"Moment-based tests under parameter uncertainty"

Christian Bontemps

Moment-based tests under parameter uncertainty

Christian Bontemps[*]

February 20, 2018

Abstract

This paper considers moment-based tests applied to estimated quantities. We propose a general class of transforms of moments to handle the parameter uncertainty problem. The construction requires only a linear correction that can be implemented in-sample and remains valid for some extended families of non-smooth moments. We reemphasize the attractiveness of working with robust moments, which lead to testing procedures that do not depend on the estimator. Furthermore, no correction is needed when considering the implied test statistic in the out-of-sample case. We apply our methodology to various examples with an emphasis on the backtesting of value-at-risk forecasts.

Keywords: moment-based tests, parameter uncertainty, out-of-sample, discrete distri-

1

butions, value-at-risk, backtesting.

JEL codes: C12, C18.

1 Introduction

Moment-based tests for assessing distributions or particular distribution features (tail properties, kurtosis) are particularly attractive because of their implementation simplicity. These tests are universal because they can consider univariate or multivariate parametric distributions, discrete or continuous distributions, and independent or serially correlated data in the same setting. Moment-based tests have therefore been extensively used in recent papers related to financial econometrics (Amengual and Sentana, 2011; Amengual et al., 2013; Bai and Ng, 2005; Bontemps and Meddahi, 2005, 2012; Candelon et al., 2011; Dufour et al., 2003; Fiorentini et al., 2004; Mencia and Sentana, 2012), forecasting (Diebold and Mariano, 1995, West, 1996, West and Mc Cracken, 1998, McCracken, 2000), and microeconometrics (Butler and Chatterjee, 1997, Tauchen, 1985, Mora and Moro-Egido, 2008). Additionally, structural econometric models provide moment-based equations from testable overidentifying restrictions. For example, structural search models (see Jolivet et al., 2006) provide testable implications for job durations and job offer arrival rates and their relations to the wage distribution.

Very often, these moment equations involve quantities or parameters that must be estimated, generally with the same data set. This is the parameter uncertainty problem that generally modifies the asymptotic distribution of the implied test statistic. Ignoring this issue would lead to an invalid procedure. This is a known problem that has been resolved in different ways. For example, Lilliefors (1980) retabulates the critical values of the Kolmogorov-Smirnov statistic using simulation methods. When introducing their portmanteau test for the white noise process in ARMA models, Box and Pierce (1970) use an approximation of the true distribution, integrating the fact that the parameters of the ARMA process are estimated. Using a moment-based test approach, Tauchen (1985) and Newey (1985) evaluate and correct for the impact of estimation noise. The correction method is explained in Subsection 2.2.2. However, a tractable expression is required for the esti-

mated parameters. This is a regularity condition that may not be satisfied in some contexts, such as for two-step estimators or semiparametric estimations.

In this paper, we use an alternate approach to address this problem by transforming the moment of interest into one that is orthogonal to the underlying score function. We call such a moment a *robust* moment. As further explained below, this transform is a linear correction of the moment of interest for which the weights can be estimated (or calculated) easily. Orthogonality to the score function ensures that local variations of the estimate around the true value (these variations belong to the space spanned by the score) do not affect, at the first order, the information measured by the moment. Our framework can handle both smooth and non-smooth moments. The literature shows that what matters is not the smoothness of the moment but the smoothness of its expectation around the true value (see Tauchen, 1985 or Andrews, 1994). Our orthogonalization strategy systematically exploits the generalized information matrix equality that remains valid for a large class of non-smooth moments, an attractive feature when testing discrete distributions.

Several methods can be used to orthogonalize a given function with respect to the score function. In recent contributions, Bontemps and Meddahi (2012) use the orthogonal projection method, and Wooldridge (1990) modifies the instruments in a conditional distribution setting. Here, we consider a general class of oblique projections. Interestingly, we obtain an analytical expression explicitly involving not the score function but the derivatives of some functionals of interest, which is of particular interest when the score function is difficult to characterize in a closed form. Additionally, it enables us to consider moments in semiparametric models in which one does not have to specify the full structure of the data. We consider such an example in Section 3.

Working with robust moments is of particular interest and is useful for testing purposes. First, we do not have to explicitly characterize the first-order expansion of the estimate because the implied test statistic does not change whether the researcher plugs in the true value of the parameter or a consistent[1] estimate. Second, we can allow for a slower convergence speed than the usual standard square root convergence rate, which is particularly interesting if part of the parameter is estimated at the nonparametric rate. Additionally, a robust moment is robust whether the data

---

[1]See Proposition 3 for the required convergence speed.

are serially correlated or not; therefore, handling dependence is not complicated. The alternative to correcting the statistic, when feasible, requires many calculations to compute the asymptotic distribution of the test statistic, which we avoid here.

Finally, we also prove that working with robust moments is particularly appealing for out-of-sample evaluation. The forecasting literature (see, in particular, West and McCracken, 1998) shows that out-of-sample correction depends on the estimation scheme. With a robust moment, no correction is required, and one can use this robust moment indiscriminately for both in-sample and out-of-sample cases. In Section 3, we consider an out-of-sample case, including, in a Monte Carlo subsection, an evaluation of the small sample properties of our approach and a comparison with existing correction methods.

We also study the power implications of our orthogonalization strategy. First, there is no trivial loss of power when working with robust moments, in comparison to the correction strategy. Second, there is no optimal transform in our projection class with no precise knowledge of the alternative because a particular choice can always be dominated by (or dominate) another choice for another local alternative. The tractability of the test procedure is ultimately the major guideline.

We organize the rest of this paper as follows. Section 2 develops the general framework and characterizes the class of our orthogonalization methods. We then expose their theoretical properties and conduct a local power study. Section 3 characterizes the advantages of using robust moments in out-of-sample contexts and presents some examples. Section 4 describes in detail the backtesting of value-at-risk (VaR) models. In particular, we derive easy-to-compute procedures to test the accuracy of VaR forecasts from a GARCH model. These tests are valid regardless of the true conditional mean and variance used to generate the GARCH. We focus, in particular, on two popular models, the normal GARCH model and the T-GARCH model. Monte Carlo simulations of the proposed tests suggest that the tests perform well in the setups traditionally considered in the literature. Finally, Section 5 considers an empirical application to test the VaR forecasts derived from a T-GARCH(1,1) model for daily exchange rate data. Section 6 concludes the paper. The Supplemental Material contains appendices that provide the proofs (Appendix B) and additional analysis.

2 General results

2.1 Set-up and notations

We consider a sample of T independent or serially correlated observations $(y_1, \ldots, y_T)$, drawn from a univariate random variable Y for which stationarity is assumed. Our goal is to test moment restrictions on these data.

Generally these moment restrictions are derived from an assumption on the distribution of Y. For example, assume that the probability density function of Y belongs to a parametric family of discrete or continuous distributions P_θ indexed by $\theta \in \Theta \subset \mathbb{R}^r$.[2] This assumption implies restrictions that are testable in the data. For example, a Poisson assumption implies that mean and variance of Y are equal. If we assume a Bernoulli distribution with parameter α known, it implies that the mean of Y is equal to α. Note that the resulting test is generally not an omnibus test for the distributional assumption since we select a finite number of moments. Most of the leading tests in the literature are not omnibus either. For example, when testing normality, tests based on skewness and kurtosis measures cannot detect deviation from moments greater than five. However, these tests are frequently used because they are intuitive, easy to implement, and sufficiently powerful for the standard alternatives of interest. Furthermore, one of the advantages of moment-based tests is that we can always adapt the moment to the alternative of interest.[3]

Additionally, our setup includes the case where we test particular features of the data without relying on full distributional assumptions. For example, in forecasting, one is interested in testing whether the one-step-ahead forecast error is orthogonal to the previous period's forecast error, and the marginal distribution is generally left unspecified.

In this paper, θ denotes the vector of parameters that are estimated and is generally estimated

---

[2]We can adapt our framework to the conditioning case in which X gathers explanatory variables that may or may not contain past values of Y in the time series case. In this case, P_θ would become $P_{\theta,x}$, and we would be able to test unconditional moments implied by the conditional distribution of $Y \mid X = x$.

[3]Bontemps et al. (2017) study point optimal moment-based procedures.

in-sample from $y_1, \ldots, y_T$.[4] The true value of parameter θ is denoted by θ^0, and $\mathbb{E}_0$ and $\mathbb{V}_0$ denote, respectively, the expectation and variance under the true data generating process (DGP). The symbol $\top$ denotes the transpose operator, and for two vector-valued functions $h_1(y, \theta)$ and $h_2(y, \theta)$, we denote by $\mathbb{E}_0 \left[h_1 h_2^\top \right]$ the matrix $\mathbb{E}_0 \left[h_1(y, \theta^0) h_2^\top(y, \theta^0) \right]$.

The moment restrictions that we consider are denoted by $m(\cdot)$, a particular k-dimensional vector[5] chosen by the researcher. Under the null hypothesis,

$$\mathbb{E}_0 \left[m(y_t, \theta^0) \right] = 0.$$

Our procedure consists of testing whether the empirical average of these moments is close to zero when θ is also estimated.

2.1.1   One leading example

Financial institutions use VaR forecasts as a measure of risk exposure. Generally, backtesting procedures are required to assess the reliability of the models used to compute VaR forecasts. Following the recent financial crisis, it has become important for financial institutions to hold sufficient capital to sustain potential losses. Although other risk measures can be used in empirical finance, VaR is the most common. Most existing tests are based on the sequence of hits, I_t, of VaR violations. Under perfect accuracy, I_t is i.i.d. Bernoulli distributed with parameter α, the coverage level of the VaR. It implies some moment restrictions that we test, which is one of our leading examples that is addressed in detail in Section 4.

2.1.2   Test statistic when the parameter is known

For a benchmark, we first present the hypothetical case in which the true value θ^0 of the parameter θ is known.

---

[4]See Section 3 for the out-of-sample case.

[5]The k components of $m(\cdot)$ are assumed to be free, that is, the variance of $m(\cdot)$ under the null hypothesis is of full rank.

Assumption CLT - Central limit theorem.

Throughout this paper, we assume that the long-run covariance matrix of $m(\cdot)$, Σ, is finite and positive definite and that the CLT applies.[6]

Under Assumption CLT, a test statistic ξ_m can be constructed from any consistent estimator $\hat{\Sigma}^{-1}$ of Σ^{-1}:

$$\xi_m = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(y_t, \theta^0) \right)^{\top} \hat{\Sigma}^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(y_t, \theta^0) \right). \tag{1}$$

Under the null hypothesis, this statistic asymptotically follows a chi-squared distribution with k degrees of freedom.

In this paper, we do not focus on how to select the moment. First, in many contexts, the researcher has an idea about which moment to test (such as skewness and kurtosis in the normal case). Moreover, the choice of moment is linked to optimality concepts and requires a separate and longer treatment (see Bontemps et al., 2017). Appendix E in the Supplemental Material provides a discussion on how to generate a moment with zero expectation when one considers testing a discrete distribution.

In the above VaR example, a natural moment to check first is to compare the frequency of VaR violations with the expected value, α. The corresponding moment is $m_t = I_t - \alpha$. Additionally, one can consider the property that the hit sequence should be independent of the past by considering the moment $m_t Z_{t-1}$ for any random variable Z_{t-1} in the past information set.

2.2   Test statistic when the parameter is estimated

Next we consider the case in which θ^0 is estimated in the sample. Let $s_\theta(\cdot)$ be the score function of the model. We now detail the regularity assumptions that we impose to write our first-order expansion in Equation (4). These regularity conditions are standard in the GMM literature.

---

[6]Lower-level assumptions that ensure Assumption CLT for $m(\cdot)$ can be found, for example, in Corollary 5.3 of Hall and Heyde (1980).

2.2.1 Regularity conditions

Assumption REG - Regular estimator.

We assume that $\hat{\theta}$, an estimator of θ^0 based on $y_1, \ldots, y_T$, converges almost surely (the regularity conditions can be found in Hansen, 1982) and satisfies the following expansion:

$$\sqrt{T}(\hat{\theta} - \theta^0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w(y_t, \theta^0) + o_P(1),$$

where $w(\cdot)$ is an estimating function that satisfies the CLT and therefore ensures the asymptotic normality of $\hat{\theta}$. The influence function, $w(\cdot)$, can be derived from maximum likelihood (ML) or GMM estimation. In our procedure, we do not need to explicitly know $w(\cdot)$.

Assumption GIM - Generalized information matrix equality.

The GIM equality

$$\left(\frac{\partial \mathbb{E}_0\left[m(y_t, \theta)\right]}{\partial \theta^\top} \right)_{\theta=\theta^0} + \mathbb{E}_0\left[ms_\theta^\top\right] = 0 \tag{2}$$

is satisfied.

The conditions for Assumption GIM to be valid can be found in Tauchen (1985) (it is proved in Theorem 5 under Assumptions 2 and 4) and requires, in particular, the continuous differentiability of $\mathbb{E}_0\left[m(y, \theta)\right]$ with respect to θ in some open neighborhood of θ^0.

First, any moment $m(\cdot)$ continuously differentiable in a neighborhood of θ^0 satisfies Assumption GIM; this expression is used, for example, in Newey and McFadden (1994). Additionally, this paper considers the following class of non-smooth moments:

$$m(y, \theta) = \mathbf{1}\{y \in [l(\theta), u(\theta)]\} - p(\theta), \tag{3}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and l, u, and p are continuously differentiable functions of θ. Such a moment estimates the frequency of a given interval/class and compares it with the expected frequency and is often used in discrete distributions; the Pearson's chi-squared test is a famous example. Following Tauchen (1985), any moment in this class satisfies Assumption GIM.

2.2.2 Asymptotic expansion

The next proposition characterizes the asymptotic distribution of the average of the moment evaluated at the estimated parameter, $\hat{\theta}$.

Proposition 1. *Let $m(\cdot, \theta^0)$ be a moment with zero expectation under the null that satisfies Assumption CLT, and let $\hat{\theta}$ be a square-root consistent estimator of θ^0 that satisfies Assumption REG. Under Assumption GIM, the sequence $m(y_1, \hat{\theta}), \ldots, m(y_T, \hat{\theta})$ satisfies the following expansion:*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(y_t, \hat{\theta}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(y_t, \theta^0) - \mathbb{E}_0 \left[m s_\theta^\top \right] \sqrt{T}(\hat{\theta} - \theta^0) + o_P(1). \tag{4}$$

Equation (4) is generally known in the differentiable case because it is the first-order expansion in which $\mathbb{E}_0 \left[m s_\theta^\top \right]$ is replaced by $-\mathbb{E}_0 \left[\frac{\partial m}{\partial \theta^\top} \right]$, which is proved by Tauchen (1985) in the non-smooth case (see Theorem 2 of Tauchen). We are not the first to use this equation, but we exploit it here systematically for testing in an alternate form.

In standard cases, plugging in $\hat{\theta}$ for θ^0 generally modifies the asymptotic variance, as the above equation indicates. Ignoring this change would lead to size distortion, a problem of empirical relevance because assumptions that should be rejected might not be and conversely. This distortion level depends on the covariance between $m(\cdot)$ and the score function as well as on the estimating function used to estimate parameter θ^0. Equation (4) highlights the two strategies to address the impact of parameter uncertainty.

The first strategy, which we call *correcting* hereafter, consists of deriving the joint asymptotic distribution of the two terms on the right-hand side of (4), as in Newey, 1985, Mora and Moro-Egido, 2008, for the probit case, and Escanciano and Olmo, 2010, for the VaR example. However, this is not always possible because the score may not be properly defined, as in a semiparametric GARCH model, or when the influence function $w(\cdot)$ is not easy to derive, as in two-step estimators. This strategy can also be very cumbersome in some cases, for example in time series.

The second strategy involves working with moments $m(\cdot)$ orthogonal to the true score function, which we call *robust* moments. For robust moments, the asymptotic distribution of $\frac{1}{\sqrt{T}} m(y_t, \hat{\theta})$ is the same, at the first order, as the asymptotic distribution of $\frac{1}{\sqrt{T}} m(y_t, \theta^0)$ because the second term

on the right-hand side of Equation (4) is equal to zero. Thus, we do not have to consider the estimation impact. For example, Bontemps and Meddahi (2005) find that Hermite polynomials of degree 3 or more can be used for the normality testing of generalized regression model (including GARCH) residuals.

In this paper, we transform any moment $m(\cdot)$ into a moment that is robust. We propose general projection methods that can transform any moment into a moment orthogonal to the score function. Note that any method in the literature that builds robust moments implicitly or explicitly transforms a moment into a moment orthogonal to the score function. For example, Wooldridge (1990) considers moment-based tests for conditional distributions. In his framework, the matrix involved is the full expectation with respect to the joint distribution of Y and X. He proposes transformation of the instruments $h(X)$ to obtain orthogonality with respect to this joint distribution and does not refer to the score function. Bontemps and Meddahi (2012) propose projection of the moment of interest $m(\cdot)$ orthogonally onto the space $S^{\perp}$, the space orthogonal to the space spanned by the score. Specifically, the transformed moment is

$$m^{\perp}(y, \theta) = m(y, \theta) - \mathbb{E}_0\left[ms_\theta^{\top}\right] \mathbb{V}_0[s_\theta]^{-1} s_\theta(y). \tag{5}$$

2.3 Orthogonalization methods

In this subsection, we introduce our general class of oblique projection transforms. These transforms generalize the orthogonal projection of Bontemps and Meddahi (2012). Interestingly, the robust moment can be characterized without explicitly mentioning the score function.

2.3.1 Robustification by oblique projection

Consider an estimating function $g(\cdot)$ that can identify[7] parameter θ and satisfy Assumption CLT and Assumption GIM. This estimating function can be used to estimate θ, but we do not impose it

---

[7]It means

$$\mathbb{E}_0\left[g(y, \theta)\right] = 0 \text{ iff } \theta = \theta^0.$$

here. We assume that $g(\cdot)$ has the same dimension as θ, as in the identifying restrictions of a GMM procedure. We denote by $\tilde{m}_g$ the projection of $m(\cdot)$ onto $S^\perp$ parallel to direction g.

Proposition 2. *Let $\tilde{m}_g$ be the projection of $m(\cdot)$ onto $S^\perp$ parallel to direction g. $\tilde{m}_g$ can be expressed as*

$$\tilde{m}_g(y,\theta) = m(y,\theta) - \left(\frac{\partial \mathbb{E}_0\left[m(y_t,\theta)\right]}{\partial \theta^\top}\right)_{\theta=\theta^0} \left(\frac{\partial \mathbb{E}_0\left[g(y_t,\theta)\right]}{\partial \theta^\top}\right)_{\theta=\theta^0}^{-1} g(y,\theta), \tag{6}$$

and this moment is robust to parameter estimation uncertainty.

Proof. Equation (6) exploits the GIM equality for $m(\cdot)$,

$$\left(\frac{\partial \mathbb{E}_0\left[m(y_t,\theta)\right]}{\partial \theta^\top}\right)_{\theta=\theta^0} = -\mathbb{E}_0\left[ms_\theta^\top\right],$$

and for $g(\cdot)$. Therefore, $\tilde{m}_g$ can also be expressed as

$$\tilde{m}_g(y,\theta) = m(y,\theta) - \mathbb{E}_0\left[ms_\theta^\top\right]\mathbb{E}_0\left[gs_\theta^\top\right]^{-1}g(y,\theta). \tag{7}$$

This moment is clearly orthogonal to the true score function. Note that (5), the orthogonal projection onto the orthogonal space spanned by the score, is a specific case of (7) with $g = s_\theta$. □

Equation (6) in Proposition 2 is one of our important results. Observe first that this transform is a simple linear correction (that exploits the generalized information equality) that depends on only $m(\cdot)$, the moment tested, and $g(\cdot)$, the estimating function chosen. Moreover, the expression in (6) does not use the score function but the derivatives of the functions of interest with respect to θ. In many cases, these quantities are easy to derive analytically (see the examples in Section 3 and 4). If it is not possible to obtain a closed form for the expectation, it is still possible to estimate these quantities in the data. Moreover, if $m(\cdot)$ is smooth, one can simplify the first matrix in Equation (6) because

$$\left(\frac{\partial \mathbb{E}_0\left[m(y_t,\theta)\right]}{\partial \theta^\top}\right)_{\theta=\theta^0} = \mathbb{E}_0\left[\frac{\partial m}{\partial \theta^\top}\right],$$

and similarly for $g(\cdot)$.

As we discuss later in Subsection 2.3.4, many choices exist for $g(\cdot)$. The empirical researcher should be aware that there is no "best choice" for $g(\cdot)$ without a specific alternative. According to

the testing literature, a closed-form expression provides better small-sample performance because it avoids imprecise quantity estimates. The ultimate guideline is to choose the estimating function $g(\cdot)$ that appears to be most tractable.

2.3.2 Advantages of working with robust moments

Working with robust moments has several advantages. Since the test statistic is insensitive to the quality of the estimates, it depends on only the choice of the moment. Therefore, the critical values of the test statistic can be tabulated using either the asymptotic distribution or by simulation (bootstrap or Monte Carlo techniques can be used to improve the small-sample properties).

Additionally, a robust moment is robust whether the data are i.i.d. or serially correlated. The alternative, which consists of correcting the statistic, could require numerous calculations to compute the covariance between the first and second terms in Equation (4), which we avoid here. Moreover, the same argument holds for two-step estimators, in cases where the influence function is not easy to derive, and when the asymptotic distribution is non-standard.

We now present another interesting property of robust moments; we can indeed loosen Assumption REG.

Proposition 3. *Let* $\tilde{m}_g$ *be a robust moment defined as above. When* $T^\alpha(\hat{\theta} - \theta^0) = O_P(1)$ *for* $\alpha > 1/4$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{m}_g(y_t, \hat{\theta}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{m}_g(y_t, \theta^0) + o_P(1).$$

In some cases, the parameters of interest have slower convergence rates. For example, Manski's maximum score estimator converges at a slower rate than $1/2$. Additionally, the convergence rate of the estimates of private values in auction models estimated nonparametrically is also slower than the standard square-root rate. In these cases, the usual correction strategy requires further investigation (Expansion (4) is indeed not valid when θ has a convergence rate slower than $T^{1/2}$). Proposition 3 shows that a testing procedure derived from a robust moment remains a solution for testing in less-regular cases.

2.3.3 A simplified procedure: building robustness from an auxiliary model

While Proposition 2 provides a strategy for building a robust moment, its attractiveness depends on the choice of $g(\cdot)$. Here, we propose a simplified procedure to construct robust moments. This procedure can be used when the parameters of interest can be concentrated out.

Consider a simple model $\tilde{\mathcal{M}}$ (the auxiliary model) defined by the parametric family of distributions $\tilde{P}(y_t; \beta)$, and let $s_\beta(\cdot)$ be the score for this auxiliary model. Assume further that our true model can be concentrated and linked to this auxiliary model by $\beta = h(X_{t-1}, \theta)$, where $h(X_{t-1}, \cdot)$ is a smooth function in the neighborhood of the true value and X_{t-1} is a collection of variables such that, conditional on X_{t-1}, the distribution of y_t is in $\tilde{P}(y_t; \beta)$. A moment orthogonal to $s_\beta(\cdot)$ in the auxiliary model is also orthogonal to the true score in the true model.[8] This approach is particularly appealing because in some cases, it is easier to build a moment orthogonal to the score for an auxiliary model than for the true model. Interestingly, such a moment remains robust regardless of the functional form $h(\cdot)$. We illustrate this result with two examples.

VaR Example Consider the following model for financial returns:

$$r_t = \mu(J_{t-1}, \beta) + \sigma(J_{t-1}, \beta)\varepsilon_t,$$

where J_{t-1} is the information set at time $t-1$, ε_t is an i.i.d. variable with a known distribution and β is a vector of parameters. Here, we can define $\mu = \mu(J_{t-1}, \beta)$ and $\sigma = \sigma(J_{t-1}, \beta)$; our auxiliary model is therefore the constant location-scale model

$$r_t = \mu + \sigma\varepsilon_t.$$

In this auxiliary model, one can apply Proposition 2 from an estimating function $g(\cdot)$ (the first two moments, for example) to build a robust moment. Following our result above, this robust moment is also robust for the GARCH model just introduced.

This characteristic is particularly important in practice because the robustness of this moment is valid regardless of the specification of $\mu(\cdot)$ and $\sigma(\cdot)$ of the GARCH model, which makes our

---

[8]See proof in subsection B.2 in the Supplemental Material.

approach interesting for financial regulators. Ignoring parameter uncertainty may distort the results and lead to not rejecting a VaR model that should be rejected. The systematic use of this robust moment approach controls for this problem even without precise knowledge of the true underlying models. We detail the implementation in Section 4.

Testing a moment in a parametric family The previous example can be generalized to the case of any robust moment with zero expectation under a given parametric distribution. If the parameter of this distribution β is linked to some exogenous variables X, $\beta = h(X, \theta)$, where θ is a parameter vector to be estimated, the same moment remains robust when θ is estimated from the data. Following Proposition 3, the convergence rate for θ can also be the nonparametric rate.

2.3.4 Local power properties and choice of g

We presented above the main advantages of using a robust moment in a moment-based test. However, a successful testing procedure must control the size to ensure validity and also have good power properties, at least with respect to the usual alternatives. Here, we might wonder whether the projection strategy may systematically decrease the power compared to the correction strategy. Additionally, the choice of $g(\cdot)$ could influence the power properties of the testing procedure. We investigate these two questions in this section.[9]

First, we need to stretch the fact that a moment is robust independently of the choice of the parameter estimator. If the estimating function $g(\cdot)$ used to estimate the parameters is the one used to make the moment robust, i.e., we use $\tilde{m}_g(\cdot)$ for our robust moment, the two test statistics result in the same numerical value. Consequently, there is no trivial loss of power for our strategy because it coincides with the correcting strategy for some specific choice of the estimating function $g(\cdot)$ in Equation (6). Again, remember that the correction strategy, which depends on the estimator, can be numerically challenging when the influence function of the estimator is not tractable.

Next, we turn to the choice of $g(\cdot)$, the direction of the oblique projection. Proposition 2 does not impose any particular requirement on $g(\cdot)$. Equation (6) illustrates that some choices of

---

[9]The proof is provided in Section B.3 of the Supplemental Material.

$g(\cdot)$ can provide closed-form expressions. Clearly, a closed-form expression helps to improve the small-sample properties of the test derived from this moment. The next proposition calculates the parameter that drives the power property of a test based on g.

Proposition 4. *Let $g(\cdot)$ be some estimating function for θ like in Proposition 2 and $\tilde{m}_g$ the robust version of $m(\cdot)$ after projection along the direction g (see Equation (6)). Assume that, under the (local) alternative, the p.d.f is $q_1 = q_0(1 + h(y)/\sqrt{T})$. The power function from the test based on $\tilde{m}_g$ is an increasing function of the parameter:*

$$a(g) = \frac{\mathbb{E}_0\left[\tilde{m}_g h\right]^2}{\mathbb{V}_1[\tilde{m}_g]}.$$

Proposition 4 allows us to prove that no optimal choice of $g(\cdot)$ for power maximization exists uniformly. Indeed, the power of the test for a specific choice of $m(\cdot)$ depends on the local alternative, $h(\cdot)$, considered. Consequently, for any choices of $g_1(\cdot)$ and $g_2(\cdot)$, there exist two local alternatives such that $\tilde{m}_{g_1}$ is better than $\tilde{m}_{g_2}$ in the first case and the reverse is true for the second case. Without any specific direction of departure from the null, our suggestion is to select the estimating function $g(\cdot)$ that appears to be the most tractable.

3 Out of sample evaluation of robust moments

In this section, we focus on moments evaluated out-of-sample. As noted by West and McCracken (1998) in particular, the estimated parameters modify, as in the in-sample case, the asymptotic distribution of the test statistic. Moreover, the correction they provide depends on the estimation scheme (recursive, rolling, or fixed) and the ratio between the number of out-of-sample observations and the sample size used for the parameter estimation. McCracken (2000) extends the approach to the case of non-smooth moments.

3.1 Invariance of the robust moments to the estimation scheme

One may use our robust moments instead of correcting. The following proposition states that a robust moment leads to invariant statistics, even for out-of-sample evaluations.

Proposition 5. *Let $\hat{\theta}_t$ be a sequence of square-root-consistent GMM-type estimators of θ^0 using the data $y_{t-R}, \ldots, y_{t-1}$ (rolling estimator), $y_1, \ldots, y_{t-1}$ (recursive estimator), or $y_1, \ldots, y_R$ (fixed estimator). We assume that $\hat{\theta}_t$ satisfies Assumption REG for the corresponding values of the time index. We also assume that R and P tend to ∞ while $\sqrt{P}/R$ tends to 0 and that $m(\cdot)$ satisfies Assumptions CLT and GIM. If $m(\cdot)$ is a robust moment,*

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{R+P} m(y_t, \hat{\theta}_t) = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{R+P} m(y_t, \theta^0) + o_P(1). \tag{8}$$

The proof is a direct consequence of the fact that the second term in the asymptotic expansion vanishes owing to orthogonality of m to the score function. Observe, however, that working with a robust moment allows us to loosen the requirement of a finite limit of P/R in West and McCracken (1998), which is generalized in McCracken (2000) for non-smooth moments.

The intuition is the same as for the in-sample properties. A robust moment is orthogonal to the score and is therefore uncorrelated with the local deviations of $\hat{\theta}$ around θ^0.

Therefore, when the moments are robust, the asymptotic variance of the out-of-sample averages of these moments is the standard long-run variance. We do not have to correct for the estimation scheme, which further demonstrates why robust moments are attractive.

3.2 Derivation of robust moments in different examples

We now complement this result by deriving robust moments for some of the tests proposed in West and McCracken (1998) and in McCracken (2000). For simplicity, we omit, in our notations, the dependence of the functions on y and θ.

3.2.1 Testing for first-order correlation in a regression model

Consider the stationary model[10]

$$y_t = x_t^\top \theta^0 + \varepsilon_t.$$

The goal is to test whether $\mathbb{E}_0 \left[\varepsilon_t \varepsilon_{t-1} \right] = 0$ from the estimated residuals; $\hat{\varepsilon}_t = y_t - x_t^\top \hat{\theta}_t$ is computed using one of the three different schemes.

---

[10]Model 6.1 in West and McCracken (1998).

West and McCracken (1998) propose a scheme-dependent correction method and a simple procedure to correct for parameter uncertainty through auxiliary regressions. In our approach, we transform the moment $m = \varepsilon_t \varepsilon_{t-1}$ into one that is robust to parameter uncertainty. On the basis of Proposition 2 and because $g = (y_t - x_t^\top \theta)x_t$, one can derive the matrices of interest as follows, noting that the moments and the estimating function are both smooth functions of the parameter.

$$\mathbb{E}_0 \left[\frac{\partial m}{\partial \theta} \right] = -\mathbb{E}_0 \left[x_t^\top \varepsilon_{t-1} \right] \text{ and } \mathbb{E}_0 \left[\frac{\partial g}{\partial \theta^\top} \right] = -\mathbb{E}_0 \left[x_t x_t^\top \right].$$

Thus, the robust version of m is

$$m^\perp = \varepsilon_t \varepsilon_{t-1} - \mathbb{E}_0 \left[x_t^\top \varepsilon_{t-1} \right] \mathbb{E}_0 \left[x_t x_t^\top \right]^{-1} x_t \varepsilon_t. \tag{9}$$

The two expectations above can be simultaneously estimated in the sample with the estimation of θ^0. Closed-form expressions can be derived with more structure. For example, an $AR(1)$ model without a constant term, $y_t = \rho y_{t-1} + \varepsilon_t$, leads to the following moment:

$$m^\perp = \varepsilon_t \varepsilon_{t-1} - (1 - \rho^2) y_{t-1} \varepsilon_t.$$

3.2.2 Encompassing test

Following West and McCracken (1998), we consider the encompassing test. Consider model 1, $y_t = x_{1t}^\top \beta_1^* + v_{1t}$, and the encompassing test with model 2, $y_t = x_{2t}^\top \beta_2^* + v_{2t}$. This process consists of testing

$$\mathbb{E}_0 \left[v_{1t}(x_{2t}^\top \beta_2^*) \right] = 0.$$

The algebra is similar to that of the previous case. The robust version of the test replaces $m = v_{1t}(x_{2t}^\top \beta_2^*)$ with

$$m^\perp = v_{1t}(x_{2t}^\top \beta_2^*) - \mathbb{E}_0 \left[x_{1t}(x_{2t}^\top \beta_2^*) \right] \left(\mathbb{E}_0 \left[x_{1t} x_{1t}^\top \right] \right)^{-1} x_{1t} v_{1t}.$$

3.2.3 Test of equal MAE

We next consider a non-smooth moment and test for equal mean absolute error between model 1, $y_t = x_{1t}^\top \beta_1^* + v_{1t}$, and model 2, $y_t = x_{2t}^\top \beta_2^* + v_{2t}$. The moment considered in McCracken (2000) is

$$m = |u_{1t}| - |u_{2t}|,$$

which satisfies the regularity conditions given in Section 2. We define the robust version according to Equation (6) as follows. For each model i, $i = 1$ or 2, the first matrix $\left(\frac{\partial \mathbb{E}_0 \left[|y_t - x_{it}^\top \beta_i^*| \right]}{\partial \beta_i^\top} \right)_{\beta_i = \beta_i^*}$ is equal to $-\mathbb{E}_0 \left[sgn(u_{it}) x_{it}^\top \right]$. Because $g(\cdot)$ is smooth, we obtain the usual expression for the second matrix; that is, $-\mathbb{E}_0 \left[x_{it} x_{it}^\top \right]$. Consequently,

$$m^\perp = u_{1t}^\perp - u_{2t}^\perp,$$

where $u_{it}^\perp = |u_{it}| - \mathbb{E}_0 \left[sgn(u_{it}) x_{it}^\top \right] \left(\mathbb{E}_0 \left[x_{it} x_{it}^\top \right] \right)^{-1} u_{it} x_{it}$, $i = 1, 2$.

3.3 A small Monte Carlo exercise

Now, we consider the first-order serial correlation test when one estimates an $AR(1)$ model with mean:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t.$$

This is a Monte Carlo exercise developed in West and McCracken (1998). The experiment involves 10 000 replications. The robust moment, as calculated above, is $\varepsilon_t \varepsilon_{t-1} - (1 - \rho^2) \varepsilon_t (y_{t-1} - \mu)$, which has long-run variance $\rho^2 \sigma^4$, where σ^2 is the variance of ε_t. In all the tables, we report the rejection frequencies for a 5% level test.

We first present in-sample results where we compare our test with the famous Box-Pierce (1970) test[11] with two correlations and its corrected version proposed by Ljung and Box (1978). Both tests consider parameter uncertainty; however, we focus on only the first autocorrelations for

---

[11] $BP(K) = T \sum_{h=1}^K \hat{\rho}_\varepsilon^2(h)$, which follows a χ^2 distribution with $K - 1$ degrees of freedom in our case. We present the results for $K = 2$.

a fair comparison.[12] Under the null hypothesis, as in our proposed test, $BP(2)$ is asymptotically $\chi^2(1)$ distributed when μ and ρ are estimated.

Table 1 shows the size results for four sample sizes: $T = 50,\ 100,\ 250$ and 500. We generate observations from an autoregressive process with mean $\mu = 0$, $\varepsilon_t \sim N(0, 1)$ and autocorrelation ρ, where ρ takes values from 0.2 to 0.99. Both μ and ρ are estimated in-sample. Overall, our robust test has good small-sample-size properties, similar to the Box-Pierce and Ljung-Box versions. However, as ρ increases and approaches one, size distortion occurs for both the Box-Pierce and Llung-Box tests.

Table 2 displays the power results for the same test when the true DGP for y_t is the $AR(2)$ process $y_t = (\rho_0 + \rho_1)y_{t-1} - \rho_0\rho_1 y_{t-2} + \varepsilon_t$. The two inverse roots of the process are ρ_0 and ρ_1. Here, we take $\rho_0 = 0.5$, as in West and McCracken (1998), and the results are qualitatively equivalent for other values. We let ρ_1 increase from 0.1 to 0.5. The larger ρ_1 is, the further the process is from an AR(1) process. Therefore, we observe higher rejection rates as ρ_1 increases. Furthermore, our robust moment greatly outperforms the standard Box-Pierce test.[13]

[insert Tables 1 and 2 here]

We now present the out-of-sample properties of the same tests. West and McCracken (1998) show that there is no parameter uncertainty problem for the recursive scheme. We consider here the rolling scheme.[14] We present the results for various values of R (the sample size used for estimating the parameters, from 50 to 500) and P (the sample size used to evaluate our test, from 50 to 500). We compare our robust moment-based tests with the correction proposed in West and McCracken (1998). The results are displayed in Table 3. The size distortion for $R = 50$ is severe in both cases. Our test statistic has much better size properties and also performs better for detecting departure from the null.

---

[12]The power of a chi-squared test decreases as the number of degrees of freedom increases.

[13]We did not correct for the size distortion of the BP test; the comparison is therefore in favor of the BP test.

[14]The fixed scheme results, available in Appendix C of the Supplemental Material, lead to similar conclusions.

4   Application to the backtesting of VaR models

In 1996, the Basel Committee on Banking Supervision proposed the use of VaR models as a possible risk management measure. There is a debate on what characterizes a good risk measure and whether VaR is adequate (for example, see Artzner et al., 1999). However, VaR is the measure commonly used by financial institutions. Let r_t be the daily log return of some given portfolio, and let VaR_t^α be the one-day-ahead VaR forecast (computed at time $t-1$) for a given level of risk α (the value considered is generally 5% or 1%). With an abuse of notation, we consider the VaR measure, VaR_t^α, as the negative of the α-quantile of the conditional distribution of r_t given J_{t-1}, the information set at date $t-1$:[15]

$$P\left(r_t \leq -VaR_t^\alpha | J_{t-1}\right) = \alpha. \tag{10}$$

Backtesting techniques attempt to check the accuracy of the models used by a given institution, in most cases observing only the VaR forecasts, the returns and the distribution assumed for the innovation terms. It is particularly appealing for regulators to measure the adequacy of these risk measures.

Let I_t be the hit, that is, the indicator of VaR violation.[16] Under H_0, i.e., the VaR parametric model used by the financial institution is the right model, I_t is i.i.d. Bernoulli distributed with parameter α.

This section presents some feasible tests that are robust to the parameter uncertainty introduced by the estimation of the conditional variance for the returns. This parameter estimation uncertainty has rarely been considered in the literature.

---

[15]In fact, the VaR measure is the potential loss induced by this negative return.

[16]I_t is defined as

$$I_t = \begin{cases} 1 & \text{if } r_t \leq -VaR_t^\alpha, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Christoffersen (1998) considers a likelihood ratio test in a Markov framework. Christoffersen and Pelletier (2004) and Candelon et al. (2011) consider tests based on the distribution of the duration between two consecutive hits without parameter uncertainty. Escanciano and Olmo (2010) characterize the potential size distortion that could arise from ignoring its impact and use the correction strategy. However, as discussed above, this strategy depends on the underlying model used for the returns. Additionally, they do not consider the parameter uncertainty with respect to the estimation of the number of degrees of freedom in the T-GARCH model. Note also that a moment framework can handle the case where the number of actual hits is equal to zero without any modification, which is particularly interesting when one backtests VaR forecasts with low coverage rate, α.

4.1 Robust moments for backtesting VaR models

In this subsection, we detail the construction of a robust moment. We also show how to build several test statistics from one robust moment.

4.1.1 Building a robust moment for backtesting in practice

Assume that the model for returns is the following constant location-scale model

$$r_t = \mu^0 + \sigma^0 \varepsilon_t, \tag{12}$$

where $\varepsilon_t \sim i.i.d.\ D(0,1)$, a continuous distribution with mean 0 and variance 1. We assume that the parameter for this distribution is known, as for the standard normal variable, or estimated with no parameter uncertainty, as for a standardized Student's distribution where the number of degrees of freedom is constrained to be an integer (this is the case considered in Escanciano and Olmo, 2010).

Assume that we want to backtest the VaR sequence in this model. The moment $m_t = I_t - \alpha$, which compares the VaR violation frequency with the expected one α, is not robust to parameter uncertainty. Following Proposition 2, we detail the steps to transform m_t into a robust moment:

1. Choose an estimating function $g(\cdot)$ for the parameters.

21

2. Calculate $\frac{\partial \mathbb{E}_0[g(r_t,\theta)]}{\partial \theta^\top}$ at $\theta = \theta^0$, which can be done numerically or explicitly.

3. Similarly, calculate $\frac{\partial \mathbb{E}_0[I_t - \alpha]}{\partial \theta^\top}$ at $\theta = \theta^0$.

4. Apply Equation (6) and calculate the test statistic ξ in (1).

We now apply this method to our specific case. First, a simple estimating function for μ^0 and σ^0 is

$$g = \begin{pmatrix} r_t - \mu \\ (r_t - \mu)^2 - \sigma^2 \end{pmatrix}. \tag{13}$$

For the second step, observe that because $g(\cdot)$ is smooth, we need to compute only the expectation of its derivative with respect to parameter θ, $\theta = (\mu, \sigma)^\top$:

$$\frac{\partial g}{\partial \theta^\top}(r_t, \theta) = \begin{bmatrix} -1 & 0 \\ -2(r_t - \mu) & -2\sigma \end{bmatrix}.$$

The expectation at $\theta = \theta^0$ is therefore $V = diag(-1, -2\sigma^0)$.

Third, we compute $P = \left(\frac{\partial \mathbb{E}_0[I_t - \alpha]}{\partial \theta^\top} \right)_{\theta = \theta^0}$,

which is equal to:

$$P = \left[\frac{1}{\sigma^0} f(q_\alpha); \frac{q_\alpha}{\sigma^0} f(q_\alpha) \right],$$

where q_α is the α quantile of the distribution of ε, and $f(\cdot)$ is the probability distribution function.[17]

Finally, applying (6) yields a robust version of $m_t = I_t - \alpha$:

$$m_t^\perp = I_t - \alpha + f(q_\alpha)\varepsilon_t + \frac{q_\alpha f(q_\alpha)}{2} \left(\varepsilon_t^2 - 1 \right), \tag{14}$$

---

[17]Here are the main steps to calculate P; observe that

$$\mathbb{E}_0[I_t - \alpha] = \mathbb{E}_0 \left[\mathbf{1}\{r_t \leq (\mu + \sigma q_\alpha)\} - \alpha \right]$$

$$= \mathbb{E}_0 \left[\mathbf{1}\{ \frac{r_t - \mu^0}{\sigma^0} \leq (\frac{\mu - \mu^0}{\sigma^0} + \frac{\sigma}{\sigma^0} q_\alpha)\} - \alpha \right]$$

$$= F \left(\frac{\mu - \mu^0}{\sigma^0} + \frac{\sigma}{\sigma^0} q_\alpha \right) - \alpha.$$

Consequently, $\frac{\partial \mathbb{E}_0[I_t - \alpha]}{\partial \mu} = \frac{1}{\sigma^0} f \left(\frac{\mu - \mu^0}{\sigma^0} + \frac{\sigma}{\sigma^0} q_\alpha \right)$ and $\frac{\partial \mathbb{E}_0[I_t - \alpha]}{\partial \sigma} = \frac{q_\alpha}{\sigma^0} f \left(\frac{\mu - \mu^0}{\sigma^0} + \frac{\sigma}{\sigma^0} q_\alpha \right)$.

where the variance, V_α, depends on the distribution D assumed for ε_t. Observe that we do not manipulate the score function to build this robust moment.

$m_t^\perp$ is the robust moment in the constant location-scale model built with the estimating function $g(\cdot)$ in (13). Now, we apply the result of Subsection 2.3.3. $m_t^\perp$ is also robust for any GARCH model

$$r_t = \mu(J_{t-1}, \theta) + \sigma(J_{t-1}, \theta)\varepsilon_t, \tag{15}$$

where J_{t-1} is the information set at time $t-1$ and θ is a vector of parameters. Different choices of $g(\cdot)$ generate different robust moments. Additionally, we can also consider the true model (15) to orthogonalize.

4.2 Building test statistics for backtests from a robust moment

Let e_t be a transform of $I_t - \alpha$ that is robust in the GARCH model (15). V_α denotes the variance under the null hypothesis. Let Z_{t-1} be any squared integrable random variable belonging to the information set at date $t-1$. The orthogonalized moment

$$m_t^\perp = Z_{t-1}e_t \tag{16}$$

satisfies $\mathbb{E}_0\left[m_t^\perp\right] = 0$ and is robust to parameter uncertainty.

Thus, the corresponding test statistic

$$\xi_Z = T \left(\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}e_t\right)^\top \left(\mathbb{E}_0\left[Z_{t-1}Z_{t-1}^\top\right] V_\alpha\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} Z_{t-1}e_t\right) \tag{17}$$

is asymptotically distributed as a $\chi^2(k)$, where k is the dimension of Z_{t-1}, whether the parameters of the GARCH model are estimated or known.

In the next Monte Carlo subsection, we study different choices for the above instruments in the past information set. $Z_{t-1} = 1$ corresponds to the unconditional test (i.e., we test that the frequency of hits is the expected one, α); Z_{t-1} can also be past values, e_{t-1}, e_{t-2}, etc. Furthermore, linear combinations of past values of e_t are also possible.

4.3 A Monte Carlo exercise

We now examine the size and power properties of our test procedure and compare them with the correcting strategy. The returns of a fictive portfolio/asset are assumed to follow a GARCH (1,1) model with zero mean and i.i.d. innovations:

$$r_t = \sqrt{\sigma_t^2(\theta)}\varepsilon_t, \ \sigma_t^2(\theta) = \omega + \gamma r_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{18}$$

with $\varepsilon_t \sim D(0,1)$, $\omega = 0.0001$, $\gamma = 0.045$ and $\beta = 0.95$. We successively consider the standard normal distribution and the standardized Student's distribution for the distribution of ε_t.

We simulate samples with $T = 250$, 500 and 750 observations. For each sample, after estimation of the model by maximum likelihood, we compute the series of one-day-ahead VaR forecasts, VaR_t^α, for α equal to 5%. All the results displayed are based on 10 000 replications, and each table reports the rejection frequencies for a 5% level test.

4.3.1 The Normal GARCH model

We first consider the case of a Gaussian innovation process. In Table 4, we study the out-of-sample properties, evaluating first the size, then the power of different competing tests.[18] One-day-ahead VaR forecasts are computed with a rolling estimator assuming normality for the innovation term. This forecasting scheme is the most appropriate for this financial example. We use $R = 500$ values to estimate the parameters. We evaluate our moments on $P = 100$, $P = 250$ and $P = 500$ observations. As emphasized earlier, robust moment tests do not require additional correction, even when studying out-of-sample performance. The tests are detailed here and are also presented in Section A.1 in the Supplemental Material. e_t is the robust version of $I_t - \alpha$ calculated in Equation (14), and $e_t^\star$ is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the space spanned by the true score function of the Normal GARCH model. Finally, $e_t^c = I_t - \alpha$ is a non-robust moment for which we correct for parameter uncertainty, as shown in Subsection 2.2.2. We also consider covariance tests based on the product of these moments with their past values, that is, $e_t e_{t-h}$ for

---

[18]The in-sample results are given in Appendix C of the Supplemental Material.

$h = 1, \ 2, \ 3$ and similarly for the two other moments.

The size properties are good, despite the slight over-rejection in the covariance tests.

The power properties are studied for two alternatives. In the first one, the distribution for the innovation terms ε_t is a standardized Student distribution with 3 degrees of freedom. When computing the VaR measure, Gaussianity is assumed, wrongly. In the second one, we simulate an EGARCH model[19] with T(4) innovations, estimating the standard normal GARCH(1,1) model to compute the VaR forecasts. Both the distributional assumption and the volatility model are wrong. The power is good for the robust moments given the alternative considered. The most important outcome is the comparison with the power properties of the correcting strategy. Correcting deteriorates the power of the test for all values of P considered and for both alternatives. Furthermore, we recall that the test based on e_t is valid for any GARCH specification. The performances of this moment are very close to the performances of the one build from considering the real score function.

[insert Table 4 here]

4.3.2   The T-GARCH model

The T-GARCH (1,1) model is a popular model in empirical finance because it accurately fits most financial data, especially the tail properties. Our new DGP maintains the same conditional variance model as in the GARCH normal case, but the distribution of ε_t is the standardized Student's distribution with $\nu = 6$ degrees of freedom.

We have now one additional parameter to estimate, ν, and we need to consider the parameter uncertainty generated by this additional estimation. We again consider the same three moments as in the normal case. For e_t, the first moment, the estimating function $g(\cdot)$ used for the orthogonalization is the score in the constant variance auxiliary model. $e_t^{\star}$ is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the space spanned by the true score of the full T-GARCH model. Finally e_t^c denotes, as before, $I_t - \alpha$, which is a non-robust moment. We use the correction strategy to take

---

[19]$\sigma_t^2 = \exp(0.0001 + 0.9 \ln \sigma_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1})$.

into account the parameter uncertainty. The analytical expressions of these moments are given in Subsection A.2 of the Supplemental Material. As before, we present only the out-of-sample properties with a rolling scheme, the natural framework for VaR forecasts.[20] The results are displayed in Table 5 for $P = 100, 250$ and 500. Again, we choose $R = 500$, which corresponds to the value chosen in the empirical application. We first present the size and then the power properties. For the power, we consider the historical simulation scheme and a skewed t-distribution.[21] As in the normal case, the tests based on e_t or e_t^* appear to be the best, although e_t does not exploit the full GARCH structure and is valid for any GARCH specification.

The correction strategy is dominated for the skewed t alternative. For the historical simulation, in the out-of-sample case, there is power from the unconditional moments and e_t^c performs well.

[insert Table 5 here]

5 Empirical Application

We illustrate our methodology in an empirical application related to VaR forecasts. We consider the exchange rate data considered previously in Kim, Shephard, and Chib (1998) and also in Bontemps and Meddahi (2005, 2012). These data comprise observations of weekday close exchange rates[22] from 10/01/1981 to 06/28/1985. Bontemps and Meddahi (2005) strongly reject the normality assumption for a GARCH(1,1), whereas Bontemps and Meddahi (2012) do not reject the

---

[20]The in-sample properties are given in Appendix C of the Supplemental Material.

[21]The skewed t-distribution with ν degrees of freedom and parameter γ has the following density g, where f_ν is the density of the standard Student's distribution:

$$g(x) = \frac{2}{\gamma + 1/\gamma} f(\gamma x) \text{ when } x < 0, \ g(x) = \frac{2}{\gamma + 1/\gamma} f(x/\gamma) \text{ when } x \geq 0.$$

In the Monte Carlo exercise, we take $\gamma = 0.5$. Additional simulations not provided here show similar patterns, with increasing power as we depart from $\gamma = 1$.

[22]The U.K. Pound, French Franc, Swiss Franc, and Japanese Yen rates, all versus the U.S. Dollar.

T-GARCH(1,1) model for all but the SF-US$ series.

The T-GARCH (1,1) model is estimated by maximum likelihood, and the parameter estimates are used to compute the one-day-ahead VaR forecast for any value of α, the risk exposure. The in-sample estimates are shown in the first part of Table 6 with degrees of freedom varying from 6.73 to 12.25.

We first test the accuracy of the in-sample VaR forecasts for the four series for three risk levels, $\alpha = 0.5\%$, $\alpha = 1\%$ and $\alpha = 5\%$, using the moments from Section 4.3.2. We also include the non-robust tests based on the number of VaR violations, $I_t - \alpha$, ignoring (wrongly) the parameter uncertainty issue. The p-values of the tests are presented in Table 6.

Note that, for each exchange rate, there is always one risk level α for which our backtesting procedure is rejected. The number of degrees of freedom of the Student's innovations captures the behavior of the left tail, which is why the T-GARCH model is popular. Globally, two series pass the unconditional tests (FF/US and Yen/US). For $\alpha = 0.5\%$, no unconditional test is rejected. The unconditional tests are rejected for the SF/US series with $\alpha = 1\%$ and for the UK/US series with $\alpha = 5\%$. For the last series, although the T-GARCH assumption is not rejected globally, the Student's assumption captures the tail behavior for low-risk values but fails to measure the risk for higher values. The percentages of VaR violations in the following table show that there are too many VaR violations (6.6% instead of 5%) for this exchange rate and this risk exposure.

	UK/US$	FF/US$	SF/US$	Yen/US$
α				
0.005	0.004	0.003	0.005	0.006
0.010	0.013	0.010	0.018	0.011
0.050	0.066	0.044	0.055	0.048

Empirical frequencies of VaR violations

The covariance tests are often rejected except for at low values of α (on the other hand, for $\alpha = 0.05\%$, we expect approximately one VaR violation per year, which reduces the power of the covariance tests). For the SF/US series, many rejections occur for $\alpha = 1\%$; we know that, globally, the T-GARCH assumption is rejected. The same is true for the Yen/US exchange rate, where the

covariance tests are systematically rejected. Therefore, the conditional variance model should be adapted.

In Table 7, we perform the same exercise, out-of-sample, using a T-GARCH(1,1) model-based rolling estimator on the last 445 observations. With 945 observations, we test our model using 500 out-of-sample one-day-ahead VaR forecasts. Note that this is how VaR forecasts are often calculated in practice. Unsurprisingly, the out-of-sample behavior of the tests is different than the in-sample behavior.

First, globally, tests based on e_t^c are more conservative than tests based on e_t and e_t^*. For the unconditional tests, we have similar behavior as before. The unconditional tests are rejected only for the SF/US exchange rate with $\alpha = 1\%$, and the results are close to rejection for the UK/US exchange rate with $\alpha = 5\%$, as for the in-sample results. Note that the difference between I_t and e_t^c is small; simply counting (i.e., ignoring parameter uncertainty or using I_t without any correction) generally decreases the power of the test.

The covariance tests lead to slightly different conclusions. Overall, one should definitively not draw any conclusion from the in-sample properties, especially when the degree of persistence is very high (the estimates of β for the conditional variance are all greater than 0.9). The covariance tests for the Yen/US series are rejected for all values of α. The VaR forecasts for the FF/US exchange rate appear to be accurate, and no test is rejected. By contrast, there is a problem with the local dynamics of the UK/US exchange rate as many covariance tests are rejected for all risk levels α.

[insert Tables 6 and 7 here]

In appendix D of the Supplemental Material, we explore additional series by considering three other exchange rates, the US dollar versus the Yen, the British Pound and the Euro, for the period 2010-2015. We also consider three stock indexes[23] for the same period. Stock indexes have different behavior: the estimates of β are lower than those for exchange rates and the tails are thicker (lower estimates of ν). However, we obtain similar conclusions; that is, the T-GARCH

---

[23]S&P 500, NIKKEI and NASDAQ

model globally fails to accurately forecast VaR measures for different values of α. Additionally, Berkowitz et al. (2011) also rejects the accuracy of VaR forecasts based on T-GARCH models for data from three of four business lines within a large commercial bank.

6 Conclusion

This paper considers moment-based tests for testing features of univariate distributions. The moments of interest can be smooth or non-smooth. In our setup, some parameters must be estimated, which generally modifies the asymptotic distribution of the underlying test. We address the parameter uncertainty problem by projecting the moment onto the orthogonal of the space spanned by the score. Therefore, we consider a class of oblique projections from which we pick any transform. Interestingly, we provide an alternative expression that does not depend on the score function. Our framework is therefore semiparametric because we do not need to specify the full structure of the model.

This paper shows that a robust moment can be built simply by applying a linear correction in which the coefficients can be estimated in-sample. Moreover, robust moments have attractive features and lead to testing procedures that are as powerful as existing ones, even better in many of the examples considered in this paper. For example, the testing procedure does not have to be changed when the estimator of the parameters changes; additionally, a moment-based test based on a robust moment is valid in some cases where the parameters are estimated with slower rates of convergence than the standard square-root rate. Finally, our method can handle out-of-sample evaluations without further correction.

We apply our method to different examples, namely, out-of-sample evaluations and backtesting different GARCH models. When proposing new test procedures, it is particularly important to first check that the small-sample-size properties are good and that the power properties are at least competitive with those of the existing alternative procedures. Our Monte Carlo experiments suggest that our tests behave well for both in-sample and out-of-sample cases, even better than the existing ones in most of the cases considered.

Applied econometrics requires distributional assumptions to compute forecasts or to derive tractable results in structural models. However, these assumptions should be tested whenever possible because they can lead to biased results in the case of misspecification. Moment-based procedures are standard. They have been widely used for estimation, and they can similarly be systematically used to test these assumptions. Parameter uncertainty, which is often ignored in empirical applications, can be easily addressed with the methodology derived in this paper.

REFERENCES

Amengual, D. and E. Sentana (2011), "Inference in Multivariate Dynamic Models with Elliptical Innovations", mimeo CEMFI.

Amengual, D., Fiorentini G. and E. Sentana (2013), "Sequential estimation of shape parameters in multivariate dynamic models", *Journal of Econometrics*, 177, 233-249.

Andrews, D. (1994), "Empirical Process Methods in Econometrics", Handbook of Econometrics, vol 4, Elsevier.

Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999), "Coherent Measures of Risk", *Mathematical Finance*, 9, 203-228.

Berkowitz, J., Peter F. Christoffersen and Denis Pelletier (2011), "Evaluating Value-at-Risk Models with Desk-Level Data", *Management Science*.

Butler J. S. and P. Chatterjee (1997), "Tests of the Specification of Univariate and Bivariate Ordered Probit", *The Review of Economics and Statistics*, 79, 343-347.

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

Bontemps, C. and N. Meddahi (2012), "Testing Distributional assumptions: A GMM Approach", *Journal of Applied Econometrics*, 27, 978-1012.

Bontemps, C., Dufour J.-M. and N. Meddahi (2017), "Optimal moment-based tests", mimeo Toulouse School of Economics.

Box, G. and D. Pierce, (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, 65, 15091526.

Candelon B., Colletaz G., Hurlin C. et Tokpavi S. (2011), "Backtesting Value-at-Risk: a GMM duration-based test", *Journal of Financial Econometrics*, 1-30.

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

Christoffersen, P. F. and D. Pelletier (2004), "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics*, 2, 84-108.

Diebold, F.X. and R.S. Mariano, (1995), "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, 13, 253-263.

Dufour, J.-M., L. Khalaf and M.-C. Beaulieu (2003), "Exact Skewness-Kurtosis Tests for Multivariate Normality and Goodness-of-Fit in Multivariate Regressions with Application to Asset Pricing Models", Oxford Bulletin of Economics and Statistics, 65(s1), 891-906.

Escanciano, J.C. and J. Olmo (2010), "Estimation Risk Effects on Backtesting For Parametric Value-at-Risk Models", Journal of Business and Economic Statistics, 28, 36-51.

Fiorentini, G., E. Sentana and G. Calzolari (2004), "On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models", Economics Letters, 83 307-312.

Francq, C. and J.M. Zakoian, (2010), GARCH Models, Wiley.

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

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

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

Jolivet, G., F. Postel-Vinay and J.M. Robin, (2006), "The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US", *European Economic Review*, 50, 877-907.

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

Ljung, G. M. and G. E. P. Box, (1978), "On a Measure of a Lack of Fit in Time Series Models", *Biometrika*, 65, 297303.

McCracken, M. W. (2000), "Robust out-of-sample inference", *Journal of Econometrics*, 99, 195-223.

Mencia J. and E. Sentana (2012), "Distributional tests in multivariate dynamic models with normal and Student t innovations", *The Review of Economics and Statistics*, 94, 133-152.

Mora, J. and A. Moro-Egido (2008), "On specification testing of ordered discrete choice models", Journal of Econometrics, 143, 191-205.

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

Newey, W.K. and D. McFadden, (1994) "Large sample estimation and hypothesis testing", Handbook of Econometrics, vol 4, Elsevier.

Skeels, C. L. and F. Vella (1999), "A Monte Carlo investigation of the sampling behavior of conditional moment tests in Tobit and Probit models", Journal of Econometrics,, 92, 275-294.

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

West, K. D. (1996), "Asymptotic Inference about Predictive Ability", *Econometrica*, 64, 1067-1084.

West, K. D. and M. W. McCracken, (1998), "Regression-Based Tests of Predictive Ability", *International Economic Review*, 39, 817-840.

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

Tables

ρ_0	Robust moment				Box-Pierce				Llung-Box			
T	50	100	250	500	50	100	250	500	50	100	250	500
0.2	4.81	4.73	4.73	4.64	5.98	5.46	5.14	4.82	7.08	5.94	5.30	4.88
0.5	4.98	4.94	4.78	4.62	5.98	5.41	5.29	4.98	7.13	5.97	5.49	5.13
0.7	4.86	4.80	4.81	4.37	7.02	6.23	6.08	5.53	8.23	6.72	6.26	5.62
0.95	4.90	4.82	5.21	4.52	12.09	11.70	11.92	11.29	13.97	12.80	12.27	11.44
0.97	5.54	5.16	5.24	4.52	12.88	12.63	12.85	12.19	14.89	13.72	13.20	12.30
0.99	4.73	5.03	5.07	4.52	13.80	13.59	13.85	13.14	15.81	14.72	14.05	13.28

Note: the DGP is a univariate AR(1) of size T. The inverse root of the process is ρ_0. We estimate an AR(1) process with mean and test the first order autocorrelation in-sample. We report the rejection frequencies from 10000 simulations. We compare our robust moment $m_t^\perp$ defined in Equation (9) with the Box-Pierce and Llung-Box tests for two autocorrelations.

Table 1: Size of the first order autocorrelation test in-sample

ρ_1	Robust moment				Box-Pierce				Llung-Box			
T	50	100	250	500	50	100	250	500	50	100	250	500
0.1	7.18	8.65	13.26	21.21	7.29	8.11	11.14	16.94	8.73	8.82	11.58	17.13
0.2	12.16	18.54	37.90	62.74	10.28	14.44	28.55	51.17	12.10	15.49	29.10	51.44
0.3	20.19	34.35	68.28	92.72	15.17	25.21	56.55	86.26	17.17	26.55	57.15	86.41
0.5	44.81	72.52	98.05	99.99	34.57	63.03	96.03	100.00	37.99	64.61	96.14	100.00

Note: the DGP is a univariate AR(2) of size T. The two inverse roots of the process are $\rho_0 = 0.5$ and ρ_1. We estimate an AR(1) process with mean and test the first order autocorrelation of the innovation process in-sample. We report the rejection frequencies from 10000 simulations. We compare our robust moment $m_t^{\perp}$ with the Box-Pierce and Llung-Box tests for two autocorrelations.

Table 2: Power of the first order autocorrelation test in-sample

A. Size, $\rho_1 = 0$	R			P		
		50	100	150	250	500
Robust moment $m_t^\perp$	50	7.14	7.87	8.59	8.64	10.11
	100	5.45	5.54	5.62	5.57	5.89
	150	5.61	5.21	5.23	5.44	5.70
	250	5.34	5.55	5.05	5.72	5.57
	500	5.36	5.18	5.24	4.96	5.48
Correction	50	6.77	8.77	11.50	19.15	41.15
	100	4.03	4.02	4.07	5.26	9.67
	150	3.14	2.75	2.77	3.06	4.60
	250	2.52	2.39	2.29	2.53	2.85
	500	2.39	2.05	2.12	1.85	2.06
B. Power, $\rho_1 = 0.2$	R			P		
		50	100	150	250	500
Robust moment $m_t^\perp$	50	12.51	19.04	25.53	37.33	61.48
	100	12.18	17.69	24.07	36.37	62.21
	150	11.22	17.62	23.60	35.23	61.49
	250	11.30	16.78	23.25	35.82	61.16
	500	11.10	17.37	23.29	35.36	61.08
Correction	50	11.88	23.00	35.29	60.34	93.67
	100	7.55	12.71	19.15	34.53	70.33
	150	6.30	10.58	15.49	28.01	58.85
	250	5.70	9.08	12.88	22.77	49.25
	500	5.17	8.31	11.70	19.69	43.03
C. Power, $\rho_1 = 0.5$	R			P		
		50	100	150	250	500
Robust moment $m_t^\perp$	50	42.87	71.03	86.95	97.57	99.98
	100	41.48	70.17	86.47	97.58	99.99
	150	40.87	68.43	84.98	97.39	99.98
	250	39.86	66.97	84.47	97.29	99.97
	500	40.05	67.55	83.25	96.73	99.95
Correction	50	28.21	61.28	84.61	98.75	100.00
	100	21.88	48.96	71.96	94.24	99.98
	150	19.95	44.78	67.38	92.04	99.86
	250	17.79	40.65	63.61	89.91	99.87
	500	16.76	40.11	61.61	87.82	99.66

Note: the DGP is a univariate AR(2). The two inverse roots of the process are $\rho_0 = 0.5$ and ρ_1. We estimate an $AR(1)$ process with mean and test the first order autocorrelation out-of-sample for various values of R and P. We report the rejection frequencies from 10000 simulations. $\rho_1 = 0$ corresponds to the size. We compare our robust moment $m_t^\perp$ with the out-of-sample correction of West and McCracken (1998).

| | Size | | | Power | | | | | |
| | | | | T-GARCH | | | EGARCH | | |
	$P = 100$	$P = 250$	$P = 500$	$P = 100$	$P = 250$	$P = 500$	$P = 100$	$P = 250$	$P = 500$
e_t	5.44	5.44	5.36	22.60	36.09	51.68	26.68	43.61	62.69
e_t^*	5.45	5.48	5.35	22.85	36.57	52.22	26.80	43.47	62.62
e_t^c	5.23	6.31	6.91	4.71	11.04	20.17	11.89	13.14	13.67
$e_t e_{t-1}$	6.82	6.58	6.13	12.44	18.45	25.44	25.34	38.59	51.98
$e_t e_{t-2}$	6.89	6.37	6.14	12.67	18.98	24.94	21.26	32.30	43.25
$e_t e_{t-3}$	6.63	6.31	5.83	13.50	19.45	26.72	19.83	29.01	38.04
$e_t^* e_{t-1}^*$	6.78	6.57	6.02	12.11	18.19	25.19	25.19	38.34	51.58
$e_t^* e_{t-2}^*$	6.83	6.25	6.12	12.56	18.81	25.19	21.20	32.19	42.99
$e_t^* e_{t-3}^*$	6.78	6.19	5.88	13.39	19.33	26.31	19.78	29.10	38.09
$e_t^c e_{t-1}^c$	4.93	5.62	5.00	2.78	2.78	2.20	17.38	26.73	38.50
$e_t^c e_{t-2}^c$	4.40	5.62	4.93	3.10	3.22	2.66	12.59	18.66	26.57
$e_t^c e_{t-3}^c$	4.33	5.05	4.67	2.83	3.34	2.80	9.55	14.42	17.74

Note: for each sample size T, we report the rejection frequencies for a 5% significance level test of the accuracy of the one-day-ahead VaR forecasts computed out-of-sample from a rolling estimation of a GARCH normal model. e_t is the robust version of $I_t - \alpha$ calculated in Equation (14), $e_t^\star$ is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the space spanned by the true score function of the Normal GARCH model, $e_t^c = I_t - \alpha$ is a non-robust moment for which we correct for the parameter uncertainty, as shown in Subsection 2.2.2. See Subsection 4.3.1 for further details.

Table 4: Backtesting VaR measures, $\alpha = 5\%$ - Normal GARCH model - Out of sample properties, rolling scheme, $R = 500$ observations.

	Size			Power					
				HS			Skewed-T		
	$P=100$	$P=250$	$P=500$	$P=100$	$P=250$	$P=500$	$P=100$	$P=250$	$P=500$
e_t	4.86	4.98	5.46	26.92	38.30	38.25	19.99	41.82	70.48
e_t^*	4.94	5.13	5.56	27.01	38.05	38.11	19.89	41.70	69.89
e_t^c	4.98	6.00	6.46	26.19	40.61	41.52	18.73	35.14	61.04
$e_t e_{t-1}$	6.49	5.79	5.47	13.55	16.65	22.56	7.04	7.10	7.52
$e_t e_{t-2}$	6.64	6.01	6.08	13.54	17.05	22.36	6.98	7.16	7.67
$e_t e_{t-3}$	6.62	6.28	6.27	13.17	17.02	22.41	6.60	6.70	7.88
$e_t^* e_{t-1}^*$	6.63	5.86	5.80	13.71	16.67	22.91	6.96	7.12	7.56
$e_t^* e_{t-2}^*$	6.70	5.99	6.22	13.61	17.21	22.42	7.02	7.28	7.76
$e_t^* e_{t-3}^*$	6.71	6.47	6.37	13.05	17.08	22.40	6.48	6.53	7.65
$e_t^c e_{t-1}^c$	4.78	4.98	4.31	12.37	15.91	21.47	10.68	13.74	15.59
$e_t^c e_{t-2}^c$	4.78	5.23	4.89	12.12	15.44	21.56	10.36	13.61	15.88
$e_t^c e_{t-3}^c$	4.63	5.09	5.19	11.57	16.20	21.66	10.56	14.12	16.10

Note: for each sample size T, we report the rejection frequencies for a 5% significance level test of the accuracy of the one-day-ahead VaR forecasts computed from the estimation of a T-GARCH(1,1) model. We study respectively the size, the power with the Historical Simulation, the power with the skewed-T distribution with $\gamma = 0.5$. e_t is the robust version of $I_t - \alpha$ calculated from the projection in the auxiliary model with constant variance, $e_t^\star$ is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the space spanned by the true score function of the T-GARCH model, $e_t^c = I_t - \alpha$ is a non-robust moment for which we correct for the parameter uncertainty, as shown in Subsection 2.2.2. See Subsection 4.3.2 for further details.

Table 5: Backtesting VaR measures, $\alpha = 5\%$ - T-GARCH model - Out of sample properties, rolling scheme, $R = 500$.

	UK/US$	FF/US$	SF/US$	Yen/US$
$\hat{\omega}$	7.20e-07	1.51e-06	9.20e-08	2.95e-06
$\hat{\gamma}$	0.07	0.08	0.04	0.09
$\hat{\beta}$	0.92	0.89	0.96	0.86
$\hat{\nu}$	8.81	12.25	6.73	7.42

Note: MLE of the T-GARCH(1,1) model for daily exchange rates.

$r_t = \sqrt{\sigma_t^2(\theta)}\varepsilon_t$, $\sigma_t^2(\theta) = \omega + \gamma r_{t-1}^2 + \beta\sigma_{t-1}^2$, $\varepsilon_t \sim T(\nu)$.

	$\alpha = 0.5\%$				$\alpha = 1\%$				$\alpha = 5\%$			
	UK/US$	FF/US$	SF/US$	Yen/US$	UK/US$	FF/US$	SF/US$	Yen/US$	UK/US$	FF/US$	SF/US$	Yen/US$
e_t	0.644	0.284	0.972	0.563	0.357	0.784	0.003	0.897	0.011	0.291	0.277	0.502
e_t^*	0.709	0.397	0.918	0.546	0.392	0.861	0.011	0.856	0.013	0.347	0.286	0.558
e_t^c	0.728	0.404	0.896	0.541	0.375	0.874	0.010	0.848	0.011	0.360	0.261	0.570
I_t	0.728	0.404	0.896	0.541	0.375	0.874	0.010	0.848	0.011	0.360	0.261	0.570
$e_t e_{t-1}$	0.837	0.839	0.427	<0.001	0.942	0.004	0.636	<0.001	0.097	0.824	0.544	0.058
$e_t e_{t-2}$	0.309	0.700	0.986	0.340	<0.001	0.911	0.765	0.296	0.823	0.569	0.468	0.070
$e_t e_{t-3}$	0.926	0.988	0.312	0.814	0.755	0.802	<0.001	0.728	0.465	0.283	0.417	0.057
$e_t^* e_{t-1}^*$	0.822	0.374	0.557	0.061	0.635	0.020	0.193	0.047	0.082	0.827	0.975	0.038
$e_t^* e_{t-2}^*$	0.580	0.704	0.906	0.327	0.108	0.726	0.883	0.194	0.813	0.643	0.171	0.047
$e_t^* e_{t-3}^*$	0.723	0.871	0.927	0.546	0.896	0.847	<0.001	0.528	0.331	0.546	0.145	0.010
$e_t^c e_{t-1}^c$	0.906	0.966	0.820	0.811	0.631	0.002	0.356	0.753	0.029	0.884	0.173	0.494
$e_t^c e_{t-2}^c$	0.912	0.966	0.857	0.801	0.631	0.827	0.388	0.742	0.913	0.859	0.172	0.005
$e_t^c e_{t-3}^c$	0.914	0.966	0.857	0.811	0.620	0.828	<0.001	0.752	0.139	0.385	0.169	0.182
$I_t I_{t-1}$	0.908	0.963	0.855	0.797	0.587	<0.001	0.377	0.720	0.004	0.845	0.095	0.352
$I_t I_{t-2}$	0.907	0.963	0.855	0.796	0.587	0.802	0.377	0.720	0.886	0.810	0.090	<0.001
$I_t I_{t-3}$	0.907	0.963	0.855	0.796	0.586	0.802	<0.001	0.719	0.049	0.241	0.088	0.067

Note: we test the accuracy of the one-day-ahead VaR forecasts computed from a T-GARCH(1,1) model for different level of risk α for the four daily exchanges rates. The p-values of the test statistics are reported. The notations are defined in Section 4.3.

Table 6: Backtesting of VaR forecasts for the T-GARCH(1,1) model, In sample evaluation.

	$\alpha = 0.5\%$				$\alpha = 1\%$				$\alpha = 5\%$			
	UK/US$	FF/US$	SF/US$	Yen/US$	UK/US$	FF/US$	SF/US$	Yen/US$	UK/US$	FF/US$	SF/US$	Yen/US$
e_t	0.474	0.443	0.218	0.416	0.948	0.919	0.137	0.990	0.078	0.432	0.866	0.927
e_t^*	0.083	0.320	0.673	0.479	0.131	0.899	0.014	0.280	0.051	0.444	0.732	0.572
e_t^c	0.107	0.105	0.745	0.334	0.168	0.351	0.022	0.167	0.083	0.663	0.829	0.996
I_t	0.113	0.113	0.751	0.342	0.178	0.369	0.025	0.178	0.101	0.682	0.837	1.000
$e_t e_{t-1}$	0.482	0.138	0.385	< 0.001	0.489	0.489	0.142	< 0.001	0.841	0.136	0.480	0.057
$e_t e_{t-2}$	< 0.001	0.540	0.886	0.464	< 0.001	0.925	0.944	0.601	0.724	0.840	0.598	0.078
$e_t e_{t-3}$	0.287	0.969	0.405	0.714	0.527	0.686	0.628	0.495	0.203	0.388	0.373	0.640
$e_t^* e_{t-1}^*$	0.378	0.282	0.601	< 0.001	0.428	0.473	0.402	< 0.001	0.542	0.102	0.316	< 0.001
$e_t^* e_{t-2}^*$	0.159	0.830	0.878	0.295	0.094	0.788	0.809	0.032	0.926	0.571	0.257	0.024
$e_t^* e_{t-3}^*$	0.958	0.704	0.794	0.402	0.945	0.589	0.521	0.368	0.035	0.348	0.229	< 0.001
$e_t^c e_{t-1}^c$	0.716	0.716	0.866	0.805	0.615	0.681	0.416	0.615	0.960	0.669	0.120	0.234
$e_t^c e_{t-2}^c$	0.725	0.735	0.873	0.805	0.610	0.682	0.487	0.557	0.957	0.665	0.738	< 0.001
$e_t^c e_{t-3}^c$	0.727	0.734	0.870	0.799	0.615	0.681	0.432	0.616	0.005	0.655	0.734	0.485
$I_t I_{t-1}$	0.736	0.736	0.875	0.804	0.619	0.684	0.497	0.619	0.961	0.670	0.120	0.238
$I_t I_{t-2}$	0.735	0.735	0.874	0.804	0.618	0.683	0.497	0.618	0.959	0.668	0.738	< 0.001
$I_t I_{t-3}$	0.735	0.735	0.874	0.804	0.617	0.682	0.496	0.617	0.005	0.666	0.736	0.483

Note: we test the accuracy of the one-day-ahead VaR forecast computed from a T-GARCH(1,1) model for different level of risk α for the four daily exchanges rates. The p-value of the test statistics are reported. The notations are defined in Section 4.3.

Table 7: Backtesting of VaR forecasts, Out-of-sample evaluation.

SUPPLEMENT TO "MOMENT-BASED TESTS UNDER PARAMETER UNCERTAINTY"

by Christian BONTEMPS

A   Moments used in the Monte Carlo exercise for the backtests

We detail here the moments used for backtesting in the Monte Carlo experiments in Subsection 4.3. We consider a GARCH(1,1) model without mean

$$r_t = \sqrt{\sigma_t^2(\theta)}\varepsilon_t, \ \sigma_t^2(\theta) = \omega + \gamma r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where the innovations ε_t follow respectively a standard normal distribution and a standardized student distribution with parameter ν. Hereafter, θ denotes the parameter vector related to the variance equation: $\theta = (\omega, \gamma, \beta)^\top$.

A.1   Moments used for the GARCH Normal case

All the moments used are based on a robustification of the centered hit value, $e_t^c = I_t - \alpha$, which is not robust to parameter uncertainty.

- e_t is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the score in the auxiliary model with constant variance. Following Subsection 2.3.3, it is also robust in the GARCH model. Bringing back Equation (14), we have:

$$e_t = I_t - \alpha + \frac{n_\alpha \varphi(n_\alpha)}{2} \left(\varepsilon_t^2 - 1\right),$$

where n_α is the α quantile of the standard normal distribution and $\varphi(\cdot)$ its p.d.f. The variance of e_t is equal to

$$\mathbb{V}_0[e_t] = \alpha(1 - \alpha) - \frac{(n_\alpha \varphi(n_\alpha))^2}{2}.$$

- e_t^* is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the score in the **full GARCH Normal** model:

$$e_t^* = I_t - \alpha + \frac{n_\alpha \varphi(n_\alpha)}{2} \mathbb{E}_0[dl_t] \, \mathbb{E}_0\left[dl_t dl_t^\top\right]^{-1} dl_t \left(\varepsilon_t^2 - 1\right),$$

where dl_t denotes $\frac{\partial \ln \sigma_t(\theta)}{\partial \theta^\top}$. Its variance is equal to:

$$\mathbb{V}_0[e_t^*] = \alpha(1 - \alpha) - \frac{(n_\alpha \varphi(n_\alpha))^2}{2} \mathbb{E}_0[dl_t] \, \mathbb{E}_0\left[dl_t dl_t^\top\right]^{-1} \mathbb{E}_0\left[dl_t^\top\right].$$

The expectation above can be estimated in-sample from standard formulas (see, for example, Franck and Zakoian, 2010).

- $e_t^c = I_t - \alpha$ is not robust. We need to correct its variance.

- We also use joint moments $e_t e_{t-h}$ and similarly for the two other moments, that test for uncorrelation with the past values.

A.2 Moments used for the T-GARCH case

Following the expression of the score function for the constant variance model:

$$s_{\sigma^2,\nu}(\varepsilon_t) = \begin{bmatrix} -\frac{1}{2\sigma^2}\left(1 + \varepsilon_t \frac{\partial \log f_\nu}{\partial \varepsilon_t}(\varepsilon_t)\right) \\ \frac{\partial \log f_\nu}{\partial \nu}(\varepsilon_t) \end{bmatrix}, \tag{A.1}$$

where $f_\nu(\cdot)$ denotes the p.d.f. of the standardized Student distribution with ν degrees of freedom, we can derive the expression of the moments used in the Monte Carlo Section for the T-GARCH model.

- e_t is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the score in the auxiliary model with constant variance, i.e. $s_{\sigma^2,\nu}(\cdot)$ expressed above. It is robust in the T-GARCH model. Adapting Equation (14) with $\tilde{s}_\theta(\cdot)$ as the new estimating function for the parameters, we have:

$$e_t = I_t - \alpha - PV_s^{-1}s_{\sigma^2,\nu}(\varepsilon_t),$$

where $s_{\sigma^2,\nu}(\cdot)$ the score function in the auxiliary model, V_s its variance and P the covariance between I_t and the score $s_{\sigma^2,\nu}(\cdot)$. Its variance is equal to

$$\mathbb{V}_0[e_t] = \alpha(1 - \alpha) - PV_s^{-1}P^\top.$$

Both P and V_s are computed in-sample.

- e_t^* like before is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the score in the **full T-GARCH** model:

$$e_t^* = I_t - \alpha - P^*(V_s^*)^{-1}s_{\theta,\nu}^*(\varepsilon_t),$$

where $s_{\theta,\nu}^*(\cdot)$ is the full score function, P^* is the covariance between I_t and this score and V_s^* its variance. These last two quantities are estimated in-sample. The variance of e_t^* is equal to

$$\mathbb{V}_0[e_t^*] = \alpha(1 - \alpha) - P^*(V_s^*)^{-1}(P^*)^\top.$$

Note that e_t and e_t^* have the same expression but the first considers the score of the constant scale model whereas the second considers the score of the true GARCH model.

- Finally e_t^c is not robust and we need to correct for the parameter uncertainty.

- Like in the normal case, we also consider joint tests.

B Proof of the propositions

B.1 Proof of Proposition 3

We prove here the result in the smooth case where $m(\cdot)$ is assumed to be continuously differentiable in a neighborhood of the true value θ^0. The non-smooth case is build similarly from Tauchen (1985)'s arguments. We assume here that

(i) $m(\cdot)$ is twice continuously differentiable in a neighborhood $\mathcal{N}_0$ of θ^0,

(ii) its second derivative with respect to θ is uniformly bounded on this neighborhood, i.e. $\mathbb{E}_0\left[|\partial^2 m/\partial\theta_j\partial\theta_i|\right] \leq M$ for all i,j in $[1,...,r]^2$, where r is the dimension of θ,

(iii) each partial derivative of m satisfies Assumption CLT.

For convenience we write the proof assuming that the moment and the parameter θ are unidimensional without loss of generality. Let m be a robust moment.

For any $\hat{\theta}$, and for any $t \in [1;T]$, there is $\overline{\theta}_t$ between $\hat{\theta}$ and θ^0 such that

$$m(y_t,\hat{\theta}) = m(y_t,\theta^0) + \frac{\partial m}{\partial\theta}(y_t,\theta^0)(\hat{\theta}-\theta^0) + \frac{1}{2}\frac{\partial^2 m}{\partial\theta^2}(y_t,\overline{\theta}_t)(\hat{\theta}-\theta^0)^2. \tag{B.2}$$

Averaging Equation (B.2) from $t=1$ to T and multiplying by $\sqrt{T}$, we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}m(y_t,\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}m(y_t,\theta^0) + \sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial m}{\partial\theta}(y_t,\theta^0)\right](\hat{\theta}-\theta^0)$$
$$+ \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{2}\frac{\partial^2 m}{\partial\theta^2}(y_t,\overline{\theta}_t)\right]\sqrt{T}(\hat{\theta}-\theta^0)^2.$$

The second term on the right-hand side which multiplies $(\hat{\theta}-\theta^0)$ is equal to $Z + O_p(1)$, where Z is a standard normal variable, because of assumption (iii). The third term which multiplies $\sqrt{T}(\hat{\theta}-\theta^0)^2$ has a bounded expectation because of assumption (ii). Therefore we can rewrite the expansion above using $(\hat{\theta}-\theta^0) = O_P\left(\frac{1}{T^\alpha}\right)$. We obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}m(y_t,\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}m(y_t,\theta^0) + O_P\left(\frac{1}{T^\alpha}\right) + O_P\left(\frac{1}{T^{2\alpha-1/2}}\right). \tag{B.3}$$

When $2\alpha - 1/2 > 0$, the conclusion follows.

B.2 Robustness from an auxiliary model (subsection 2.3.3)

The conditional score function with respect to θ, $s_\theta(y_t|X_{t-1})$, in the true model, is equal to

$$s_\theta(y_t|X_{t-1}) = \sum_{j=1}^{l}\frac{\partial h_j}{\partial\theta^\top}(X_{t-1},\theta)s_{\beta_j}(y_t),$$

where l is the dimension of β. The conditional score is therefore a linear combination of the components of the score in the auxiliary model; the weights are functions of X_{t-1}. Consequently,

$$\mathbb{E}_0\left[ms_\theta|X_{t-1}\right] = \sum_{j=1}^{l}\frac{\partial h_j}{\partial\theta^\top}(X_{t-1},\theta)\mathbb{E}_0\left[ms_\beta\right] = 0,$$

which implies $\mathbb{E}_0\left[ms_\theta\right] = 0$.

B.3 Choice of $g(\cdot)$ and local power properties

B.3.1 Correcting for parameter uncertainty or working with a robust moment

We assume that parameter θ is estimated from the estimating function $g(\cdot)$. We also assume that all the functions involved are continuously differentiable w.r.t. θ in a neighborhood of the true value and that the data are i.i.d.

We first derive the asymptotic variance of $\frac{1}{\sqrt{T}} m(y_t, \hat{\theta})$. Equation (4) combined with Assumption REG gives the following expansion

$$\frac{1}{\sqrt{T}} m(y_t, \hat{\theta}) = \frac{1}{\sqrt{T}} m(y_t, \theta^0) + \mathbb{E}_0 \left[\frac{\partial m}{\partial \theta^\top} \right] \frac{1}{\sqrt{T}} w(y_t, \theta^0) + o_P(1)$$

$$= \frac{1}{\sqrt{T}} (m + D_m w)(y_t, \theta^0) + o_P(1),$$

where D_m denotes $\mathbb{E}_0 \left[\frac{\partial m}{\partial \theta^\top} \right]$.

Its asymptotic variance is therefore $\Omega = \mathbb{V}_0(m + D_m w)$. Note that under the standard regularity conditions of the GMM literature (here $g(\cdot)$ can be seen as the set of identifiying restrictions), $w(\cdot, \theta)$, the influence function, is equal to $\mathbb{E}_0 \left[\frac{\partial g}{\partial \theta^\top} \right]^{-1} g(\cdot, \theta)$.

The correcting strategy consists of calculating the test statistic

$$\xi_c = T \left(\frac{1}{T} \sum_{t=1}^T m(y_t, \hat{\theta}) \right)^\top \Omega^{-1} \left(\frac{1}{T} \sum_{t=1}^T m(y_t, \hat{\theta}) \right).$$

Ω is generally estimated in-sample.

• Assume now that we use the robust moment $\tilde{m}_g$ in Equation (6) for the same estimating function $g(\cdot)$. For any t and θ in a neighborhood of θ^0:

$$\tilde{m}_g(y_t, \theta) = m(y_t, \theta) + \mathbb{E}_0 \left[\frac{\partial m}{\partial \theta^\top} \right] \mathbb{E}_0 \left[\frac{\partial g}{\partial \theta^\top} \right]^{-1} g(y_t, \theta).$$

First, the variance of $\tilde{m}_g$, when $\theta = \theta_0$, is Ω introduced above. Then, write the empirical average:

$$\frac{1}{T} \sum_{t=1}^T \tilde{m}_g(y_t, \hat{\theta}) = \frac{1}{T} \sum_{t=1}^T m(y_t, \hat{\theta}) + \mathbb{E}_0 \left[\frac{\partial m}{\partial \theta^\top} \right] \mathbb{E}_0 \left[\frac{\partial g}{\partial \theta^\top} \right]^{-1} \frac{1}{T} \sum_{t=1}^T g(y_t, \hat{\theta}),$$

$$= \frac{1}{T} \sum_{t=1}^T m(y_t, \hat{\theta}).$$

The second term on the right-hand side is indeed equal to zero because $g(\cdot)$ is the estimating function used to estimate θ. Consequently,

$$\frac{1}{T} \sum_{t=1}^T g(y_t, \hat{\theta}) = 0.$$

Observe that this is not true if θ^0 is estimated by another estimating function. Now, using the property, that $\tilde{m}_g$ is a robust moment, the asymptotic variance of $\frac{1}{\sqrt{T}}\tilde{m}_g(y_t, \hat{\theta})$ is the one of $\frac{1}{\sqrt{T}}\tilde{m}_g(y_t, \theta^0)$, i.e. Ω. The test statistic based on $\tilde{m}_g$ is therefore

$$\xi_{\tilde{m}_g} = T \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{m}_g(y_t, \hat{\theta}) \right)^{\top} \Omega^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \tilde{m}_g(y_t, \hat{\theta}) \right),$$

$$= T \left(\frac{1}{T} \sum_{t=1}^{T} m(y_t, \hat{\theta}) \right)^{\top} \Omega^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} m(y_t, \hat{\theta}) \right).$$

It is the same as ξ_c above.

Consequently, there is no loss of power for our strategy because it coincides with the correcting strategy for some specific choice of the estimating function $g(\cdot)$ in Equation (6). Remark that if the two statistics coincide in $\hat{\theta}$, it it not true for any other value, and especially θ^0. This is the advantage of using a robust moment: for any estimator of θ^0 (including the true value itself), one manipulates the same expression.

B.3.2 Proof of Proposition 4

We first calculate the behavior of the test statistic under the local alternative q_1. First,

$$\mathbb{E}_1 [\tilde{m}_g] = \int \tilde{m}_g(y, \theta^0)(q_0(y) + h(y)q_0(y)/\sqrt{T})dy$$

$$= \frac{1}{\sqrt{T}} \int \tilde{m}_g(y, \theta^0)h(y)q_0(y)dy$$

$$= \frac{1}{\sqrt{T}} \mathbb{E}_0 [\tilde{m}_g h].$$

Second,

$$\mathbb{V}_1 [\tilde{m}_g] = \int \tilde{m}_g^2(y, \theta^0)(q_0(y) + h(y)q_0(y)/\sqrt{T})dy - \frac{1}{T} \mathbb{E}_0 [\tilde{m}_g h]^2$$

$$= \mathbb{V}_0 [\tilde{m}_g] + \frac{1}{\sqrt{T}} \mathbb{E}_0 [\tilde{m}_g^2 h] - \frac{1}{T} \mathbb{E}_0 [\tilde{m}_g h]^2.$$

$\xi_{\tilde{m}_g}$, defined in Equation (1), is asymptotically χ^2 distributed, under the null, for any choice of $g(\cdot)$. Moreover, under the local alternative,

$$\xi_{\tilde{m}_g} = T \frac{\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{m}_g(y_t, \theta^0) \right)^2}{\mathbb{V}_0 [\tilde{m}_g]}$$

$$= \left(\sqrt{T} \frac{\frac{1}{T} \sum_{t=1}^{T} \tilde{m}_g(y_t, \theta^0) - \mathbb{E}_1 [\tilde{m}_g]}{\sqrt{\mathbb{V}_0 [\tilde{m}_g]}} + \sqrt{T} \frac{\mathbb{E}_1 [\tilde{m}_g]}{\sqrt{\mathbb{V}_0 [\tilde{m}_g]}} \right)^2,$$

$$= \left(\sqrt{T} \frac{\frac{1}{T} \sum_{t=1}^{T} \tilde{m}_g(y_t, \theta^0) - \mathbb{E}_1 [\tilde{m}_g]}{\sqrt{\mathbb{V}_1 [\tilde{m}_g]}} + \sqrt{T} \frac{\mathbb{E}_1 [\tilde{m}_g]}{\sqrt{\mathbb{V}_1 [\tilde{m}_g]}} + o_P(1) \right)^2,$$

$$= \left(Z_1 + \varepsilon_h \sqrt{a(g)} + o_P(1) \right)^2.$$

In the third line above, we use the fact that the variances of the moment under the null and the local alternative are equal at the first order. Then, we apply the central limit theorem under the alternative, Z_1 is therefore a standard normal variable. Finally we replace, $\sqrt{T}\mathbb{E}_1 [\tilde{m}_g]$ by $\mathbb{E}_0 [\tilde{m}_g h]$ to get the result (ε_h is equal to the sign of $\mathbb{E}_0 [\tilde{m}_g h]$).

Consequently, $\xi_{\tilde{m}_g}$ is asymptotically distributed as a noncentral chi-squared distribution under the alternative. $a(g) = \frac{\mathbb{E}_0 [\tilde{m}_g h]^2}{\mathbb{V}_1 [\tilde{m}_g]}$, the noncentrality parameter, drives the power properties. Maximizing the power is maximizing $a(g)$ and the solution depends on h.

B.4 Proof of Proposition 5

We follow here the proof of Lemma 4.1 in West and McCracken (1998) assuming that the moment $m(\cdot)$ is smooth. Following McCracken (2000) and Tauchen (1985), we can adapt the proof to non-smooth moments under our regularity conditions (and specifically the regular Assumptions 2 and 4 in Tauchen, 1985).

We now detail the proof for the rolling scheme under Assumptions 1 to 5 of West and McCracken (1998). It adapts easily for the other schemes. We also assume for notational simplicity that both m and θ are scalar. Let m be a robust moment and let d_t denote $\frac{\partial m}{\partial \theta}(y_t, \theta^0)$. Following Assumption 2 of West and McCracken (1998), the estimate of θ_0, $\hat{\theta}$, satisfies our regularity assumption (R) in Section 2: $\hat{\theta} - \theta^0 = \frac{1}{T} \sum_{t=1}^{T} w(y_t, \theta^0) + o_P(1)$. In the following, w_t denotes $w(y_t, \theta^0)$.

For any P, R and $\tau \geq 1$, we can write the following expansion:

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{R+P-1} m(y_{t+\tau}, \hat{\theta}_R) = \frac{1}{\sqrt{P}} \sum_{t=R}^{R+P-1} m(y_{t+\tau}, \theta^0) + \frac{1}{\sqrt{P}} \sum_{t=R}^{R+P-1} \frac{\partial m}{\partial \theta}(y_{t+\tau}, \theta^0)(\hat{\theta}_R - \theta^0) + o_P(1).$$

(B.4)

The difference with respect to West and McCracken (1998), Equation (A1), is that, due to the fact that $m(\cdot)$ is robust, there is no other term. We now work with the second term of the right hand

side.

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{R+P-1} \frac{\partial m}{\partial \theta}(y_{t+\tau}, \theta^0)(\hat{\theta}_R - \theta^0) = \frac{1}{\sqrt{P}} \sum_{t=R}^{R+P-1} d_{t+\tau} \left(\frac{1}{R} \sum_{s=t-R+1}^{t} w(z_s, \theta^0) \right) + o_P(1),$$

$$= \frac{1}{\sqrt{PR}} \sum_{t=R}^{R+P-1} d_{t+\tau}(w_{t-R+1} + w_{t-R+2} + \ldots + w_t) + o_P(1).$$

Let us compute the expectation of each term of the right hand side. Let $\gamma_s = E_0(d_t w_{t-s})$. We first assume that $P < R$ (the bound is similar for $R \le P$).

$$\mathbb{E}_0 \left(\sum_{t=R}^{R+P-1} d_{t+\tau}(w_{t-R+1} + w_{t-R+2} + \ldots + w_t) \right) = P(\gamma_\tau + \gamma_{\tau+1} + \ldots + \gamma_{\tau+R-1})$$

$$\le P \sum_{s=-\infty}^{+\infty} \gamma_s.$$

The expectation of the second term of the right hand side of Equation (B.4) tends to zero when $\sqrt{P}/R$ tends to zero. Following the argument in West (1996), page 1080, (the same argument is used in West and McCracken, page 837), the second term tends to zero under the regularity conditions.

C Additional Monte Carlo results

C.1 Out-of-sample properties, fixed scheme, first order autocorrelation test

A. Size, $\rho_1 = 0$	R			P		
		50	100	150	250	500
Robust moment $m_t^{\perp}$	50	12.38	15.54	16.89	20.54	26.21
	100	7.73	8.89	9.39	10.61	13.25
	150	6.63	7.42	7.70	7.70	9.25
	250	6.03	5.71	6.61	6.12	6.60
	500	6.28	5.50	5.51	5.50	5.65
Correction	50	2.61	2.40	2.25	2.01	1.94
	100	2.42	2.37	2.22	1.82	1.90
	150	2.43	2.41	1.81	1.73	1.96
	250	2.24	2.03	2.08	2.05	1.78
	500	2.68	1.81	1.88	1.96	1.77
B. Power, $\rho_1 = 0.2$	R			P		
		50	100	150	250	500
Robust moment $m_t^{\perp}$	50	13.97	18.15	21.05	28.48	43.44
	100	11.75	16.32	21.22	28.89	47.67
	150	11.77	16.24	21.06	30.89	51.13
	250	11.49	17.08	21.29	32.04	55.18
	500	11.46	16.90	21.81	33.64	57.09
Correction	50	5.62	7.82	11.91	19.01	38.53
	100	5.30	8.15	11.41	18.24	38.49
	150	5.19	7.70	10.72	17.70	37.58
	250	5.21	7.56	10.39	18.04	38.10
	500	5.21	7.46	10.27	18.02	37.71
C. Power, $\rho_1 = 0.5$	R			P		
		50	100	150	250	500
Robust moment $m_t^{\perp}$	50	37.69	58.74	72.14	86.69	95.48
	100	38.84	62.61	78.37	92.88	99.31
	150	38.77	63.36	80.10	94.62	99.78
	250	39.57	65.29	81.67	96.03	99.93
	500	40.78	65.42	83.32	96.07	99.93
Correction	50	18.38	40.27	61.03	87.04	99.61
	100	17.33	39.69	60.48	87.04	99.61
	150	16.16	39.11	59.79	86.38	99.76
	250	16.70	38.72	60.62	86.77	99.59
	500	17.61	37.52	59.78	86.30	99.63

Note: the DGP is a univariate AR(2). The two inverse roots of the process are $\rho_0 = 0.5$ and ρ_1. We estimate an AR(1) process with mean and test the first order autocorrelation out-of-sample for various values of R and P. We report the rejection frequencies from 10000 simulations. $\rho_1 = 0$ corresponds to the size. We compare our robust moment $m_t^{\perp}$ with the out-of-sample correction of West and McCracken.

Table VIII: Size and power of first order autocorrelation test - fixed scheme

C.2 In-sample properties, GARCH Normal example

	Size			Power								
				Hist. Simulation			T-GARCH			EGARCH		
T	250	500	750	250	500	750	250	500	750	250	500	750
e_t	4.74	5.00	5.07	0.00	0.00	0.00	16.05	34.54	50.36	33.95	56.64	72.17
e_t^*	4.27	4.93	4.71	0.00	0.00	0.00	14.05	32.65	47.69	33.33	56.45	72.35
e_t^c	3.80	5.95	4.67	0.00	0.00	0.00	11.49	32.51	44.87	11.27	18.16	15.86
$e_t e_{t-1}$	4.71	4.81	4.87	10.81	15.12	21.16	9.76	13.91	17.79	33.38	47.91	56.84
$e_t e_{t-2}$	4.82	4.80	4.89	10.94	14.23	20.70	11.52	15.72	19.57	26.99	37.20	43.93
$e_t e_{t-3}$	5.20	4.85	4.72	9.91	14.30	19.38	11.91	15.74	19.27	23.70	32.34	38.49
$e_t^* e_{t-1}^*$	4.69	4.79	4.80	10.97	15.13	21.39	8.78	12.41	16.41	32.68	47.44	56.64
$e_t^* e_{t-2}^*$	4.78	4.71	5.06	10.80	14.30	20.78	10.57	14.57	18.14	26.67	36.98	43.67
$e_t^* e_{t-3}^*$	5.12	5.03	4.59	9.95	14.40	19.53	11.15	14.75	17.89	23.45	32.14	38.36
$e_t^c e_{t-1}^c$	3.27	3.53	4.14	4.30	10.54	17.88	2.47	2.00	1.96	24.74	37.51	46.34
$e_t^c e_{t-2}^c$	3.92	4.08	4.11	4.78	10.42	18.24	3.41	2.16	2.07	18.31	24.69	30.04
$e_t^c e_{t-3}^c$	3.87	4.03	4.05	3.91	9.95	16.87	3.00	2.19	2.09	13.17	16.48	19.39

Note: for each sample size T, we report the rejection frequencies for a 5% significance level test of the accuracy of the one-day-ahead VaR forecasts computed from the estimation of a GARCH normal model. e_t is the robust version of $I_t - \alpha$ calculated in Equation (14), $e_t^\star$ is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the space spanned by the true score function of the Normal GARCH model, $e_t^c = I_t - \alpha$ is a non-robust moment for which we correct for the parameter uncertainty, as shown in Subsection 2.2.2. See Subsection 4.3.1 for further details.

Table IX: Backtesting VaR measures, $\alpha = 5\%$ - Normal GARCH model - In sample properties.

C.3 In-sample properties, T-GARCH example

T	Size 250	500	750	HS 250	500	750	Skewed t 250	500	750
e_t	4.92	4.82	4.62	0.00	0.00	0.00	41.75	71.57	88.50
e_t^*	5.10	5.01	4.73	0.00	0.00	0.00	44.44	73.55	89.35
e_t^c	4.42	4.47	4.52	0.00	0.00	0.00	40.48	76.04	88.25
$e_t e_{t-1}$	4.67	4.88	4.93	10.31	14.48	21.44	5.93	6.96	7.83
$e_t e_{t-2}$	5.05	5.08	5.02	10.33	14.35	21.28	6.59	7.37	7.98
$e_t e_{t-3}$	5.11	4.96	4.99	10.17	15.00	20.68	7.02	7.12	8.18
$e_t^* e_{t-1}^*$	5.10	5.03	5.02	10.98	14.47	21.75	6.72	7.50	8.37
$e_t^* e_{t-2}^*$	5.49	5.23	5.19	10.75	14.86	21.44	7.85	8.07	8.46
$e_t^* e_{t-3}^*$	5.46	5.22	5.14	10.49	15.37	20.94	7.67	7.81	8.72
$e_t^c e_{t-1}^c$	5.22	4.82	4.56	10.70	11.80	19.04	9.94	16.35	16.82
$e_t^c e_{t-2}^c$	5.93	5.07	5.06	11.02	11.23	18.91	11.54	16.96	17.41
$e_t^c e_{t-3}^c$	6.00	4.76	4.59	10.67	11.57	18.42	12.85	17.25	17.59

Note: for each sample size T, we report the rejection frequencies for a 5% significance level test of the accuracy of the one-day-ahead VaR forecasts computed from the estimation of a T-GARCH(1,1) model. We study respectively the size , the power with the Historical Simulation, the power with the skewed t-distribution with $\gamma = 0.5$. e_t is the robust version of $I_t - \alpha$ calculated from the projection in the auxiliary model with constant variance, e_t^* is the orthogonal projection of $I_t - \alpha$ onto the orthogonal of the space spanned by the true score function of the T-GARCH model, $e_t^c = I_t - \alpha$ is a non-robust moment for which we correct for the parameter uncertainty, as shown in Subsection 2.2.2. See Subsection 4.3.2 for further details.

Table X: Backtesting VaR measures, $\alpha = 5\%$ - T-GARCH model- In sample properties.

D   Additional Backtests evaluations

	Yen/US$	UK/US$	EUR/US$	S&P 500	NIKKEI	NASDAQ
$\hat{\omega}$	5.20e-07	9.31e-08	3.77e-08	3.77e-06	7.53e-06	4.64e-06
$\hat{\gamma}$	0.040	0.034	0.034	0.134	0.106	0.107
$\hat{\beta}$	0.945	0.963	0.966	0.831	0.857	0.855
$\hat{\nu}$	5.018	15.565	8.203	6.407	9.449	7.634

Note: MLE of the T-GARCH(1,1) model for daily exchange rates or stock indexes.
$r_t = \sqrt{\sigma_t^2(\theta)}\varepsilon_t,\ \sigma_t^2(\theta) = \omega + \gamma r_{t-1}^2 + \beta\sigma_{t-1}^2,\ \varepsilon_t \sim T(\nu).$

	$\alpha = 0.5\%$			$\alpha = 1\%$			$\alpha = 5\%$		
	Yen/US	UK/US	EUR/US	Yen/US	UK/US	EUR/US	Yen/US	UK/US	EUR/US
e_t	0.125	0.514	0.654	0.011	0.585	0.421	0.550	0.003	0.011
e_t^*	0.176	0.601	0.595	0.026	0.625	0.564	0.604	0.005	0.032
e_t^c	0.183	0.580	0.586	0.028	0.599	0.574	0.638	0.004	0.034
I_t	0.060	0.496	0.864	0.002	0.562	0.399	0.407	0.006	0.007
$e_t e_{t-1}$	0.983	0.871	0.984	0.341	0.907	0.678	0.243	0.834	0.111
$e_t e_{t-2}$	0.972	0.276	0.941	0.991	0.148	0.953	0.988	0.988	0.852
$e_t e_{t-3}$	0.969	0.481	0.270	0.964	0.872	0.541	0.936	0.925	0.080
$e_t^* e_{t-1}^*$	0.808	0.696	0.959	0.862	0.621	0.778	0.270	0.756	0.255
$e_t^* e_{t-2}^*$	0.929	0.406	0.928	0.890	0.071	0.881	0.821	0.754	0.628
$e_t^* e_{t-3}^*$	0.923	0.698	0.763	0.859	0.384	0.849	0.890	0.695	0.314
$e_t^c e_{t-1}^c$	0.991	0.786	0.774	0.978	0.623	0.776	0.836	0.286	0.311
$e_t^c e_{t-2}^c$	0.990	0.784	0.783	0.978	0.035	0.773	0.378	0.971	0.534
$e_t^c e_{t-3}^c$	0.990	0.782	0.771	0.977	0.618	0.772	0.452	0.972	0.639
$I_t I_{t-1}$	0.983	0.963	0.967	0.250	0.945	0.809	0.503	0.863	0.073
$I_t I_{t-2}$	0.936	0.393	0.978	0.978	0.125	0.994	0.903	0.980	0.922
$I_t I_{t-3}$	0.882	0.860	0.678	0.928	0.827	0.720	0.910	0.915	0.049

Note: we test the accuracy of the one-day-ahead VaR forecast computed from a T-GARCH(1,1) model for different levels of risk α for three daily exchange rates - period 2010-2015. The p-value of the test statistics are reported. The moments are defined in Section 4.3.

Table XI: Backtesting of VaR forecasts for the T-GARCH(1,1) model, In sample evaluation.

	$\alpha = 0.5\%$			$\alpha = 1\%$			$\alpha = 5\%$		
	S& P 500	NIKKEI	NASDAQ	S& P 500	NIKKEI	NASDAQ	S& P 500	NIKKEI	NASDAQ
e_t	0.494	0.434	0.803	0.004	0.086	< 0.001	0.050	0.167	0.015
e_t^*	0.582	0.579	0.864	0.031	0.174	0.007	0.069	0.190	0.026
e_t^c	0.585	0.580	0.864	0.032	0.175	0.007	0.070	0.191	0.026
I_t	0.338	0.726	0.490	0.022	0.265	0.007	0.079	0.165	0.029
$e_t e_{t-1}$	0.004	< 0.001	0.250	< 0.001	0.142	0.208	0.427	0.255	0.502
$e_t e_{t-2}$	0.007	0.962	0.468	0.003	0.974	0.429	0.500	0.025	0.249
$e_t e_{t-3}$	0.707	0.639	0.604	0.946	0.825	0.307	0.265	0.577	0.624
$e_t^* e_{t-1}^*$	0.813	< 0.001	0.910	< 0.001	< 0.001	0.034	0.307	0.178	0.271
$e_t^* e_{t-2}^*$	< 0.001	0.573	< 0.001	< 0.001	0.337	< 0.001	0.479	0.012	0.853
$e_t^* e_{t-3}^*$	0.734	0.670	0.910	0.384	0.676	0.148	0.018	0.650	0.118
$e_t^c e_{t-1}^c$	0.786	< 0.001	0.826	< 0.001	< 0.001	0.090	0.540	0.489	0.438
$e_t^c e_{t-2}^c$	< 0.001	0.786	< 0.001	< 0.001	0.513	< 0.001	0.027	< 0.001	0.389
$e_t^c e_{t-3}^c$	0.784	0.776	0.826	0.419	0.509	0.363	0.089	0.344	0.436
$I_t I_{t-1}$	0.523	< 0.001	0.675	0.001	0.098	0.198	0.346	0.244	0.307
$I_t I_{t-2}$	0.001	0.932	< 0.001	< 0.001	0.862	0.037	0.521	0.024	0.243
$I_t I_{t-3}$	0.718	0.967	0.961	0.584	0.866	0.464	0.154	0.573	0.325

Note: we test the accuracy of the one-day-ahead VaR forecast computed from a T-GARCH(1,1) model for different level of risk α for three stock indexes - period 2010-2015. The p-value of the test statistics are reported. The moments are defined in Section 4.3.

Table XII: Backtesting of VaR forecasts for the T-GARCH(1,1) model, In sample evaluation.

	$\alpha = 0.5\%$			$\alpha = 1\%$			$\alpha = 5\%$		
	Yen/US	UK/US	EUR/US	Yen/US	UK/US	EUR/US	Yen/US	UK/US	EUR/US
e_t	0.472	0.346	0.024	0.266	0.051	0.001	0.378	0.001	0.046
e_t^*	0.497	0.091	0.338	0.307	0.015	0.062	0.460	0.001	0.351
e_t^c	0.340	0.112	0.340	0.176	0.024	0.071	0.213	0.004	0.407
I_t	0.342	0.113	0.342	0.178	0.025	0.072	0.218	0.004	0.412
$e_t e_{t-1}$	0.725	0.675	0.736	0.015	0.625	0.668	0.567	0.991	0.012
$e_t e_{t-2}$	0.942	0.581	0.839	0.928	0.342	0.898	0.461	0.214	0.333
$e_t e_{t-3}$	0.910	0.878	0.641	0.987	0.423	0.878	0.638	0.887	0.800
$e_t^* e_{t-1}^*$	0.617	0.699	0.753	< 0.001	0.238	0.655	0.580	0.420	0.069
$e_t^* e_{t-2}^*$	0.914	0.720	0.789	0.933	0.153	0.723	0.430	0.747	0.738
$e_t^* e_{t-3}^*$	0.888	0.208	0.834	0.926	0.570	0.674	0.350	0.891	0.982
$e_t^c e_{t-1}^c$	0.982	0.735	0.982	< 0.001	0.496	0.892	0.743	0.744	0.043
$e_t^c e_{t-2}^c$	0.982	0.735	0.982	0.965	< 0.001	0.893	0.537	0.204	0.891
$e_t^c e_{t-3}^c$	0.982	0.733	0.981	0.965	0.495	0.892	0.535	0.565	0.893
$I_t I_{t-1}$	0.982	0.736	0.982	< 0.001	0.497	0.892	0.743	0.743	0.043
$I_t I_{t-2}$	0.982	0.735	0.982	0.965	< 0.001	0.893	0.537	0.204	0.891
$I_t I_{t-3}$	0.983	0.735	0.983	0.965	0.496	0.893	0.535	0.566	0.893

Note: we test the accuracy of the one-day-ahead VaR forecast computed from a T-GARCH(1,1) model for different levels of risk α for three daily exchanges rates - period 2010-2015. The out-of-sample evaluation is based on 1000 observations to estimate the parameters and 500 observations to test the moment. The p-value of the test statistics are reported. The moments are defined in Section 4.3.

Table XIII: Backtesting of VaR forecasts for the T-GARCH(1,1) model, Out of sample evaluation.

	α = 0.5%			α = 1%			α = 5%		
	S& P 500	NIKKEI	NASDAQ	S& P 500	NIKKEI	NASDAQ	S& P 500	NIKKEI	NASDAQ
e_t	0.009	0.741	0.032	0.015	0.018	0.010	0.078	0.077	0.024
e_t^*	0.089	0.042	0.120	0.132	0.001	0.075	0.121	0.016	0.056
e_t^c	0.112	0.112	0.112	0.176	0.007	0.071	0.214	0.147	0.062
I_t	0.113	0.113	0.113	0.178	0.007	0.072	0.218	0.151	0.065
$e_t e_{t-1}$	< 0.001	< 0.001	< 0.001	< 0.001	0.564	0.149	0.172	0.979	0.509
$e_t e_{t-2}$	0.413	0.492	0.414	0.004	0.642	0.778	0.649	0.038	0.459
$e_t e_{t-3}$	0.694	0.870	0.407	0.742	0.690	0.847	0.485	0.411	0.997
$e_t^* e_{t-1}^*$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.613	0.510	0.896
$e_t^* e_{t-2}^*$	0.771	0.754	0.936	< 0.001	0.743	0.852	0.524	0.007	0.551
$e_t^* e_{t-3}^*$	0.937	0.760	0.786	0.822	0.595	0.422	0.155	0.926	0.642
$e_t^c e_{t-1}^c$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.890	0.927	0.424
$e_t^c e_{t-2}^c$	0.735	0.735	0.735	< 0.001	0.440	0.555	0.043	< 0.001	0.883
$e_t^c e_{t-3}^c$	0.735	0.733	0.735	0.617	0.439	0.555	0.418	0.930	0.882
$I_t I_{t-1}$	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.889	0.927	0.424
$I_t I_{t-2}$	0.735	0.735	0.735	< 0.001	0.441	0.556	0.043	< 0.001	0.884
$I_t I_{t-3}$	0.735	0.735	0.735	0.617	0.440	0.555	0.418	0.930	0.882

Note: we test the accuracy of the one-day-ahead VaR forecast computed from a T-GARCH(1,1) model for different levels of risk α for three stock indexes - period 2010-2015. The out-of-sample evaluation is based on 1000 observations to estimate the parameters and 500 observations to test the moment. The p-value of the test statistics are reported. The moments are defined in Section 4.3.

Table XIV: Backtesting of VaR forecasts for the T-GARCH(1,1) model, Out of sample evaluation.

E Choice of the moments

In this section, we provide a guideline on how to derive moment conditions from a discrete distributional assumption. The support of Y is supposed to be discrete and, without loss of generality, we assume it is $\mathbb{N}$. $p_i(\theta)$ denotes the probability that $Y = i$ under P_θ.

E.1 Adhoc choices

Adhoc choices of moments are always possible. For standard distributions, one generally knows the first moments (mean, variance, skewness, and kurtosis) as functions of the parameters. For discrete distributions, one can also compare the expected frequency of counts with the actual ones (this is the rationale of the standard Pearson's chi-squared test).

For the Poisson distribution, we know that it has the property of equidispersion, i.e. the mean and the variance are equal. This gives us the opportunity to test H_0 from the first and second moments together. We could alternatively use the sequence of moments $m_i(y, \theta) = \mathbf{1}\{Y = i\} - p_i(\theta)$ for different i.

E.2 Orthogonal polynomials and Ord's family of discrete distributions

The Ord's family is a well-known extension of the famous Pearson's family to the case of discrete distributions. This family includes the Poisson, binomial, Pascal (or negative binomial), and hypergeometric distributions, as particular examples.

A discrete distribution belongs to the Ord's family if the ratio (we omit the dependence of p_i in θ) $\frac{p_{y+1} - p_y}{p_y}$ equals the ratio of two polynomials $A(.)$ and $B(.)$, where $A(.)$ is affine and $B(.)$ is quadratic.

$$\frac{\Delta p_y}{p_y} = \frac{p_{y+1} - p_y}{p_y} = \frac{A(y)}{B(y)} = \frac{a_0 + a_1 y}{b_0 + b_1 y + b_2 y^2}, \tag{E.5}$$

where Δ is the forward difference operator: $\Delta p_y = p_{y+1} - p_y$.

We can build the associated orthonormal polynomial family Q_j, $j \in \mathbb{N}$, where each polynomial is derived using an analogue of the Rodrigues' formula on finite difference (see Weber and Erdelyi, 1952 or Szegö, 1967):

$$Q_j(y) = \lambda_j \frac{1}{p_y} \Delta^j \left[p_{y-j} B(y) B(y-1)...B(y-j+1) \right],$$

where λ_j is a constant which ensures that the variance of Q_j is equal to 1.

These orthonormal polynomials can be used for our testing procedure. They are not necessarily the best in terms of power or robust to parameter uncertainty. However, one advantage is that the variance is known, equal to one.

Another advantage is that the family of orthogonal polynomials is complete in L^2 (see, for example, Gallant, 1980, in a continuous case). Testing the distribution or testing the full sequence of polynomials is therefore equivalent. The next Section presents some particular examples of Ord's distributions and related polynomial families of interest. Candelon et al. (2011) used, for example, the Meixner polynomials to test the geometric distributional assumption in a VaR framework.

E.3 Examples of Ord's distributions

We provide here particular examples of discrete distributions. The definition of the orthonormal polynomial family is provided in Table XV.

E.3.1 The Poisson distribution

When $Y \sim \mathcal{P}o(\theta)$, the probability distribution function of Y is:

$$p_y = e^{-\theta} \frac{\theta^y}{y!}$$

The orthonormal family associated to the Poisson distribution is the family of Charlier polynomials $C_j^\theta(y)$. As

$$\frac{\partial \ln p_y}{\partial \theta} = -1 + \frac{y}{\theta} = -\frac{C_1^\theta(y)}{\sqrt{\theta}},$$

Charlier polynomials of degree greater or equal to 2 are robust to parameter uncertainty when one estimates the parameter θ.

E.3.2 The Pascal distribution

The Pascal distribution is also known as the negative binomial distribution. It extends the Poisson distribution to some cases where the variance could be greater than the mean of the distribution (the overdispersion that Poisson counting processes fail to fit). The negative binomial distribution is also known as a Poisson-Gamma mixture.

When $Y \sim \mathcal{P}a(\mu, \delta)$,

$$p_y = \left(\frac{\mu}{\mu + \delta}\right)^y \left(\frac{\delta}{\mu + \delta}\right)^\delta \frac{\Gamma(y + \delta)}{\Gamma(\delta)\Gamma(y + 1)}$$

When $\delta \to +\infty$, the Pascal distribution tends to the Poisson distribution. The orthonormal polynomials associated to this distribution are the Meixner polynomials $M_j(y, \mu, \delta)$.

When $\delta = 1$, the Pascal distribution is the geometric distribution ($\alpha = \frac{1}{\mu+1}$). Candelon <u>et al.</u> (2011) test this discrete distribution in a context of backtesting.

E.3.3 The binomial distribution

The probability distribution function of the Binomial distribution is:

$$p_y = \binom{N}{y} p^y (1 - p)^{N-y}$$

where $p \leq 1$

In this case, the orthogonal polynomials $K_j(y, N, p)$ are the Krawtchouk polynomials. They can be used for testing probit and logit models.

Name	p_y	A	B	Q_1
	Recursive relationship			
Poisson	$e^{-\mu}\frac{\mu^y}{y!}$	$-(y-\mu+1)$	$y+1$	$\frac{\mu-y}{\sqrt{\mu}}$
	$Q_{j+1}(y) = \frac{\mu+j-y}{\sqrt{\mu(j+1)}}Q_j(y) - \sqrt{\frac{j}{j+1}}Q_{j-1}(y)$			
Pascal	$\left(\frac{\mu}{\mu+\delta}\right)^y\left(\frac{\delta}{\mu+\delta}\right)^\delta\frac{\Gamma(y+\delta)}{\Gamma(\delta)\Gamma(y+1)}$	$\frac{\mu}{\mu+\delta}(y+\delta)-(y+1)$	$y+1$	$\frac{\mu\delta-\delta y}{\sqrt{\mu\delta(\mu+\delta)}}$
	$Q_{j+1}(y) = \frac{\mu(2j+\delta)+\delta(j-y)}{\sqrt{\mu(\mu+\delta)(j+\delta)(j+1)}}Q_j(y) - \sqrt{\frac{j(\delta+j-1)}{(j+1)(\delta+j)}}Q_{j-1}(y)$			
Geometric	$(1-\alpha)^y\alpha$	$-\alpha(y+1)$	$y+1$	$\frac{1-\alpha-\alpha y}{\sqrt{1-\alpha}}$
	$Q_{j+1}(y) = \frac{(1-\alpha)(2j+1)+\alpha(j-y)}{\sqrt{1-\alpha}(j+1)}Q_j(y) - \frac{j}{j+1}Q_{j-1}(y)$			
Binomial	$\binom{N}{y}p^y(1-p)^{N-y}$	$-(y-Np+q)$	$q(y+1)$	$\frac{pN-y}{\sqrt{pqN}}$
	$Q_{j+1}(y) = \frac{p(N-j)+qj-y}{\sqrt{pq(N-j)(j+1)}}Q_j(y) - \sqrt{\frac{j(N-j+1)}{(j+1)(N-j)}}Q_{j-1}(y)$			

$\frac{p_{y+1}-p_y}{p_y} = \frac{A(y)}{B(y)}$. Q_j is the orthogonal polynomial of degree j, normalized.

E.4   A general class of moments

The two previous sections present some particular moments that can be used for testing purposes. There are however some cases where such moments are not so easy to derive. We propose here a general rule for constructing any moment for which the expectation under the null is equal to zero. Let ψ be a function defined on $\mathbb{N} \times \Theta$ and such that its expectation under P_θ is finite. We assume that $\psi(0, \theta) = 0$, this is just a normalization.

Proposition 6. *Let $m(y, \theta)$ be the function defined by*

$$m(y,\theta) = \left[\psi(y+1,\theta) - \psi(y,\theta) + \frac{p_{y+1}(\theta)-p_y(\theta)}{p_y(\theta)}\psi(y+1,\theta)\right]. \tag{E.6}$$

We have:

$$\mathbb{E}_0\left[m(y,\theta^0)\right] = 0. \tag{E.7}$$

Proof. We first prove the proposition in the case where N is infinite. Take first the expectation of $\Delta\psi(\cdot)(y,\theta^0)$ under the null:

$$\mathbb{E}_0[\Delta\psi(y,\theta^0)] = \sum_{i=0}^{+\infty}\left(\psi(i+1,\theta^0) - \psi(i,\theta^0)\right)p_i(\theta^0) \tag{E.8}$$

Reordering the second term of the last expression yields to

$$
\begin{aligned}
\mathbb{E}_0[\Delta\psi(y,\theta^0)] &= \sum_{i=0}^{+\infty}\psi(i+1,\theta^0)p_i(\theta^0) - \sum_{i=0}^{+\infty}\psi(i,\theta^0)p_i(\theta^0) \\
&= \sum_{i=0}^{+\infty}\psi(i+1,\theta^0)p_i(\theta^0) - \sum_{i=1}^{+\infty}\psi(i,\theta^0)p_i(\theta^0) \text{ because } \psi(0,\theta^0)=0, \\
&= \sum_{i=0}^{+\infty}\psi(i+1,\theta^0)p_i(\theta^0) - \sum_{i=0}^{+\infty}\psi(i+1,\theta^0)p_{i+1}(\theta^0) \\
&= -\sum_{i=0}^{+\infty}\psi(i+1,\theta^0)\left(p_{i+1}(\theta^0) - p_i(\theta^0)\right) \\
&= -\mathbb{E}_0\left[\psi(y+1,\theta^0)\frac{p_{y+1}(\theta^0) - p_y(\theta^0)}{p(y,\theta^0)}\right].
\end{aligned}
$$

Consequently

$$
\mathbb{E}_0\left[\Delta\psi(y,\theta^0) + \left(\psi(y+1,\theta^0)\frac{p_{y+1}(\theta^0) - p_y(\theta^0)}{p(y,\theta^0)}\right)\right] = 0.
$$

Observe that the quantity inside the brackets is exactly $m(y,\theta^0)$. When N is finite, the proof is similar as $p_i(\theta)$ is equal to zero when $i \geq (N+1)$.

$\square$

We illustrate the usefulness of Proposition 6 previous results by considering the geometric distribution with parameter θ. In this case, $p_y(\theta) = (1-\theta)^y\theta$ and $\frac{p_{y+1}(\theta)-p_y(\theta)}{p_y(\theta)} = -\theta$. When $\psi(y,\theta) = y$, we obtain the first Meixner polynomial, up to some scale factor, $1 - \theta - \theta y$. When $\psi(y,\theta) = y^2$, the moment derived from (E.6) is a linear combination of the first two Meixner polynomials. The family of functions y^k generates the first k terms of the Meixner family.

More generally, Proposition 6 generates a set of moments when one does not have any obvious moment to use.

One could argue that focusing on this class could restrict the range of the tests derived from these moment conditions. It might be the case that the set of moments generated by Equation (E.6) could be a small subset of the set of moments with zero expectation. The next proposition shows that any moment with zero expectation can be generated by the construction presented above.

Proposition 7. *Let $m(y,\theta)$ be a moment such that*

$$
\mathbb{E}_0\left[m(y,\theta^0)\right] = 0. \tag{E.9}
$$

Let $\psi(y,\theta)$ be a function defined on S by:

$$
\psi(0,\theta) = 0,
$$
$$
\psi(y,\theta) = \frac{1}{p_y(\theta)}\sum_{k=0}^{y-1}m(k,\theta)p_k(\theta) \text{ for } y \geq 1 \tag{E.10}
$$

Then, $m(\cdot)$ satisfies the equality in Eq. (E.6).

Proof. Following the definition of $\psi(\cdot)$,

$$\Delta\psi(y,\theta) + \psi(y+1,\theta)\frac{\Delta p_y(\theta)}{p_y(\theta)}$$

$$= \psi(y+1,\theta) - \psi(y,\theta) + \psi(y+1,xy,\theta)\left(\frac{p_{y+1}(\theta)}{p_y(\theta)} - 1\right)$$

$$= \psi(y+1,\theta)\frac{p_{y+1}(\theta)}{p_y(\theta)} - \psi(y,\theta)$$

$$= \frac{1}{p_y(\theta)}\sum_{k=0}^{y} m(k,\theta)p_k(\theta) - \frac{1}{p_y(\theta)}\sum_{k=0}^{y-1} m(k,\theta)p_k(\theta) \qquad \text{using Equation (E.10),}$$

$$= m(y,\theta).$$

$\square$

Observe that the last equality holds without the expectation.