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 nonsmooth 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.

I. 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 & Sentana, 2011; Amengual, Fiorentini, & Sentana, 2013; Bai & Ng, 2005; Bontemps & Meddahi, 2005, 2012; Candelon et al., 2011; Dufour, Khalaf, & Beaulieu, 2003; Fiorentini, Sentana, & Calzolari, 2004; Mencia & Sentana, 2012), forecasting (Diebold & Mariano, 1995, West, 1996; West & McCracken, 1998; McCracken, 2000), and microeconomics (Butler & Chatterjee, 1997; Skeels & Vella, 1999; Tauchen, 1985; Mora & Moro-Egido, 2008). In addition, structural econometric models provide moment-based equations from testable overidentifying restrictions. For example, structural search models (see Jolivet, Postel-Vinay, & Robin, 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, Lilliefof (1967) rebalutes 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 section IIB). However, a tractable expression is required for the estimated 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 one that is orthogonal to the underlying score function. We call such a moment a robust moment. As we explain 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 nonsmooth 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 nonsmooth 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. It also 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 III.

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 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. A robust moment is robust whether the data 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 & 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 III, we consider an out-of-sample case, including, in section IIIC, 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 II 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 III characterizes the advantages of using robust moments in out-of-sample contexts and presents some examples. Section IV 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 V considers an empirical application to test the VaR forecasts derived from a T-GARCH(1,1) model for daily exchange rate data. Section VI concludes the paper.

The supplemental material contains appendices that provide the proofs (appendix B) and additional analysis.

II. General Results

A. Setup 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 $P_\theta$ indexed by $\theta \in \Theta \subset \mathbb{R}^r$. 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 $\alpha$, it implies that the mean of $Y$ is equal to $\alpha$. 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.

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, $\theta$ denotes the vector of parameters that are estimated and is generally estimated in-sample from $y_1, \ldots, y_T$. The true value of parameter $\theta$ is denoted by $\theta^0$, and $E_0$ and $V_0$ 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 $E_0[h_1 h_2 \top]$ the matrix $E_0[h_1(y, \theta^0) h_2 \top(y, \theta^0)]$.

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

$$E_0[m(y_t, \theta^0)] = 0.$$ 

Our procedure consists of testing whether the empirical average of these moments is close to 0 when $\theta$ is also estimated.

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

---

1 Bontemps et al. (2018) study point optimal moment-based procedures.

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, $P_t$ would become $P_{t|x}$, and we would be able to test unconditional moments implied by the conditional distribution of $Y \mid X = x$.

3 See section III for the out-of-sample case.

4 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.
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 \( \alpha \), the coverage level of the VaR. It implies some moment restrictions that we test, which is one of our leading examples that we address in detail in section IV.

**Test statistic when the parameter is known.** For a benchmark, we first present the hypothetical case in which the true value \( \theta^0 \) of the parameter \( \theta \) is known.

**Assumption CLT: Central Limit Theorem.** Throughout this paper, we assume that the long-run covariance matrix of \( m(\cdot) \), \( \Sigma \), is finite and positive definite and that the CLT applies.\(^6\)

Under assumption CLT, a test statistic \( \xi_m \) can be constructed from any consistent estimator \( \hat{\Sigma}^{-1} \) of \( \Sigma^{-1} \):

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

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 empirical finance, VaR is the most common. Most existing tests are found, for example, in corollary 5.3 of Hall and Heyde (1980)).

First, any moment \( m(\cdot) \) continuously differentiable in a neighborhood of \( \theta^0 \) satisfies assumption GIM; this expression is used, for example, in Newey and McFadden (1994). In addition, this paper considers the following class of nonsmooth 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 \( \theta \). Such a moment estimates the frequency of a given interval or class and compares it with the expected frequency and is often used in discrete distributions; Pearson’s chi-squared test is a famous example. Following Tauchen (1985), any moment in this class satisfies assumption GIM.

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 \( \theta^0 \) that satisfies assumption REG. Under assumption GIM, the sequence \( m(y_1, \hat{\theta}), \ldots, m(y_T, \hat{\theta}) \) satisfies the following expansion:

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(y_t, \hat{\theta}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(y_t, \theta^0) - \mathbb{E}_0 \left[ m_{\theta_0}^\top \right] \sqrt{T} (\hat{\theta} - \theta^0) + o_P(1). \tag{4}
\]
Equation (4) is generally known in the differentiable case because it is the first-order expansion in which $E_0 \left[ ms_{\theta} \right]$ is replaced by $-E_0 \left[ m m_{\theta} \right]$, which is proved by Tauchen (1985) in the nonsmooth 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 $\theta^0$ generally modifies the asymptotic variance, as equation (4) indicates. Ignoring this correction 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 $\theta^0$. Equation (4) highlights the two strategies to address the impact of parameter uncertainty.

The first strategy, which we call correcting, consists of deriving the joint asymptotic distribution of the two terms on the right-hand side of equation (4), as in Newey (1985) and 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 and the score function, as equation (4) indicates. Ignoring this correction 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 $\theta^0$. Equation (4) highlights the two strategies to address the impact of parameter uncertainty.

In this paper, we transform any moment $m(\cdot)$ orthogonal to the true score function, which we call robust moments. For robust moments, the asymptotic distribution of $\frac{1}{\sqrt{T}} m(y, \hat{\theta})$ is the same, at the first order, as the asymptotic distribution of $\frac{1}{\sqrt{T}} m(y, \theta^0)$ because the second term on the right-hand side of equation (4) is equal to 0. 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 orthogonal projection of the moment of interest $m(\cdot)$ onto the space $S$, the space orthogonal to the space spanned by the score. Specifically, the transformed moment is

$$m^\perp(y, \theta) = m(y, \theta) - E_0 \left[ ms_{\theta}^\top \right] V_{0}[s_0]^{-1} s_0(y).$$

### C. Orthogonalization Methods

In this section, 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.

**Robustification by oblique projection.** Consider an estimating function $g(\cdot)$ that can identify parameter $\theta$ and satisfy assumption CLT and assumption GIM. This estimating function can be used to estimate $\theta$, but we do not impose it here. We assume that $g(\cdot)$ has the same dimension as $\theta$, 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

$$\tilde{m}_g(y, \theta) = m(y, \theta) - \left( \frac{\partial E_0 \left[ m(y, \theta) \right]}{\partial \theta^\top} \right)_{\theta=\theta^0} \times \left( \frac{\partial E_0 \left[ g(y, \theta) \right]}{\partial \theta} \right)^{-1}_{\theta=\theta^0} g(y, \theta),$$

and this moment is robust to parameter estimation uncertainty.

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

$$\left( \frac{\partial E_0 \left[ m(y, \theta) \right]}{\partial \theta^\top} \right)_{\theta=\theta^0} = -E_0 \left[ ms_{\theta}^\top \right],$$

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

$$\tilde{m}_g(y, \theta) = m(y, \theta) - E_0 \left[ ms_{\theta} \right] E_0 \left[ g s_{\theta}^\top \right]^{-1} g(y, \theta).$$

This moment is clearly orthogonal to the true score function. Note that equation (5), the orthogonal projection onto the orthogonal space spanned by the score, is a specific case of equation (7) with $g = s_0$.

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 equation (6) does not use the score function but the derivatives of the functions of interest with respect to $\theta$. In many cases, these quantities are easy to derive analytically (see the examples in sections III and IV). 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
\[
\frac{\partial E_0[\mu(y, \theta)]}{\partial \theta} = E_0 \left[ \frac{\partial \mu}{\partial \theta} \right].
\]
and similarly for \( g(\cdot) \).

As we discuss below, many choices exist for \( g(\cdot) \). The empirical researcher should be aware that there is no "best choice" for \( g(\cdot) \) without a specific objective. 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.

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 only on 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).

In addition, 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 nonstandard.

We now present another interesting property of robust moments: we can indeed loosen assumption REG.

Proposition 3. Let \( \hat{m}_h \) be a robust moment defined as above. When \( T^{\alpha} (\hat{\theta} - \theta^0) \sim O_P(1) \) for \( \alpha > 1/4 \),
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{m}_h(y_t, \hat{\theta}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{m}_h(y_t, \theta^0) + O_P(1).
\]

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. In addition, 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 \( \theta \) 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.

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 \( \mathcal{M} \) (the auxiliary model) defined by the parametric family of distributions \( \hat{P}(y; \beta) \), and let \( s_P(\cdot) \) 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 \( \hat{P}(y; \beta) \). A moment orthogonal to \( s_P(\cdot) \) 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 \( t - 1 \), \( \varepsilon_t \) is an i.i.d. variable with a known distribution, and \( \beta \) 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, 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 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 IV.

Example of 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 \( \beta \) is linked to some exogenous variables \( X \), \( \beta = h(X, \theta) \), where \( \theta \) is a parameter vector to be estimated, the same moment
\(^8\) See the proof in section B.2 in the supplemental material.
remains robust when $\theta$ is estimated from the data. Following proposition 3, the convergence rate for $\theta$ can also be the nonparametric rate.

**Local power properties and choice of $g$.** We have presented 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. In addition, 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 independent 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 $\hat{m}_g(\cdot)$ 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 correcting 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 $g(\cdot)$ 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 $\theta$ as in proposition 2 and $\hat{m}_g$ the robust version of $m(\cdot)$ after projection along the direction $g$ (see Equation [6]). Assume that under the (local) alternative, the pdf is $q_1 = q_0(1 + h(y) / \sqrt{T})$. The power function from the test based on $\hat{m}_g$ is an increasing function of the parameter

$$a(g) = \frac{\mathbb{E}_0[\hat{m}_g h]}{\sqrt{\text{Var}[\hat{m}_g]}}.$$

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 $\hat{m}_{g_1}$ is better than $\hat{m}_{g_2}$ 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.

**III. Out-of-Sample Evaluation of Robust Moments**

In this section, we focus on moments evaluated out-of-sample. As West and McCracken (1998), in particular, noted, 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 nonsmooth moments.

**A. 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 $\hat{\theta}_i$ be a sequence of square-root-consistent GMM-type estimators of $\theta$ using the data $y_{1-r}, \ldots, y_{i-1}$ (rolling estimator), $y_1, \ldots, y_{i-1}$ (recursive estimator), or $y_1, \ldots, y_R$ (fixed estimator). We assume that $\hat{\theta}_i$ satisfies assumption REG for the corresponding values of the time index. We also assume that $R$ and $P$ tend to $\infty$ while $\sqrt{P}/R$ tends to 0 and that $m(\cdot)$ satisfies assumptions CLT and GIM. If $m(\cdot)$ is a robust moment,

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

The proof is a direct consequence of the fact that the second term in the asymptotic expansion vanishes owing to the 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 nonsmooth 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 $\theta^0$.

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.

**B. 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 $\theta$.

---

9 The proof is provided in section B.3 of the supplemental material.
Testing for first-order correlation in a regression model. Consider the stationary model,\(^{10}\)
\[ y_t = x_t^\top \theta_0 + \epsilon_t. \]

The goal is to test whether \( E_0[\epsilon_t \epsilon_{t-1}] = 0 \) from the estimated residuals; \( \hat{\theta}_1 = y_t - x_t^\top \hat{\theta}_0 \) is computed using one of the three schemes.

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 \( m = \epsilon_t \epsilon_{t-1} \) 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:
\[
E_0 \left[ \frac{\partial m}{\partial \theta} \right] = -E_0 \left[ x_t^\top \epsilon_{t-1} \right] \quad \text{and} \quad E_0 \left[ \frac{\partial g}{\partial \theta^\top} \right] = -E_0 \left[ x_t x_t^\top \right].
\]

Thus, the robust version of \( m \) is
\[ m^\perp = \epsilon_t \epsilon_{t-1} - (1 - \rho^2) y_{t-1} \epsilon_t. \]

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
\[ E_0 \left[ v_{1t} (x_{2t}^\top \beta_2^*) \right] = 0. \]

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
\[ m^\perp = v_{1t} (x_{2t}^\top \beta_2^*) - E_0 \left[ x_{1t} (x_{2t}^\top \beta_2^*) \right] \left( E_0 \left[ x_{1t} x_{1t}^\top \right] \right)^{-1} x_{1t} v_{1t}. \]

Test of equal MAE. We next consider a nonsmooth 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 II. We define the robust version according to equation (6) as follows. For each model \( i \), \( i = 1 \) or 2, the first matrix
\[
\left( \frac{\partial E_0[|y_t - x_t^\top \beta_{i*}|]}{\partial \beta_{i*}} \right)_{\beta_{i*} = \beta_{i0}} \]
is equal to \( -E_0 \left[ sgn(u_{it}) x_{it}^\top \right] \). Because \( g(\cdot) \) is smooth, we obtain the usual expression for the second matrix—that is, \( -E_0 \left[ x_{it} x_{it}^\top \right] \). Consequently,
\[ m^\perp = u_{1t}^\top - u_{2t}^\top, \]
where \( u_{it} = |u_{it}| - E_0 \left[ sgn(u_{it}) x_{it}^\top \right] \left( E_0 \left[ x_{it} x_{it}^\top \right] \right)^{-1} u_{it} x_{it}, i = 1, 2. \]

C. A Small Monte Carlo Exercise

Now, we consider the first-order serial correlation test when one estimates an AR(1) model with mean
\[ y_t = \mu + \rho y_{t-1} + \epsilon_t. \]

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 \( \epsilon_t \epsilon_{t-1} - (1 - \rho^2) y_{t-1} \epsilon_t \), which has long-run variance \( \rho^2 \sigma^4 \), where \( \sigma^2 \) is the variance of \( \epsilon_t \). 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 a fair comparison.\(^{12}\) Under the null hypothesis, as in our proposed test, \( BP(2) \) is asymptotically \( \chi^2(1) \) distributed when \( \mu \) and \( \rho \) 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 \( \mu = 0, \epsilon_t \sim N(0,1) \) and autocorrelation \( \rho \), where \( \rho \) takes values from 0.2 to 0.99. Both \( \mu \) and \( \rho \) 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 \( \rho \) increases and approaches 1, size distortion occurs for both the Box-Pierce and Ljung-Box tests.

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

\(^{10}\) Model 6.1 in West and McCracken (1998).

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

\(^{12}\) 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.
et al., 1999). However, VaR is the measure commonly used to
manage risk and whether VaR is adequate (see Artzner
management measure. There is a debate on what characterizes a
consider here the rolling scheme. 14 We present the results for
parameter uncertainty problem for the recursive scheme. We
tests. West and McCracken (1998) show that there is no
We report the rejection frequencies from 10,000 simulations. We compare our robust moment
of the negative of the
material, lead to similar conclusions.

The DGP is a univariate AR(1) of size \(T\). The two inverse roots of the process are \(\rho_1 = 0.5\) and \(\rho_1 \). We estimate an AR(1) process with mean and test the first-order autocorrelation in-sample. We report the rejection frequencies from 10,000 simulations. We compare our robust moment \(m^*_t\) defined in equation (9) with the Box-Pierce and Llung-Box tests for two autocorrelations.

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.

IV. 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 (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 \(\text{VaR}_{t+1}^\alpha\) be the one-day-ahead VaR forecast (computed at time \(t - 1\)) for a given level of risk \(\alpha\) (the value considered is generally 5% or 1%). With an abuse of notation, we consider the VaR measure, \(\text{VaR}_{t+1}^\alpha\), as the negative of the \(\alpha\)-quantile of the conditional distribution of \(r_t\) given \(J_{t-1}\), the information set at date \(t - 1\):\(^{15}\)

\[
P \left( r_t \leq -\text{VaR}_{t+1}^\alpha \mid J_{t-1} \right) = \alpha. \tag{10} \]

\(^{14}\) The fixed scheme results, available in appendix C of the supplemental material, lead to similar conclusions.

\(^{15}\) In fact, the VaR measure is the potential loss induced by this negative return.

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. Under $H_0$—the VaR parametric model used by the financial institution is the right model—$I_t$ is i.i.d. Bernoulli distributed with parameter $\alpha$.

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. Cristofferansen (1998) considers a likelihood ratio test in a Markov framework. Cristofferansen 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, this strategy depends on the underlying model used for the returns. In addition, 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 parameter uncertainty introduced by the estimation of the conditional variance for the returns. This parameter uncertainty has rarely been considered in the literature.

A. Robust Moments for Backtesting VaR Models

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

Building a robust moment for backtesting in practice. Assume that the model for returns is the following constant location-scale model,

$$r_t = \mu^0 + \sigma^0 \epsilon_t,$$

where $\epsilon_t \sim i.i.d. \mathcal{N}(0,1)$, a continuous distribution with mean 0 and variance 1. 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 an integer (this is the case considered in Escanciano & 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 $\alpha$, 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.
2. Calculate $\frac{\partial g}{\partial \theta^0}(r_t, \theta)$ at $\theta = \theta^0$, which can be done numerically or explicitly.
3. Similarly, calculate $\frac{\partial g}{\partial \theta^0}(r_t, \theta)$ at $\theta = \theta^0$.
4. Apply equation (6) and calculate the test statistic $\xi$ in equation (1).

We now apply this method to our specific case. First, a simple estimating function for $\mu^0$ and $\sigma^0$ is

$$g = \left( \frac{r_t - \mu}{(r_t - \mu)^2 - \sigma^2} \right).$$

(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 $\theta$, $\theta = (\mu, \sigma)^T$:

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

The expectation at $\theta = \theta^0$ is therefore $V = \text{diag}(-1, -2\sigma^0)$. Third, we compute $P = \left( \frac{\partial g}{\partial \theta^0} \right)_{\theta=\theta^0}$, which is equal to

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

where $q_\alpha$ is the $\alpha$ quantile of the distribution of $\epsilon$, and $f(\cdot)$ is the probability distribution function.

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

$$m_t^\perp = I_t - \alpha + f(q_\alpha) \epsilon_t + \frac{q_\alpha f(q_\alpha)}{2} (\epsilon_t^2 - 1),$$

(14)

where the variance, $V$ of $\epsilon_t$ depends on the distribution $D$ assumed for $\epsilon_t$. 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 equation (13). Now we apply the above result. $m_t^\perp$ is also robust for any GARCH model,

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

(15)

where $J_{t-1}$ is the information set at time $t - 1$ and $\theta$ is a vector of parameters. Different choices of $g(\cdot)$ generate different robust moments. In addition, we can also consider the true model (15) to orthogonalize.
satisfies $E_0 \left[ m_1^T \right] = 0$ and is robust to parameter uncertainty. Thus, the corresponding test statistic, $\xi_Z = T \left( \frac{1}{T} \sum_{t=1}^{T} Z_{t-1} e_t^2 \right)^T \left( E_0 \left[ Z_{t-1} Z_{t-1}^T \right] V_0 \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} Z_{t-1} e_t \right)$, 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 section, we study different choices for the above instruments in the past information set. $Z_{t-1}$ corresponds to the unconditional test (i.e., we test that the frequency of hits is the expected one, $\alpha$); $Z_{t-1}$ can also be past values, $e_{t-1}$, $e_{t-2}$, and so on. Furthermore, linear combinations of past values of $e_t$ are also possible.

C. 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: $r_t = \sqrt{\sigma_t^2(\theta)} e_t$, $\sigma_t^2(\theta) = \omega + \gamma r_{t-1}^2 + \beta \sigma_{t-1}^2$, with $e_t \sim D(0, 1)$, $\omega = 0.0001, \gamma = 0.045$, and $\beta = 0.95$. We successively consider the standard normal distribution and
the standardized Student’s distribution for the distribution of $\varepsilon_t$.

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, $VaR^\alpha_t$, for $\alpha$ 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.

**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 and then the power of different competing tests, 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 $e^*_t$ 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, $e^*_t = I_t - \alpha$ is a nonrobust moment for which we correct for parameter uncertainty, as shown in section 2.B. We also consider covariance tests based on the product of these moments with their past values, that is, $\varepsilon_t \varepsilon_{t-h}$ for $h = 1$, 2, 3 and similarly for the other moments.

The size properties are good, despite the slight overrejection in the covariance tests.

The power properties are studied for two alternatives. In the first one, the distribution for the innovation terms $\varepsilon_t$ is a standardized Student’s distribution with 3 degrees of freedom. When computing the VaR measure, Gaussianity is assumed, wrongly. In the second one, we simulate an EGARCH model 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 built from considering the real score function.

**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 $\varepsilon_t$ is the standardized Student’s distribution with $\nu = 6$ degrees of freedom.

We have now one additional parameter to estimate, $\nu$, 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. $e^*_t$ 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 $e^*_t$ denotes, as before, $I_t - \alpha$, which is a nonrobust moment. We use the correction strategy to take 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. 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. 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 $e^*_t$ performs well.

---

18 The in-sample results are given in appendix C of the supplemental material.
19 $\sigma^2_t = \exp(0.0001 + 0.9 \ln \sigma^2_{t-1} + 0.3(\varepsilon_{t-1}^2 - 2\gamma/\pi) - 0.8s_{t-1})$.
20 The in-sample properties are given in appendix C of the supplemental material.
21 The skewed $t$-distribution with $\nu$ degrees of freedom and parameter $\gamma$ has the following density $g$, where $f$ is the density of the standard Student’s distribution:

$$g(x) = \frac{2}{\gamma + 1/\gamma} f(\gamma x) \text{ when } x < 0, \quad g(x) = \frac{2}{\gamma + 1/\gamma} f(x/\gamma) \text{ when } x \geq 0.$$
We test the accuracy of the one-day-ahead VaR forecasts computed from a T-GARCH(1,1) model for different levels of risk $\alpha$ for the four daily exchange rates. The $p$-values of the test statistics are reported. The notations are defined in section IV.C.

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 from October 1, 1981, to June 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 T-GARCH(1,1) model for all but the SF/U.S. 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 $\alpha$, the risk exposure. The in-sample estimates are shown in 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. We also include the nonrobust tests based on the number of VaR violations, $I_{\alpha}$, ignoring (wrongly) the parameter uncertainty issue. The $p$-values of the tests are presented in table 7.

Note that for each exchange rate, there is always one risk level, $\alpha$, 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/U.S. and yen/U.S.). For $\alpha = 0.5\%$, no unconditional test is rejected. The unconditional tests are rejected for the SF/U.S. series with $\alpha = 1\%$ and for the U.K./U.S. 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 table 8 show that there are too many VaR violations (6.6% instead of 5%) for this exchange rate and this risk exposure.

The covariance tests are often rejected except for at low values of $\alpha$ (but for $\alpha = 0.05\%$, we expect approximately one VaR violation per year, which reduces the power of the covariance tests). For the SF/U.S. series, many rejections occur for $\alpha = 1\%$. We know that, globally, the T-GARCH assumption is rejected. The same is true for the yen/U.S. exchange rate, where the covariance tests are systematically rejected. Therefore, the conditional variance model should be adapted.

In table 9, 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 from the in-sample behavior.

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

22 The U.K. pound, French franc, Swiss franc, and Japanese yen rates, all versus the U.S. dollar.

### Table 7.—Backtesting of VaR Forecasts for the T-GARCH(1,1) Model: In-Sample Evaluation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.5%$</td>
<td>0.644</td>
<td>0.284</td>
<td>0.972</td>
<td>0.563</td>
</tr>
<tr>
<td>$\alpha = 1%$</td>
<td>0.357</td>
<td>0.784</td>
<td>0.003</td>
<td>0.897</td>
</tr>
<tr>
<td>$\alpha = 5%$</td>
<td>0.011</td>
<td>0.291</td>
<td>0.277</td>
<td>0.277</td>
</tr>
</tbody>
</table>

### Table 8.—Percentage of VaR Violations: Daily Exchange Rates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 0.005$</td>
<td>0.005</td>
<td>0.004</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td>$p = 0.010$</td>
<td>0.010</td>
<td>0.013</td>
<td>0.010</td>
<td>0.018</td>
</tr>
<tr>
<td>$p = 0.050$</td>
<td>0.050</td>
<td>0.066</td>
<td>0.044</td>
<td>0.055</td>
</tr>
</tbody>
</table>

V. Empirical Application
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 $\beta$ for the conditional variance are all greater than 0.9). The covariance tests for the Yen/U.S. series are rejected for all values of $\alpha$. The VaR forecasts for the FF/U.S. exchange rate appear to be accurate, and no test is rejected. By contrast, there is a problem with the local dynamics of the U.K./U.S. exchange rate, as many covariance tests are rejected for all risk levels $\alpha$.

In appendix D of the supplemental material, we explore additional series by considering three other exchange rates—the U.S. dollar versus the yen, the British pound, and the euro—for the period 2010 to 2015. We also consider three additional series by considering three other exchange rates—the Yen/U.S. series are rejected for all values of $\alpha$ (the $p$-values of the test statistics are reported). The $\chi^2$ tests for the four daily exchange rates are based on a robust moment 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; in addition, 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: 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 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.

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 $\beta$ for the conditional variance are all greater than 0.9). The covariance tests for the Yen/U.S. series are rejected for all values of $\alpha$. The VaR forecasts for the FF/U.S. exchange rate appear to be accurate, and no test is rejected. By contrast, there is a problem with the local dynamics of the U.K./U.S. exchange rate, as many covariance tests are rejected for all risk levels $\alpha$.

In appendix D of the supplemental material, we explore additional series by considering three other exchange rates—the U.S. dollar versus the yen, the British pound, and the euro—for the period 2010 to 2015. We also consider three additional series by considering three other exchange rates—the Yen/U.S. series are rejected for all values of $\alpha$ (the $p$-values of the test statistics are reported). The $\chi^2$ tests for the four daily exchange rates are based on a robust moment 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; in addition, 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: 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 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.

<table>
<thead>
<tr>
<th>$\alpha = 0.5%$</th>
<th>$\alpha = 1%$</th>
<th>$\alpha = 5%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>0.474</td>
<td>0.443</td>
</tr>
<tr>
<td>$e_2$</td>
<td>0.083</td>
<td>0.320</td>
</tr>
<tr>
<td>$e_3$</td>
<td>0.107</td>
<td>0.105</td>
</tr>
<tr>
<td>$l_1$</td>
<td>0.113</td>
<td>0.113</td>
</tr>
</tbody>
</table>

23 S&P 500, NIKKEI, and NASDAQ.
REFERENCES


