
$m_{\frac{1}{2}}^{\perp}$	5.1	4.9	5.2	$m_{\frac{1}{2}}^{\perp}$	5.2	4.9	5.2	$m_{\frac{1}{2}}^{\perp}$	5.1	5.0	5.2
$m_1^{\perp}$	5.1	4.9	5.1	$m_1^{\perp}$	5.1	5.0	5.1	$m_1^{\perp}$	5.0	5.0	5.0
$m_2^{\perp}$	5.1	4.9	5.1	$m_2^{\perp}$	5.1	5.0	5.1	$m_2^{\perp}$	5.0	5.0	5.0
$m_3^{\perp}$	5.1	4.9	5.3	$m_3^{\perp}$	5.3	4.9	5.3	$m_3^{\perp}$	5.1	5.0	5.1
$m_4^{\perp}$	5.1	4.9	5.1	$m_4^{\perp}$	5.1	4.8	5.2	$m_4^{\perp}$	5.2	4.9	5.2
$m_j^{\perp}$	5.0	4.9	5.0	$m_j^{\perp}$	4.8	5.1	5.0	$m_j^{\perp}$	6.0	5.5	5.1
				D_1^P	5.8	5.5	4.7	D_1^S	2.0	1.7	1.5
				D_2^P	5.4	4.5	4.1	D_2^S	5.1	4.4	4.0
				D_3^P	5.2	4.5	5.3	D_3^S	5.1	4.6	5.3
				S_{Bai}^T	1.9	2.5	2.4				
				S_{Bai}	5.2	6.6	5.9				

Table 1 (cont'd): Size of the tests

ν known.				**ν estimated by the 2nd moment.**							
				in population				in sample			
Panel B: $\nu = 20$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
m_0	4.8	5.1	5.1	m_0	0.0	0.0	0.0				
$m_{\frac{1}{2}}$	4.9	5.2	5.0	$m_{\frac{1}{2}}$	0.0	0.0	0.0				
m_1	5.0	5.1	4.9	m_1	0.3	0.1	0.0				
m_2	5.1	5.1	4.9	m_2	0.8	0.3	0.2				
m_3	5.1	5.1	4.9	m_3	1.2	0.7	0.4				
m_4	5.0	5.2	5.1	m_4	1.7	1.1	0.8				
m_j	5.2	5.0	5.1	m_j	1.9	3.4	3.9				
KS	4.3	4.9	4.9	KS	4.1	4.4	4.4				
$m_0^\perp$	5.3	5.1	5.0	$m_0^\perp$	3.5	4.4	4.7	$m_0^\perp$	4.9	4.3	3.9
$m_{\frac{1}{2}}^\perp$	5.3	5.1	5.0	$m_{\frac{1}{2}}^\perp$	3.5	4.3	4.7	$m_{\frac{1}{2}}^\perp$	5.0	4.3	3.9
$m_1^\perp$	5.3	5.1	5.0	$m_1^\perp$	3.4	4.3	4.8	$m_1^\perp$	5.0	4.4	3.9
$m_2^\perp$	5.3	5.1	5.0	$m_2^\perp$	3.4	4.3	4.8	$m_2^\perp$	5.0	4.4	3.9
$m_3^\perp$	5.3	5.1	5.0	$m_3^\perp$	3.5	4.4	4.7	$m_3^\perp$	5.1	4.3	3.9
$m_4^\perp$	5.2	5.1	5.1	$m_4^\perp$	3.4	4.2	4.7	$m_4^\perp$	5.0	4.3	3.9
$m_j^\perp$	5.0	5.2	5.3	$m_j^\perp$	2.5	4.7	5.3	$m_j^\perp$	6.2	8.4	7.8
				D_1^P	5.1	5.4	5.1	D_1^S	2.0	2.1	1.9
				D_2^P	4.3	3.3	3.4	D_2^S	4.9	3.3	3.4
				D_3^P	5.5	4.9	5.0	D_3^S	5.1	5.0	5.0
				S_{Bai}^T	1.4	2.6	2.3				
				S_{Bai}	3.9	6.8	6.3				

Note: The data are i.i.d. from a $\mathcal{T}(5)$ distribution (Panel A) or a $\mathcal{T}(10)$ distribution (Panel B). We test the student distributional assumption. The degrees of freedom are either assumed known or estimated. The results are based on 10000 replications. For each sample size T (100, 500 and 1000), we provide the percentage of rejection at a 5% level. m_α corresponds to the moment test based on the test-function $\psi_\alpha(x,\nu)$, $m_\alpha^\perp$ to the moment robust to the problem of the parameter uncertainty, KS to the Kolmogorov-Smirnov test. $m_\alpha^\perp$ corresponds to the moment test based on the projection of $\psi_\alpha(x,\nu)$ on the orthogonal of the space spanned by the derivative of the score, the projection being computed theoretically (column 'in population') or within the sample (column 'in sample'). By the same way, the variance are computed theoretically or in the sample. $m_j^\perp$ corresponds to the joined test $m_0^\perp - m_1^\perp$. D_1^P, D_2^P, D_3^P correspond to the first three statistics used by Duan (2003) when the matrix of the derivative (see Duan (2003)) is estimated by simulations, D_1^S, D_2^S, D_3^S when the matrix is estimated in the sample. S_{Bai} is the statistics used by Bai (2003), S_{Bai}^T is the same statistics computed on the 10% uper-tail trimmed sample.

Table 2: Size of the tests, ν estimated by MLE.

	$\nu = 5$				$\nu = 20$			
in population	T	100	500	1000	T	100	500	1000
	$m_0^\perp$	5.1	5.0	5.1	$m_0^\perp$	3.5	4.9	4.9
	$m_{\frac{1}{2}}^\perp$	5.2	4.9	5.1	$m_{\frac{1}{2}}^\perp$	3.5	4.8	5.0
	$m_1^\perp$	5.2	5.0	5.1	$m_1^\perp$	3.5	4.8	4.9
	$m_2^\perp$	5.2	5.0	5.1	$m_2^\perp$	3.5	4.8	4.9
	$m_3^\perp$	5.1	4.9	5.2	$m_3^\perp$	3.6	4.8	4.9
	$m_4^\perp$	5.1	4.9	5.1	$m_4^\perp$	3.5	4.9	4.9
	all	4.2	5.0	5.0	all	2.1	4.0	4.2
in sample	T	100	500	1000	T	100	500	1000
	$m_0^\perp$	5.1	4.9	5.1	$m_0^\perp$	5.4	4.7	5.1
	$m_{\frac{1}{2}}^\perp$	5.2	5.0	5.1	$m_{\frac{1}{2}}^\perp$	5.5	4.8	5.0
	$m_1^\perp$	5.2	5.0	5.1	$m_1^\perp$	5.5	4.8	5.1
	$m_2^\perp$	5.2	5.0	5.1	$m_2^\perp$	5.5	4.8	5.1
	$m_3^\perp$	5.2	5.0	5.1	$m_3^\perp$	5.5	4.8	5.0
	$m_4^\perp$	5.2	4.8	5.2	$m_4^\perp$	5.4	4.7	5.1
	$m_j^\perp$	6.0	5.4	5.1	$m_j^\perp$	7.2	9.2	8.3

Note: The data are i.i.d. from a Student distribution with ν equal to 5 and 20. We test the student distributional assumption. The degrees of freedom are estimated by MLE. The results are based on 10000 replications. For each sample size T (100, 500 and 1000), we provide the percentage of rejection at a 5% level. m_α, $m_\alpha^\perp$, $m_j^\perp$ as well as the legend ('in population') or ('in sample') are defined in Table 1.

Table 3: Power of the tests

Asymetric Distribution				Mixture of normals								
				$p = 0.7$			$p = 0.8$			$p = 0.9$		

Panel A: $VX = \frac{5}{3}$**.**

T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	42.6	98.0	100.0	100.0	100.0	100.0	63.4	99.8	100.0	7.0	8.8	10.0
$m_{\frac{1}{2}}^\perp$	42.8	97.8	100.0	100.0	100.0	100.0	73.0	100.0	100.0	7.8	17.4	27.4
$m_1^\perp$	42.4	97.5	100.0	100.0	100.0	100.0	76.8	100.0	100.0	8.7	23.4	38.9
$m_2^\perp$	42.4	97.5	100.0	100.0	100.0	100.0	76.8	100.0	100.0	8.7	23.4	38.9
$m_3^\perp$	42.4	97.6	100.0	100.0	100.0	100.0	72.4	100.0	100.0	7.9	18.5	29.5
$m_4^\perp$	41.9	97.5	100.0	100.0	100.0	100.0	65.6	100.0	100.0	7.3	13.8	20.9
$m_j^\perp$	35.0	95.8	99.9	100.0	100.0	100.0	73.3	100.0	100.0	19.0	76.8	97.8
D_1^S	9.5	66.2	96.4	0.5	0.5	0.6	0.9	0.8	0.8	1.4	1.1	1.0
D_2^S	9.9	40.1	73.6	95.0	100.0	100.0	24.3	86.3	99.0	5.5	13.4	26.8
D_3^S	12.0	44.0	74.1	64.9	90.9	98.5	26.3	92.3	99.7	5.3	9.8	15.8
S_{Bai}^T	29.9	91.8	99.8	52.8	100.0	100.0	6.6	48.0	84.1	2.0	3.1	4.4
S_{Bai}	42.9	94.2	99.8	81.7	100.0	100.0	28.0	85.7	98.6	9.2	22.7	33.4

Panel B: $VX = \frac{20}{18}$**.**

T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^\perp$	9.8	23.5	38.2	4.9	4.2	4.2	5.0	4.6	5.1	5.3	5.2	6.0
$m_{\frac{1}{2}}^\perp$	9.5	24.9	42.0	5.0	4.3	4.2	5.1	4.6	5.1	5.4	5.2	6.0
$m_1^\perp$	9.4	25.6	43.9	5.0	4.3	4.3	5.1	4.6	5.1	5.4	5.2	6.0
$m_2^\perp$	9.4	25.6	43.9	5.0	4.3	4.3	5.1	4.6	5.1	5.4	5.2	6.0
$m_3^\perp$	9.5	24.3	40.6	5.0	4.2	4.2	5.1	4.6	5.1	5.3	5.2	6.0
$m_4^\perp$	9.6	22.6	36.0	5.0	4.3	4.2	5.1	4.6	5.1	5.3	5.2	6.1
$m_j^\perp$	14.5	51.6	85.5	6.2	8.1	6.8	6.0	6.6	6.4	6.7	7.5	7.1
D_1^S	9.2	73.0	98.1	1.5	1.4	1.4	1.4	1.2	1.6	1.4	1.4	1.4
D_2^S	4.4	9.6	18.1	4.5	4.1	3.8	4.8	4.1	3.8	4.3	4.4	4.4
D_3^S	4.0	3.8	3.6	4.5	4.2	4.7	5.1	4.2	4.5	5.0	4.7	5.2
S_{Bai}^T	46.5	99.6	100.0	1.7	2.3	2.5	1.4	2.2	2.2	1.6	2.2	2.5
S_{Bai}	69.1	100.0	100.0	4.3	6.6	6.5	3.7	6.4	5.9	4.0	6.0	6.5

Note: The data are i.i.d. from a $\chi(7)$ distribution standardized in order to have a zero mean and a variance equal to the variance of a Student distribution with 5 degrees of freedom (Panel A) or 20 degrees of freedom (Panel B). Data are also drawn from a mixture of two normal variables with respective weights p and $1-p$ for values of p equal to 0.7, 0.8 and 0.9. The standard deviations of the normal distributions are computed in order to obtain the first sixth moments of the mixture equal to the first sixth moments of a Student with 5 degrees of freedom (Panel A) or 20 degrees of freedom (Panel B). We test the student distributional assumption. The degrees of freedom are estimated by the second moment. The results are based on 10000 replications. For each sample size T (100, 500 and 1000), we provide the percentage of rejection at a 5% level. m_α, $m_\alpha^\perp$, $m_j^\perp$, D_1^S, D_2^S, D_3^S, S_{Bai}^T and S_{Bai} are defined in Table 1. The projection and the variance are computed in the sample and not theoretically.

Table 4: Size and Power under serial correlation

Size properties

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

Power against mixture of normals $p = 0.7$.

	$\nu=5$, $\rho=0.4$			$\nu=5$, $\rho=0.9$			$\nu=20$, $\rho=0.4$			$\nu=20$, $\rho=0.9$		
T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^{\perp}$	79.9	100.0	100.0	7.2	7.4	6.1	9.2	33.5	58.9	12.4	21.9	32.5
$m_{\frac{1}{2}}^{\perp}$	75.4	100.0	100.0	22.1	72.3	94.3	10.4	36.6	63.0	51.3	99.0	100.0
$m_1^{\perp}$	73.0	100.0	100.0	38.2	93.2	99.8	10.9	38.1	64.5	72.7	100.0	100.0
$m_2^{\perp}$	73.0	100.0	100.0	38.2	93.2	99.8	10.9	38.1	64.5	72.7	100.0	100.0
$m_3^{\perp}$	73.9	100.0	100.0	33.5	86.8	98.9	10.5	36.6	62.9	62.9	99.7	100.0
$m_4^{\perp}$	73.2	100.0	100.0	28.8	77.2	96.2	9.9	34.4	59.7	52.1	98.8	100.0
$m_j^{\perp}$	67.9	100.0	100.0	55.0	100.0	100.0	11.6	41.8	68.8	91.6	100.0	100.0

Power against mixture of normals $p = 0.9$.

	$\nu=5$, $\rho=0.4$			$\nu=5$, $\rho=0.9$			$\nu=20$, $\rho=0.4$			$\nu=20$, $\rho=0.9$		
T	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$m_0^{\perp}$	8.0	24.1	41.9	10.1	9.4	9.4	10.6	40.1	67.6	12.8	21.2	32.1
$m_{\frac{1}{2}}^{\perp}$	8.4	22.1	37.9	41.3	96.5	100.0	11.7	43.3	70.9	50.2	99.2	100.0
$m_1^{\perp}$	8.2	20.8	35.4	67.6	99.9	100.0	12.3	44.5	72.1	71.7	100.0	100.0
$m_2^{\perp}$	8.2	20.8	35.4	67.6	99.9	100.0	12.3	44.5	72.1	71.7	100.0	100.0
$m_3^{\perp}$	8.5	21.3	36.4	57.9	99.5	100.0	11.8	43.1	70.6	62.5	99.9	100.0
$m_4^{\perp}$	8.5	21.1	35.8	47.1	97.8	100.0	11.2	40.6	67.8	51.3	99.0	100.0
$m_j^{\perp}$	7.7	19.7	35.4	89.0	100.0	100.0	13.1	46.6	73.0	91.8	100.0	100.0

Note: The data follow an AR(1) process $x_t = \rho x_{t-1} + \sqrt{1-\rho^2}\varepsilon_t$ for $\rho = 0.4$ and $\rho = 0.9$. For the size properties, ε is such that the marginal distribution of x_t is a Student distribution with $\nu = 5$ degrees of freedom and $\nu = 20$. For the power properties, ε_t follows a mixture of two normal variables with respective weights p and $1-p$ for p equal to 0.7 and 0.9. The standard deviations of the two normal variables are computed in order to obtain the first sixth moments of the mixture equal to the first sixth moments of a Student with 5 degrees of freedom and 20 degrees of freedom . We test the student distributional assumption for the marginal density of x_t. The degrees of freedom are estimated by the second order moment. We take into account the serial correlation by estimating the variance matrix through a HAC procedure. The results are based on 10000 replications. For each sample size T (100, 500 and 1000), we provide the percentage of rejection at a 5% level. The notations $m_\alpha^{\perp}$ and $m_j^{\perp}$ are defined in Table 1.

Table 5: Size and Power with GARCH(1,1) DGP

Size				**Power against mixture of normals**							
				$p = 0.7$				$p = 0.9$			
Panel A: $\nu = 5$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
$m_0^\perp$	6.8	7.3	6.2	$m_0^\perp$	92.4	100.0	100.0	$m_0^\perp$	6.8	28.8	60.9
$m_{\frac{1}{2}}^\perp$	6.5	6.9	6.2	$m_{\frac{1}{2}}^\perp$	90.9	100.0	100.0	$m_{\frac{1}{2}}^\perp$	6.8	30.5	62.5
$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—
$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—
$m_3^\perp$	5.8	6.1	5.8	$m_3^\perp$	86.0	100.0	100.0	$m_3^\perp$	6.4	23.1	45.9
$m_4^\perp$	5.6	6.1	5.8	$m_4^\perp$	84.4	100.0	100.0	$m_4^\perp$	6.3	19.5	38.3
$m_j^\perp$	4.9	8.3	7.1	$m_j^\perp$	90.7	100.0	100.0	$m_j^\perp$	6.5	22.8	50.6
D_1^S	0.1	0.2	0.1	D_1^S	0.6	0.6	0.6	D_1^S	0.2	0.2	0.1
D_2^S	6.0	6.1	7.9	D_2^S	85.9	100.0	100.0	D_2^S	7.3	15.1	26.3
D_3^S	5.1	7.5	9.9	D_3^S	63.3	93.6	99.5	D_3^S	7.0	12.8	19.6
S_{Bai}^T	2.0	4.9	6.2	S_{Bai}^T	45.5	98.9	100.0	S_{Bai}^T	4.7	13.1	18.1
S_{Bai}	2.8	8.3	11.0	S_{Bai}	48.3	99.2	100.0	S_{Bai}	5.9	17.2	23.8
Panel B: $\nu = 20$.											
T	100	500	1000	T	100	500	1000	T	100	500	1000
$m_0^\perp$	2.4	5.2	5.9	$m_0^\perp$	2.3	5.0	5.9	$m_0^\perp$	2.5	4.0	5.0
$m_{\frac{1}{2}}^\perp$	2.5	5.2	5.8	$m_{\frac{1}{2}}^\perp$	2.3	4.9	5.8	$m_{\frac{1}{2}}^\perp$	2.5	4.0	5.1
$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—	$m_1^\perp$	—	—	—
$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—	$m_2^\perp$	—	—	—
$m_3^\perp$	2.5	5.0	5.5	$m_3^\perp$	2.4	4.8	5.5	$m_3^\perp$	2.7	4.0	5.1
$m_4^\perp$	2.6	5.0	5.2	$m_4^\perp$	2.4	4.7	5.3	$m_4^\perp$	2.7	3.9	5.2
$m_j^\perp$	1.6	2.7	3.9	$m_j^\perp$	1.4	2.9	4.4	$m_j^\perp$	1.8	2.2	3.2
D_1^S	0.1	0.1	0.1	D_1^S	0.1	0.1	0.1	D_1^S	0.1	0.1	0.1
D_2^S	4.4	3.0	3.0	D_2^S	4.4	3.1	2.9	D_2^S	4.3	3.0	3.0
D_3^S	4.0	3.1	2.9	D_3^S	4.8	2.5	2.9	D_3^S	4.9	2.9	3.0
S_{Bai}^T	1.4	2.9	2.7	S_{Bai}^T	1.5	2.6	2.8	S_{Bai}^T	1.6	3.0	3.1
S_{Bai}	2.2	5.3	5.3	S_{Bai}	2.3	5.0	5.4	S_{Bai}	2.2	5.2	5.4

Note: The data follow a GARCH(1,1) process : $x_t = \mu + \sqrt{h_t} u_t$ with $h_t = \omega + \alpha(\sqrt{h_{t-1}} u_{t-1})^2 + \beta h_{t-1}$ and $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$, $\beta = 0.8$. For the size properties, u_t follows a Student with 5 or 20 degrees of freedom up to some linear transformation which guarantees that $Vu_t = 1$. For the power properties, u_t follows a rescaled mixture of normal with respective weights p and $1-p$ for p equal to 0.7 and 0.9. The standard deviations of the two normal variables are computed in order to obtain the first sixth moments of the mixture equal to the first sixth moments of a Student with 5 degrees of freedom or 20 degrees of freedom. u_t is rescaled in order to have $Eu_t = 0$ and $Vu_t = 1$. μ, ω, α and β are estimated with a QMLE method. ν is estimated using the fourth moment of u_t. We test the student distributional assumption of $u_t\sqrt{\frac{\nu}{\nu-2}}$. The results are based on 10000 replications. For each sample size, we provide the percentage of rejection at a 5% level. The notations m_α, $m_\alpha^\perp$, $m_j^\perp$, D_1^S, D_2^S, D_3^S, S_{Bai}^T and S_{Bai} are defined in Table 1.

Table 6: Power of the tests with GARCH(1,1) DGP against asymmetric innovations.

T	100	500	1000
$m_0^\perp$	12.3	78.0	98.5
$m_{\frac{1}{2}}^\perp$	12.8	79.8	98.8
$m_1^\perp$	—	—	—
$m_2^\perp$	—	—	—
$m_3^\perp$	13.8	82.0	98.9
$m_4^\perp$	13.6	80.6	98.5
$m_5^\perp$	13.3	78.5	97.6
$m_{10}^\perp$	11.3	62.2	85.1
$m_{20}^\perp$	9.0	35.8	49.8
$m_j^\perp$	11.5	78.8	98.9
D_1^S	2.9	63.7	98.2
D_2^S	5.5	10.8	24.2
D_3^S	5.5	8.2	16.4
S_{Bai}^T	7.2	81.5	99.7
S_{Bai}	8.8	87.4	99.9

Note: The data follow a GARCH(1,1) process : $x_t = \mu + \sqrt{h_t}u_t$ with $h_t = \omega + \alpha(\sqrt{h_{t-1}}u_{t-1})^2 + \beta h_{t-1}$ and $\mu = 0$, $\omega = 0.2$, $\alpha = 0.1$, $\beta = 0.8$. u_t follows a chi-square distribution with 7 degrees of freedom up to some affine transformation which guarantees that $Eu_t = 0$ and $Vu_t = 1$. μ, ω, α, β are estimated with a QMLE method. ν is estimated using the fourth moment of u_t. We test the student distributional assumption of $u_t\sqrt{\frac{\nu}{\nu-2}}$. The results are based on 10000 replications. For each sample size, we provide the percentage of rejection at a 5% level. The notations m_α, $m_\alpha^\perp$, $m_j^\perp$, D_1^S, D_2^S, D_3^S, S_{Bai}^T and S_{Bai} are defined in Table 1.

Table 7: Testing the Student distributional assumption of fitted residuals for a GARCH(1,1) model

	UK-US$		FF-US$		SF-US$		Yen-US$	
$\hat{\nu}$	9.61		9.56		6.64		5.54	
$m_0^\perp$	0.09754	(0.75)	1.25273	(0.26)	0.00157	(0.97)	0.00323	(0.95)
$m_{\frac{1}{2}}^\perp$	0.12138	(0.73)	1.09922	(0.29)	0.01311	(0.91)	0.01353	(0.91)
$m_3^\perp$	0.22614	(0.63)	0.70084	(0.40)	0.45082	(0.50)	0.21660	(0.64)
$m_4^\perp$	0.23585	(0.63)	0.66540	(0.41)	0.71038	(0.40)	0.24267	(0.62)
$m_j^\perp$	0.40240	(0.82)	1.77873	(0.41)	1.81173	(0.40)	1.11926	(0.57)
S_{Bai}^T	0.69467	(—)	1.19929	(—)	2.19812	(—)	2.27336	(—)
S_{Bai}	1.03593	(—)	1.26185	(—)	2.31280	(—)	3.02817	(—)
D_1^S	1.92989	(0.38)	2.13736	(0.34)	0.78790	(0.67)	0.69509	(0.71)
D_2^S	0.27112	(0.87)	0.70956	(0.70)	0.65384	(0.72)	3.09756	(0.21)
D_3^S	3.41397	(0.18)	1.77444	(0.41)	6.08390	(0.05)	1.40669	(0.49)

Note: We tested the Student assumption of the standardized residuals. The volatility model is a GARCH(1,1) and is estimated by the Gaussian QML method. We reported the test statistics and their corresponding p-values in parentheses. The data are daily exchange rate returns used by Harvey, Ruiz and Shephard (1994) and Kim, Shephard and Chib (1998). The notations m_α, $m_\alpha^\perp$, $m_j^\perp$, D_1^S, D_2^S, D_3^S, S_{Bai}^T and S_{Bai} are defined in Table 1. The critical values of the Bai-statistics are respectively: 1.94 (1%), 2.22 (5%) and 2.80 (10%).

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

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

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

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

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

Observe that

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

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

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

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

Proof of Proposition 3.1: In the sequel, we use the following notation:

$$< f|g >_0 = \underset{0}{E}[f(x)g(x)] \quad \text{and} \quad < f|g >_a = \underset{a}{E}[f(x)g(x)]$$

1) Let $m(\cdot)$ be a function in $\mathcal{C}_1$, then we have

$$\begin{aligned} \frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{0}{V}[m(x_t)]} &= \frac{\left(\underset{0}{E}\left[m(x_t)\frac{q_a(x_t)}{q_0(x_t)}\right]\right)^2}{\underset{0}{E}[m^2(x_t)]} = \frac{\left(\underset{0}{E}\left[m(x_t)\left(\frac{q_a(x_t)}{q_0(x_t)} - 1\right)\right]\right)^2}{\underset{0}{E}[m^2(x_t)]} = \frac{\left(\underset{0}{E}\left[m(x_t)\ m_1^*(x_t)\right]\right)^2}{\underset{0}{E}[m^2(x_t)]} \\ &= \frac{(< m, m_1^* >_0)^2}{< m, m >_0} \leq\ < m_1^*, m_1^* >_0, \end{aligned} \tag{A.2}$$

where the last inequality holds due to the Cauchy-Schwartz inequality. Therefore, one has the inequality (3.10) given that applying (A.2) to $m(\cdot) = m_1^*(\cdot)$ leads to

$$< m_1^*, m_1^* >_0 \quad = \quad \frac{\left(\underset{a}{E}[m_1^*(x_t)]\right)^2}{\underset{0}{V}[m_1^*(x_t)]}.$$

Also, due to the Cauchy-Schwartz Theorem, the inequality is an equality if and only if $m(\cdot)$ is proportional to $m_1^*(\cdot)$. Finally, observe that

$$< m_1^*, m_1^* >_0 \;=\; \underset{0}{E}\left[\left(\frac{q_a(x_t)}{q_0(x_t)} - 1\right)^2\right] = \underset{0}{E}\left[\left(\frac{q_a(x_t)}{q_0(x_t)}\right)^2\right] - 2\underset{0}{E}\left[\frac{q_a(x_t)}{q_0(x_t)}\right] + 1 = \underset{0}{E}\left[\left(\frac{q_a(x_t)}{q_0(x_t)}\right)^2\right] - 1,$$

which achieves the proof of (3.10).

2) Let $m(\cdot)$ be a function in $\mathcal{C}_2$, then we have

$$\underset{a}{E}[m(x_t)m_2^*(x_t)] = \frac{\underset{a}{E}\left[m(x_t)\dfrac{q_0(x_t)}{q_a(x_t)}\right]}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} - \underset{a}{E}[m(x_t)] = \frac{\underset{0}{E}\left[m(x_t)\right]}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} - \underset{a}{E}[m(x_t)] = -\underset{a}{E}[m(x_t)].$$

Therefore,

$$\frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{a}{E}[m^2(x_t)]} = \frac{\left(\underset{a}{E}[m(x_t)m_2^*(x_t)]\right)^2}{\underset{a}{E}[m^2(x_t)]} = \frac{(< m(x_t) \mid m_2^*(x_t) >_a)^2}{< m(x_t) \mid m(x_t) >_a} \leq \; < m_2^*(x_t) \mid m_2^*(x_t) >_a, \tag{A.3}$$

where the last inequality holds due to the Cauchy-Schwartz inequality. Therefore, one has the inequality (3.11) given that applying (A.3) to $m(\cdot) = m_2^*(\cdot)$ leads to

$$< m_2^*, m_2^* >_a \; = \frac{\left(\underset{a}{E}[m_2^*(x_t)]\right)^2}{\underset{0}{E}[(m_2^*(x_t))^2]}.$$

Also, due to the Cauchy-Schwartz Theorem, the inequality is an equality if and only if $m(\cdot)$ is proportional to $m_2^*(\cdot)$. Observe that

$$\begin{aligned} < m_2^*, m_2^* >_a \; &= \underset{a}{E}\left[\left(\frac{\dfrac{q_0(x_t)}{q_a(x_t)}}{\underset{0}{E}\left[\frac{q_0(x_t)}{q_a(x_t)}\right]} - 1\right)^2\right] = \frac{\underset{a}{E}\left[\left(\dfrac{q_0(x_t)}{q_a(x_t)}\right)^2\right]}{\left(\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]\right)^2} - 2\frac{\underset{a}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} + 1 \\ &= \frac{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}{\left(\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]\right)^2} - 2\frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} + 1 = 1 - \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}, \end{aligned} \tag{A.4}$$

which achieves the proof of (3.11).

3) Let $m(\cdot)$ be a function in $\mathcal{C}_2$, then by using (3.11), one has

$$\begin{aligned} & \frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{a}{E}[m^2(x_t)]} \leq \frac{\left(\underset{a}{E}[m_2^*(x_t)]\right)^2}{\underset{a}{E}[(m_2^*(x_t))^2]} \\ \Leftrightarrow \quad & \left(\underset{a}{E}[m(x_t)]\right)^2 \underset{a}{E}[(m_2^*(x_t))^2] \leq \left(\underset{a}{E}[m_2^*(x_t)]\right)^2 \underset{a}{E}[m^2(x_t)] \\ \Leftrightarrow \quad & \left(\underset{a}{E}[m(x_t)]\right)^2 \left(\underset{a}{E}[(m_2^*(x_t))^2] - \left(\underset{a}{E}[m_2^*(x_t)]\right)^2\right) \leq \left(\underset{a}{E}[m_2^*(x_t)]\right)^2 \left(\underset{a}{E}[m^2(x_t)] - \left(\underset{a}{E}[m(x_t)]\right)^2\right) \\ \Leftrightarrow \quad & \left(\underset{a}{E}[m(x_t)]\right)^2 \underset{a}{V}[m_2^*(x_t)] \leq \left(\underset{a}{E}[m_2^*(x_t)]\right)^2 \underset{a}{V}[m(x_t)] \\ \Leftrightarrow \quad & \frac{\left(\underset{a}{E}[m(x_t)]\right)^2}{\underset{a}{V}[m(x_t)]} \leq \frac{\left(\underset{a}{E}[m_2^*(x_t)]\right)^2}{\underset{a}{V}[m_2^*(x_t)]}, \end{aligned}$$

i.e., the inequality in (3.12). This inequality is an equality if and only if $m(\cdot)$ is proportional to $m_2^*(\cdot)$, otherwise it contradicts the result in 2). Finally, we have

$$\underset{a}{E}[m_2^*(x_t)] = \frac{\underset{a}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} - 1 = \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} - 1$$

while (A.4) implies

$$\underset{a}{E}[(m_2^*(x_t))^2] = 1 - \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}.$$

Therefore,

$$\underset{a}{V}[(m_2^*(x_t))^2] = 1 - \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} - \left(\frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]} - 1\right)^2 = \left(1 - \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}\right) \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}.$$

Consequently,

$$\frac{\left(\underset{a}{E}[m_2^*(x_t)]\right)^2}{\underset{a}{V}[(m_2^*(x_t))^2]} = \underset{0}{E}\left[\frac{q_0(x_t)}{q_a(x_t)}\right]\left(1 - \frac{1}{\underset{0}{E}\left[\dfrac{q_0(x_t)}{q_a(x_t)}\right]}\right) = \underset{0}{E}\left[\frac{q_0(x_t)}{q_a(x_t)}\right] - 1,$$

which achieves the proof of (3.12).■

Proof of Proposition 4.1. For all γ, we have

$$0 = \int n(u,\gamma) f(u,\gamma) du. \tag{A.5}$$

Under some regularity conditions, one can differentiate (A.5) to get

$$\begin{aligned}
0 &= \int \frac{\partial n}{\partial \gamma^\top}(u,\gamma^0) f(u,\gamma^0) du + \int n(u,\gamma^0) \frac{\partial f}{\partial \gamma^\top}(u,\gamma^0) du \\
&= \int \frac{\partial n}{\partial \gamma^\top}(u,\gamma^0) f(u,\gamma^0) du + \int n(u,\gamma^0) \frac{\partial \log f}{\partial \gamma^\top}(u,\gamma^0) f(u,\gamma^0) du \\
&= E\left[\frac{\partial n}{\partial \gamma^\top}(u,\gamma^0)\right] + E\left[n(u,\gamma^0) s(u,\gamma^0)^\top\right],
\end{aligned}$$

which implies (4.9).■

Proof of Proposition 4.2. By applying (2.1) to the functions $\dfrac{\partial \psi^0}{\partial \beta^\top}(x_t)$ and $\dfrac{\partial \psi^0}{\partial \theta^\top}(x_t)$, one gets

$$\begin{aligned}
E\left[\frac{\partial}{\partial x}\frac{\partial \psi^0}{\partial \beta^\top}(x_t) + \frac{\partial \log q^0}{\partial x}(x_t) \frac{\partial \psi^0}{\partial \beta^\top}(x_t)\right] &= 0, \\
E\left[\frac{\partial}{\partial x}\frac{\partial \psi^0}{\partial \theta^\top}(x_t) + \frac{\partial \log q^0}{\partial x}(x_t) \frac{\partial \psi^0}{\partial \theta^\top}(x_t)\right] &= 0,
\end{aligned}$$

which yields to (4.10) and (4.11).■

Proof of Proposition 4.3. By applying (2.1) to the function $\dfrac{\partial\psi^0}{\partial x}(x)\dfrac{\partial h^0}{\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))$, one gets

$$\begin{aligned} E\left[\frac{\partial^2\psi^0}{\partial^2 x}(x)\frac{\partial h^0}{\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right] + E\left[\frac{\partial\log q^0}{\partial x}(x)\frac{\partial\psi^0}{\partial x}(x)\frac{\partial h^0}{\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right] \\ = -E\left[\frac{\partial\psi^0}{\partial x}(x)\frac{\partial^2 h^0}{\partial x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right]. \end{aligned} \tag{A.6}$$

Similarly, by applying (2.1) to the function $\psi^0(x)\dfrac{\partial^2 h^0}{\partial x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))$, one gets

$$\begin{aligned} -E\left[\frac{\partial\psi^0}{\partial x}(x)\frac{\partial^2 h^0}{\partial x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right] = E\left[\psi^0(x)\frac{\partial^3 h^0}{\partial^2 x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right] \\ + E\left[\frac{\partial\log q^0}{\partial x}(x)\psi^0(x)\frac{\partial^2 h^0}{\partial x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right]. \end{aligned} \tag{A.7}$$

By plugging (A.6) and (A.7) in (4.10), one gets

$$\begin{aligned} P_{\psi\beta} = {} & E\left[\psi^0(x)\frac{\partial^2\log q^0}{\partial x\partial\beta^\top}(x)\right] + E\left[\psi^0(x)\frac{\partial^3 h^0}{\partial^2 x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right] \\ & + E\left[\psi^0(x)\frac{\partial\log q^0}{\partial x}(x)\frac{\partial^2 h^0}{\partial x\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right] + E\left[\psi^0(x)\frac{\partial^2\log q^0}{\partial^2 x}(x)\frac{\partial h^0}{\partial\beta^\top}(h^{-1}(x,\beta^0,\theta^0))\right], \end{aligned}$$

i.e., (4.12). A similar proof leads to (4.13).■

Proof of Proposition 4.4. The model is fully parametric. Therefore,

$$\lambda_1 = E[\psi(x,\beta,\theta)\zeta^\top(x,\beta,\theta)] \quad\text{and}\quad \lambda_2 = E[\zeta(x,\beta,\theta)\zeta^\top(x,\beta,\theta)]$$

are functions of β and θ, i.e., there are no additional (nuisance) parameters that appear in the definition of $\psi^\perp(x,\beta,\theta)$. Note however that if one does not know the relationship between λ_1, λ_2, with β and θ, one can include λ_1 and λ_2 in θ. Their estimation will be obtained by doing the regression of $\psi(x,\hat\beta_T,\hat\theta_T)$ on $\zeta(x,\hat\beta_T,\hat\theta_T)$

By construction, $\psi^\perp(x,\beta^0,\theta^0)$ is a linear combination of $\psi(x,\beta^0,\theta^0)$ and the component of $\zeta(x,\beta^0,\theta^0)$ for which Assumption A1 hold. Hence, Assumption A1 holds for $\psi^\perp(x,\beta^0,\theta^0)$, and therefore (4.3) holds for $\psi^\perp(x,\beta^0,\theta^0)$. By construction, (4.16) holds for $\psi^\perp(x,\beta^0,\theta^0)$. Therefore, the test-function (4.3) based on $\psi^\perp(x,\beta,\theta)$ is robust against parameter uncertainty.■

Proof of Proposition 4.5. In this example, Q does not depend on any parameter and F only depends on θ:

$$x = Q^{-1}\circ F(z,\theta) = h(z,\theta)$$

Therefore $P_{\psi\beta} = 0$. We will derive $P_{\psi\theta}$ by using Proposition 4.3. Let us denote the partial derivative of $h(z,\theta)$ with respect to θ by $K_\theta(x)$. We have

$$K_\theta(x) = \left[\frac{\partial h}{\partial\theta}(z,\theta)\right]_{z=F_\theta^{-1}\circ Q(x)} = \left[\frac{\frac{\partial F}{\partial\theta}(z,\theta)}{q\circ Q^{-1}\circ F(z,\theta)}\right]_{z=F_\theta^{-1}\circ Q(x)} = \frac{\frac{\partial F}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta)}{q(x)}$$

Simple calculations give the first and second derivative of $K_\theta(x)$ with respect to x (denoted by $K'_\theta(x)$ and $K''_\theta(x)$):

$$K'_\theta(x) = -(\log q)'(x)K_\theta(x) + \frac{\frac{\partial f}{\partial \theta}(F_\theta^{-1}\circ Q(x),\theta)}{f(F_\theta^{-1}\circ Q(x),\theta)}$$

and

$$\begin{aligned} K''_\theta(x) = -(\log q)''(x)K_\theta(x) - (\log q)'(x)K'_\theta(x) &- \frac{q(x)(f'\circ F^{-1}\circ Q(x))}{f^3(F_\theta^{-1}\circ Q(x),\theta)}\frac{\partial f}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta) \\ &+ \frac{q(x)}{f^2(F_\theta^{-1}\circ Q(x),\theta)}\frac{\partial f'}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta) \end{aligned} \tag{A.8}$$

It is therefore straightforward to prove that

$$\begin{aligned} K''_\theta(x) + (\log q)'(x)K'_\theta(x) + (\log q)''(x)K_\theta(x) &= -\frac{q(x)(f'\circ F^{-1}\circ Q(x))}{f^3(F_\theta^{-1}\circ Q(x),\theta)}\frac{\partial f}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta) \\ &\quad + \frac{q(x)}{f^2(F_\theta^{-1}\circ Q(x),\theta)}\frac{\partial f'}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta) \\ &= \frac{\partial(\log f)'}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta)\frac{q(x)}{f(F_\theta^{-1}\circ Q(x),\theta)} \end{aligned} \tag{A.9}$$

given that

$$\frac{\partial f'}{\partial\theta}(z,\theta) = \frac{\partial(\log f)'}{\partial\theta}(z,\theta)f(z,\theta) + \frac{\partial f}{\partial\theta}(z,\theta)(\log f)'(z,\theta)$$

Hence,

$$P_{\psi\theta} = E\left[\psi(x,\beta^0,\theta^0)\frac{\partial(\log f)'}{\partial\theta}(F_\theta^{-1}\circ Q(x),\theta)\frac{q(x)}{f(F_\theta^{-1}\circ Q(x),\theta)}\right]$$

i.e., (4.20).■

Proof of Eq. (6.24). Several standard calculations give:

$$\begin{aligned} A^\nu_\alpha =< \frac{1}{(x^2+\nu)^\alpha}, 1 >_{q_\nu} &= \frac{1}{\nu^\alpha}\frac{\Gamma(\alpha+\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\alpha+\frac{\nu+1}{2})} \\ < \psi_\alpha(x,\nu), \frac{x-x^3}{(\nu+x^2)^2} >_{q_\nu} &= A_\alpha\left(-1 + \frac{2\nu+1}{\nu}\frac{\alpha+\frac{\nu}{2}}{\alpha+\frac{\nu+1}{2}} - \frac{\nu+1}{\nu}\frac{\alpha+\frac{\nu}{2}}{\alpha+\frac{\nu+1}{2}}\frac{\alpha+1+\frac{\nu}{2}}{\alpha+1+\frac{\nu+1}{2}}\right) \\ < \frac{x-x^3}{(\nu+x^2)^2}, \frac{x-x^3}{(\nu+x^2)^2} >_{q_\nu} &= \left(\frac{1}{\nu+1} - \frac{3\nu+2}{\nu}\frac{\nu+2}{(\nu+1)(\nu+3)}\right. \\ &\quad \left. + \frac{3\nu+1}{\nu^2}\frac{(\nu+4)(\nu+2)}{(\nu+5)(\nu+3)} - \frac{\nu+1}{\nu^2}\frac{(\nu+6)(\nu+4)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)}\right) \end{aligned}$$

Let $k_\alpha(\nu)$ given by

$$k_\alpha(\nu) = \frac{< \psi_\alpha(x,\nu), \frac{x-x^3}{(\nu+x^2)^2} >_{q_\nu}}{< \frac{x-x^3}{(\nu+x^2)^2}, \frac{x-x^3}{(\nu+x^2)^2} >_{q_\nu}}.$$

We can express in a simpler form the moment $m_\alpha^\perp(x,\nu)$ associated to the test function $\psi_\alpha(x,\nu)$:

$$\begin{aligned}
m_\alpha^\perp(x,\nu) &= \frac{\partial}{\partial x}\left(\psi_\alpha(x,\nu) - k_\alpha(\nu)\frac{x-x^3}{(\nu+x^2)^2}\right) - \frac{x}{\nu+x^2}\left(\psi_\alpha(x,\nu) - k_\alpha(\nu)\frac{x-x^3}{(\nu+x^2)^2}\right) \\
&= \underbrace{\frac{\nu-(2\alpha+\nu)x^2}{(\nu+x^2)^{\alpha+1}}}_{l_1(x)} - k_\alpha(\nu)\underbrace{\left(\frac{x^4(\nu+2)-4x^2(\nu+1)+\nu}{(\nu+x^2)^3}\right)}_{l_2(x)}
\end{aligned}$$

The variance of the moment can be computed using the equality:

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

A standard calculation leads to:

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

Computations for the power study.

Moments of the mixtures and of the Student

Panel A: $\nu = 5$

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

Panel B: $\nu = 10$

		EX^2	EX^4	EX^6
$T(\nu)$		1.25	6.25	78.125
$pN(0,\sigma_1^2)+$ $(1-p)N(0,\sigma_2^2)$	$p=0.7$	1.25	6.25	63.52
	$p=0.8$	1.25	6.25	67.05
	$p=0.9$	1.25	6.25	73.63

Panel C: $\nu = 20$

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

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

Computations for the Garch example.
$h(y_t,\mu,\omega,\alpha,\beta,\nu) = \sqrt{\frac{\nu}{\nu-2}}\frac{y_t-\mu}{\sqrt{v_t}}$
Following 4.3, we obtain after the calculations:

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

Hence ψ_α has to be projected on the orthogonal space spanned by ψ_1 and ψ_2 or equivalently the space spanned by ψ_2 and $\frac{x^3}{(\nu^0+x^2)^2}$.

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

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

where

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

The moment used is also expressed as:

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