CAEPR
INDIANA UNIVERSITY
CENTER FOR APPLIED ECONOMICS AND POLICY RESEARCH

CAEPR Working Paper
#2007-005_updated

Backtesting Parametric Value-at-Risk with Estimation Risk

Juan Carlos Escanciano
Indiana University Bloomington

Jose Olmo
City University, London

Original: March 19, 2007
Updated: September 4, 2008

BACKTESTING PARAMETRIC VALUE-AT-RISK WITH ESTIMATION RISK

J. Carlos Escanciano
Indiana University, Bloomington, IN, USA

Jose Olmo
City University, London, UK

This draft, 21 February 2008

Abstract

One of the implications of the creation of Basel Committee on Banking Supervision was the implementation of Value-at-Risk (VaR) as the standard tool for measuring market risk. Since then, the capital requirements of commercial banks with trading activities are based on VaR estimates. Therefore, appropriately constructed tests for assessing the out-of-sample forecast accuracy of the VaR model (backtesting procedures) have become of crucial practical importance. In this paper we show that the use of the standard unconditional and independence backtesting procedures to assess VaR models in out-of-sample composite environments can be misleading. These tests do not consider the impact of estimation risk and therefore may use wrong critical values to assess market risk. The purpose of this paper is to quantify such estimation risk in a very general class of dynamic parametric VaR models and to correct standard backtesting procedures to provide valid inference in out-of-sample analyses. A Monte Carlo study illustrates our theoretical findings in finite-samples and shows that our corrected unconditional test can provide more accurately sized and more powerful tests than the uncorrected one. Finally, an application to $S\&P500$ Index shows the importance of this correction and its impact on capital requirements as imposed by Basel Accord.

Keywords and Phrases: Backtesting; Basel Accord; Conditional Quantile; Estimation Risk; Forecast evaluation; Fixed, rolling and recursive forecasting scheme; Risk management; Value at Risk.

1 Introduction

In the aftermath of a series of bank failures during the seventies a group of ten countries (G-10) decided to create a committee to set up a regulatory framework to be observed by internationally active banks operating in these member countries. This committee coined as Basel Committee on Banking Supervision (BCBS) was intended to prevent financial institutions, in particular banks, from operating without effective supervision.

The subsequent documents derived from this commitment focused on the imposition of capital requirements for internationally active banks intending to act as provisions for losses from adverse market fluctuations, concentration of risks or simply bad management of institutions. The risk measure agreed to determine the amount of capital on hold was the Value-at-Risk (VaR). In financial terms, this is the maximum loss on a trading portfolio for a period of time given a confidence level. In statistical terms, VaR is a quantile of the conditional distribution of returns on the portfolio given agent's information set. More formally, denote the real-valued time series of portfolio returns or Profit and Losses (P&L) account by Y_t, and assume that at time $t-1$ the agent's information set is given by W_{t-1}, which may contain past values of Y_t and other relevant economic and financial variables, *i.e.*, $W_{t-1} = (Y_{t-1}, Z'_{t-1}, Y_{t-2}, Z'_{t-2} ...)'$. Henceforth, A' denotes the transpose matrix of A. Let $\mathcal{F}_{t-1}$ be the σ-algebra generated by W_{t-1}. Assuming that the conditional distribution of Y_t given W_{t-1} is continuous, we define the α-th conditional VaR of Y_t given W_{t-1} as the $\mathcal{F}_{t-1}$-measurable function $q_\alpha(W_{t-1})$ satisfying the equation

$$P(Y_t \leq q_\alpha(W_{t-1}) \mid W_{t-1}) = \alpha, \text{ almost surely (a.s.)}, \alpha \in (0,1), \ \forall t \in \mathbb{Z}. \tag{1}$$

In *parametric* VaR inference one assumes the existence of a parametric family of functions $\mathcal{M} = \{m_\alpha(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ and proceeds to make VaR forecasts using the model $\mathcal{M}$. Inference within the model, including forecast analysis, depends crucially on the hypothesis that $q_\alpha \in \mathcal{M}$, *i.e.*, if there exists some $\theta_0 \in \Theta$ such that $m_\alpha(W_{t-1}, \theta_0) = q_\alpha(W_{t-1})$ a.s. In parametric models the nuisance parameter θ_0 belongs to Θ, with Θ a compact set in an Euclidean space $\mathbb{R}^p$. Semiparametric and nonparametric specifications for $q_\alpha(\cdot)$ have also been considered, see *e.g.* Fan and Gu (2003), Martins-Filho and Yao (2006) and references therein, where θ_0 belongs to an infinite-dimensional space. This paper will focus on parametric VaR models where θ_0

is finite-dimensional and can be estimated by a $\sqrt{R}$-consistent estimator, with R denoting the (in-)sample size (cf. A4 below.) Parametric VaR models are popular since the functional form $m_\alpha(W_{t-1}, \theta_0)$, jointly with the parameter θ_0, describes in a very precise way the impact of the agent's information set on the VaR. The most popular parametric VaR models are those derived from traditional location-scale models such as ARMA-GARCH models. Our empirical analysis will be focused on these models, although our theoretical results go beyond location-scale models. Alternative parametric VaR models can be found in *e.g.* Engle and Manganelli (2004), Koenker and Xiao (2006) and Gourieroux and Jasiak (2006), among many others.

The computation of VaR measures has become of paramount importance in risk management. In fact, for banks with sufficiently highly developed risk management systems the implementation of VaR techniques was a priori the only restriction set by the Basel Accord (1996a) for computing capital reserves. Thus, in order to monitor and assess the accuracy and quality of the different VaR forecasts techniques the Basel Accord (1996a) and the Amendment of Basel Accord (1996b) developed a statistical testing device that was denominated backtesting. The essence of backtesting is the out-of-sample comparison of actual trading results with model-generated risk measures. If the comparison uncovers sufficient differences between both figures the risk model should be subject to revision by the corresponding regulatory body. From Basel Committee's perspective (unconditional) backtesting consists on statistically testing whether the observed percentage of out-of-sample returns or P&L that are less than or equal to the forecasted VaR is consistent with the VaR level $100\alpha\%$, usually 99%.

An important limitation of the standard backtesting techniques is the assumption of the parameter θ_0 being known. In practice, however, the parameter θ_0 is unknown and must be estimated from the sample at time t by an estimator, say $\widehat{\theta}_t$. The standard approach in the literature consists on performing relevant inferences replacing θ_0 by the estimator $\widehat{\theta}_t$ in the standard backtesting procedures. We stress in this article that this method of forecast evaluation can lead to invalid inferences in backtesting procedures, which in turn may imply suboptimal levels of idle capital on the bank, that is, higher or lower levels than those actually required by the Basel Accord. We do so by showing that the introduction of $\widehat{\theta}_t$, *i.e.* uncertainty about θ_0 coming from the data, adds an additional term in the unconditional and independence backtesting procedures that must be taken into account to construct valid inferences in out-of-sample VaR forecasts evaluations.

Some of the earliest work on estimation risk in VaR measures is due to Jorion (2000). Our methodology, however, for the out-of-sample analysis builds on West (1996), and also McCracken (2000), adapted to our case of estimation of a quantile. These authors also acknowledge the presence of uncertainty due to parameter estimation in out-of-sample forecast inference, but they do not consider the problem we deal with here. In addition, we consider an asymptotic theory based on martingale methods, different from the asymptotic theory advocated by these authors based on mixing conditions. The purpose of the present paper is then, first to quantify the estimation risk in the most popular backtesting procedures in out-of-sample environments, and second, to propose a correction of these methods to make them free of estimation risk.

The rest of the paper is structured as follows. Section 2 introduces the forecast environments and studies the effects of estimation risk in unconditional and independence backtesting. Section 3 studies both backtesting methods for the popular family of GARCH models, and illustrates via Monte Carlo experiments with different data generating processes our theoretical findings in finite samples. Section 4 contains an application of our procedures to quantify the implications on capital requirements of correcting the critical values of the standard backtesting tests for the $S\&P500$ Index tracking the US equity market. Finally, Section 5 concludes. Mathematical proofs are gathered into Appendix A and some figures into Appendix B. Finally, we should mention that equivalent results to those of this paper but for in-sample inferences can be found in Escanciano and Olmo (2007).

2 Backtesting Techniques Robust to Estimation Risk

2.1 Forecast Evaluation Problem

From (1), a parametric VaR model $m_\alpha(W_{t-1},\theta_0)$ is correctly specified if and only if

$$E[I_{t,\alpha}(\theta_0) \mid W_{t-1}] = \alpha \text{ a.s. for some } \theta_0 \in \Theta, \tag{2}$$

where $I_{t,\alpha}(\theta) := 1(Y_t \leq m_\alpha(W_{t-1},\theta))$, $\theta \in \Theta$, and $1(A)$ is the indicator function, i.e. $1(A) = 1$ if the event A occurs and 0 otherwise. Most of the existing inference procedures are, however, based on testing *some* of the implications of condition (2) rather than the condition itself. For instance, Engle and Manganelli (2004) used the classical augmented regression argument for testing a version of (2). This consists on regressing $I_{t,\alpha}(\theta_0)-\alpha$ against its lagged values and other variables included in W_{t-1}, and testing whether these variables are significant in the regression. But the most popular implication explored is given in Christoffersen (1998),

$$E[I_{t,\alpha}(\theta_0) \mid \widetilde{I}_{t-1,\alpha}(\theta_0)] = \alpha, \text{ a.s. for some } \theta_0 \in \Theta, \tag{3}$$

where $\widetilde{I}_{t-1,\alpha}(\theta_0) := (I_{t-1,\alpha}(\theta_0), I_{t-2,\alpha}(\theta_0)...)'$. It is important to stress that (3) is a necessary but not sufficient condition of (2). This has important consequences in terms of the power performance of backtesting procedures. The popularity of condition (3) is mostly due to the discrete character and ease of interpretation of the variables $\{I_{t,\alpha}(\theta_0)\}$, which are the so-called *hits* or *exceedances*. In particular, the discreteness of the exceedances implies that condition (3) is equivalent to

$$\{I_{t,\alpha}(\theta_0)\} \text{ are } iid \text{ Ber}(\alpha) \text{ random variables } (r.v.) \text{ for some } \theta_0 \in \Theta, \tag{4}$$

where Ber(α) stands for a Bernoulli *r.v.* with parameter α. In the VaR literature, the satisfaction of condition (4) has been taken as the criteria for the out-of-sample evaluation of VaR forecasts, leading to the so-called unconditional backtesting (i.e. tests for $E[I_{t,\alpha}(\theta_0)] = \alpha$) and tests of independence (i.e. tests for $\{I_{t,\alpha}(\theta_0)\}$ being *iid*).

Backtesting techniques check for (4) in a forecast environment that we describe as follows. We assume a given sample $\{Y_t, Z_t'\}_{t=1}^n$ of size $n \geq 1$ that is used to evaluate the VaR forecasts. For simplicity we only consider one-step-ahead predictions, generalizations to other forecast horizons are straightforward (as long as we use non-overlapping intervals). As is standard in the forecasting literature we assume that the first R observations in the sample are used to estimate the parameters in the first forecast and that there are $P = n - R$ predictions to be evaluated. That is, the first VaR forecast $VaR_{R+1,1}(\widehat{\theta}_R) = m_\alpha(W_R, \widehat{\theta}_R)$, is based upon an estimator using the first R observations. Further forecasts, $VaR_{t+1,1}(\widehat{\theta}_t) = m_\alpha(W_t, \widehat{\theta}_t)$ are constructed with parameter estimators using observations $s = 1, ..., t$, with $R \leq t \leq n-1$.

We separately discuss the two backtesting problems, the unconditional and the independence hypotheses, under the aforementioned forecast environments.

2.2 Unconditional Backtesting

The most popular unconditional backtesting technique was proposed by Kupiec (1995), see also Christoffersen (1998), based on the absolute value of the standardized sample mean

$$K_P \equiv K(P,R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} (I_{t,\alpha}(\theta_0) - \alpha). \tag{5}$$

Under appropriate regularity conditions, including (4), $(\alpha(1-\alpha))^{-1/2} K_P$ converges to a standard normal r.v. The unconditional hypothesis $E[I_{t,\alpha}(\theta_0)] = \alpha$ is then tested using the critical values from the standard normal distribution. In fact, this test is optimal if θ_0 is known. In practice, however, the parameter θ_0 is not known and the relevant test statistic becomes

$$S_P \equiv S(P,R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} (I_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha),$$

with $\widehat{\theta}_t$ satisfying certain regularity conditions (cf. A4 below).

A common approach in the empirical and theoretical literature on risk management is to carry out inferences for S_P as if it were K_P, taking the same normal critical values to evaluate the forecast performance. The main message of our paper is that such inference procedures may be misleading under very general circumstances. We show that the estimation of parameters $\widehat{\theta}_{t-1}$ introduces asymptotically an extra term in the, still normal, limiting distribution, changing the resulting asymptotic variance of S_P.

As expected, one of the main determinants of the new asymptotic variance is the forecasting scheme used to create the forecasts. For the sake of completeness, and following *e.g.* West (1996) and McCraken (2000), we discuss three different forecasting schemes, namely, the recursive, rolling and fixed forecasting schemes. They differ in how the parameter θ_0 is estimated. In the recursive scheme, the estimator $\widehat{\theta}_t$ is computed with all the sample available up to time t. In the rolling scheme only the last R values of the series are used to estimate $\widehat{\theta}_t$, that is, $\widehat{\theta}_t$ is constructed from the sample $s = t - R + 1, ..., t$. Finally, in the fixed scheme the parameter is not updated when new observations become available, i.e., $\widehat{\theta}_t = \widehat{\theta}_R$, for all t, $R \le t \le n$.

The next theorem quantifies the effect of the estimation risk in S_P in the three forecasting schemes considered. In order to see this, we need some notation and assumptions. Define the family of conditional distributions $F_x(y) := P(Y_t \leq y \mid W_{t-1} = x)$, and let $f_x(y)$ be the associated conditional densities.

Assumption A1: $\{Y_t, Z_t'\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic.

Assumption A2: The family of distributions functions $\{F_x, x \in \mathbb{R}^\infty\}$ has Lebesgue densities $\{f_x, x \in \mathbb{R}^\infty\}$ that are uniformly bounded $\sup_{x\in\mathbb{R}^\infty, y\in\mathbb{R}} |f_x(y)| \leq C$, and equicontinuous: for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_{x\in\mathbb{R}^\infty, |y-z|\leq\delta} |f_x(y) - f_x(z)| \leq \epsilon.$$

Assumption A3: The model $m_\alpha(W_{t-1}, \theta)$ is continuously differentiable in θ (a.s.) with derivative $g_\alpha(W_{t-1}, \theta)$ such that $E\left[\sup_{\theta\in\Theta_0} |g_\alpha(W_{t-1}, \theta)|^2\right] < C$, for a neighborhood Θ_0 of θ_0.

Assumption A4: The parameter space Θ is compact in $\mathbb{R}^p$. The true parameter θ_0 belongs to the interior of Θ. The estimator $\widehat{\theta}_t$ satisfies the asymptotic expansion $\widehat{\theta}_t - \theta_0 = H(t) + o_P(1)$, where $H(t)$ is a $p\times 1$ vector such that $H(t) = t^{-1}\sum_{s=1}^{t} l(Y_s, W_{s-1}, \theta_0)$, $R^{-1}\sum_{s=t-R+1}^{t} l(Y_s, W_{s-1}, \theta_0)$ and $R^{-1}\sum_{s=1}^{R} l(Y_s, W_{s-1}, \theta_0)$ for the recursive, rolling and fixed schemes, respectively. We assume that $E[l(Y_t, W_{t-1}, \theta_0) \mid W_{t-1}] = 0$ a.s. and $V := E[l(Y_t, W_{t-1}, \theta_0)l'(Y_t, W_{t-1}, \theta_0)]$ exists and is positive definite. Moreover, $l(Y_t, W_{t-1}, \theta)$ is continuous (a.s.) in θ in Θ_0 and $E\left[\sup_{\theta\in\Theta_0} |l(Y_t, W_{t-1}, \theta)|^2\right] \leq C$, where Θ_0 is a small neighborhood around θ_0.

Assumption A5: $R, P \to \infty$ as $n \to \infty$, and $\lim_{n\to\infty} P/R = \pi$, $0 \leq \pi < \infty$.

Assumption A1 is made here for simplicity in the exposition. Our results are also valid for some non-stationary and non-ergodic sequences, see Escanciano (2007a) for details. Assumption A2 is required as in Koul and Stute (1999). Assumption A3 is classical in inference on nonlinear models. Assumption A4 is satisfied for most estimators considered in the literature, including maximum likelihood and generalized method of moments estimators. Assumption A5 is assumed in West (1996) and McCraken (2000), see e.g. the discussion in McCraken (2000, p. 200). With these assumptions in place we are in position to establish the first important result of the paper.

Theorem 1: *Under Assumptions A1-A5,*

$$\begin{aligned} S_P &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[I_{t,\alpha}(\theta_0) - F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) \right] \\ &\quad \underbrace{+ E\left[g'_\alpha(W_{t-1}, \theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) \right] \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} H(t-1)}_{\text{Estimation Risk}} \\ &\quad \underbrace{+ \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) - \alpha \right]}_{\text{Model Risk}} + o_P(1). \end{aligned}$$

Theorem 1 quantifies both *estimation risk* and *model risk* in the unconditional coverage backtest introduced before. It also has several important implications for our testing problems. Note that Theorem 1 does not assume either the correct specification of the parametric VaR model nor *iid* exceedances. Also, Theorem 1 does not require any mixing condition in contrast to related papers dealing with estimation risk in evaluation of forecasts, e.g. West (1996) and McCraken (2000). These mixing assumptions are difficult to verify in practice and are not satisfied for some simple models. The proof of Theorem 1 is based on applications of the modern theory of empirical processes under martingale conditions, see Delgado and Escanciano (2006) and references therein.

Under correct specification of the parametric VaR model, *i.e.* $F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) = \alpha$ *a.s.*, model risk vanishes. In contrast, under misspecification, even if $E\left[F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))\right] = E[I_{t,\alpha}(\theta_0)] = \alpha$ holds, model risk does not vanish and has a non-negligible effect on the unconditional test. In this case, unconditional backtesting tests based on S_P are inconsistent for testing (2). On the other hand if $E\left[F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))\right] \neq \alpha$, under some regularity conditions, Theorem 1 yields that

$$\frac{1}{P} \sum_{t=R+1}^{n} [I_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha] \xrightarrow{P} E[F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) - \alpha] \neq 0,$$

and the unconditional test based on S_P is consistent as a specification test of the parametric VaR model. In this paper, however, we do not make a thorough study of model risk as our main focus is on the estimation risk, thus we shall assume hereafter that $F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) = \alpha$ *a.s.* whenever is necessary.

The first term in the expansion of Theorem 1 has martingale difference sequence (*mds*) summands, so applying a Martingale Central Limit Theorem, see *e.g.* Hall and Heyde (1980), this term converges to a Gaussian distribution. The second term is the estimation risk. The analysis of this part has to be made on a case-by-case basis, *i.e.*, for a particular estimator $\widehat{\theta}_t$, model, true data generating process (DGP) and forecast scheme. To illustrate our theoretical findings we shall study in Section 4 the widely used GARCH(1,1) models with a fixed forecasting scheme.

To simplify notation we write $A := E\left[g'_\alpha(W_{t-1},\theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\right]$ in the expression for the estimation risk. Next corollary provides the necessary corrections to carry out valid asymptotic inference for unconditional backtests free of estimation risk.

Corollary 1: *Under Assumptions A1-A5 and (2),* $S_P \xrightarrow{d} N(0,\sigma_u^2)$, *where* $\sigma_u^2 = \alpha(1-\alpha) + 2\lambda_{hl} A\rho + \lambda_{ll} AVA'$, *with* $\rho = E[(I_{t,\alpha}(\theta_0) - \alpha)\, l(Y_t, W_{t-1}, \theta_0)]$ *and where*

Scheme	λ_{hl}	λ_{ll}
Recursive	$1 - \pi^{-1}\ln(1+\pi)$	$2\left[1 - \pi^{-1}\ln(1+\pi)\right]$
Rolling, $\pi \leq 1$	$\pi/2$	$\pi - \pi^2/3$
Rolling, $1 < \pi < \infty$	$1 - (2\pi)^{-1}$	$1 - (3\pi)^{-1}$
Fixed	0	π

(6)

From our Corollary 1 we obtain that the (asymptotic) size-distortion at $\upsilon\%$ of the two-sided (uncorrected) unconditional backtesting test is, as $n \to \infty$,

$$P\left(\frac{1}{\alpha^{1/2}(1-\alpha)^{1/2}}\left|S_P\right| > z_{\upsilon/2}\right) - \upsilon \;=\; P\left(\frac{1}{\sigma_u}\left|S_P\right| > \frac{\alpha^{1/2}(1-\alpha)^{1/2}}{\sigma_u} z_{\upsilon/2}\right) - \upsilon \;\to\; 2\left(1 - \Phi\left(\frac{\alpha^{1/2}(1-\alpha)^{1/2}}{\sigma_u} z_{\upsilon/2}\right)\right) - \upsilon, \quad (7)$$

where Φ is the cdf of the standard normal *r.v.* and $z_{\upsilon/2}$ is such that $\Phi\left(z_{\upsilon/2}\right) = 1 - \upsilon/2$.

In general, σ_u^2 may be greater, equal or smaller than $\alpha(1-\alpha)$. For instance, the presence of estimation risk is asymptotically irrelevant if $2\lambda_{hl}A\rho + \lambda_{ll}AVA' = 0$, that is, the variance induced by error in estimation of θ_0 is offset by the covariance between such terms and terms that would be present even if the parameter were known. Note that if R is arbitrarily large relative to P, i.e. $\pi = 0$, there is "infinite" information contained in $\widehat{\theta}_{t-1}$ about θ_0 relative to S_P, and as a result the estimation risk asymptotically vanishes. In practice, however, the backtesting experiments usually consider P of similar size of R. For specific cases we can be more precise, for instance, for the fixed scheme, $\sigma_u^2 > \alpha(1-\alpha)$, provided $A \neq 0$ and $\pi > 0$. In this case, traditional unconditional backtesting techniques will be oversized since

$$2\left(1-\Phi\left(\frac{\alpha^{1/2}(1-\alpha)^{1/2}}{\sigma_u}z_{\upsilon/2}\right)\right)-\upsilon>0. \quad (8)$$

We now turn into the problem of estimating the asymptotic variance of S_P. The vector A can be consistently estimated by

$$\widehat{A}_\tau = -\frac{1}{P}\sum_{t=R+1}^{n}\frac{1}{\tau}\exp\left[\left(Y_t - m_\alpha(W_{t-1},\widehat{\theta}_{t-1})\right)/\tau\right] I_{t,\alpha}(\widehat{\theta}_{t-1})g'_\alpha(W_{t-1},\widehat{\theta}_{t-1}), \quad (9)$$

with $\tau \to 0$ as $n \to \infty$. This estimator is introduced in Giacomini and Komunjer (2005) for encompassing tests of different conditional quantile forecasts. Alternative nonparametric methods for estimating A can be found in *e.g.* Engle and Manganelli (2004), or in Li and Racine (2006) using kernel smoothers or local polynomials. For certain models, *e.g.* GARCH models, simpler estimators for A are available, see Section 3 and Appendix A in the working paper version of this article. Methods for estimating the variance-covariance matrix V are abundant in the literature, including bootstrap techniques. The parameters $\lambda_{hl} = \lambda_{hl}(\pi)$ and $\lambda_{ll} = \lambda_{ll}(\pi)$ in (6) depend on the forecasting scheme and are functions of π. Therefore, their natural estimators are $\widehat{\lambda}_{hl} = \lambda_{hl}(\widehat{\pi})$ and $\widehat{\lambda}_{ll} = \lambda_{ll}(\widehat{\pi})$, where the parameter π is approximated by $\widehat{\pi} = P/R$. Hence, the asymptotic variance σ_u^2 can be consistently estimated by

$$\widehat{\sigma}_u^2 := \alpha(1-\alpha) + 2\widehat{\lambda}_{hl}\widehat{A}_\tau\widehat{\rho} + \widehat{\lambda}_{ll}\widehat{A}_\tau\widehat{V}\widehat{A}'_\tau,$$

where

$$\widehat{\rho} = \frac{1}{P}\sum_{t=R+1}^{n}\left(I_{t,\alpha}(\widehat{\theta}_{t-1})-\alpha\right) l(Y_t, W_{t-1},\widehat{\theta}_{t-1}),$$

and

$$\widehat{V} = \frac{1}{P}\sum_{t=R+1}^{n} l(Y_t, W_{t-1},\widehat{\theta}_{t-1})l'(Y_t, W_{t-1},\widehat{\theta}_{t-1}),$$

are consistent estimators for ρ and V, respectively. Then, valid inference can be accomplished by the *corrected unconditional backtesting test* statistic

$$\widetilde{S}_P \equiv \widetilde{S}(P,R,\widehat{\theta}_{t-1}) = \frac{1}{\widehat{\sigma}_u\sqrt{P}}\sum_{t=R+1}^{n}(I_{t,\alpha}(\widehat{\theta}_{t-1})-\alpha),$$

which converges to a standard normal $r.v$ as shown in the next corollary.

COROLLARY 2: *Under Assumptions A1-A5, (2) and that* $\tau \to 0$ *as* $n \to \infty$, $\widetilde{S}_P \xrightarrow{d} N(0,1)$.

2.3 Independence and Joint Tests

This section is devoted to the hypothesis of independence, i.e.

$$\{I_{t,\alpha}(\theta_0)\}_{t=R+1}^{n} \text{ are } iid. \tag{10}$$

Christoffersen (1998) introduces in his seminal paper a likelihood ratio (LR) test for testing (10). This author embedded the sequence of hits $\{I_{t,\alpha}(\theta_0)\}_{t=R+1}^{n}$ in a first-order Markov model and construct a LR test within this family. Recently, more general tests for (10) have been based on the autocovariances

$$\xi_j = Cov(I_{t,\alpha}(\theta_0), I_{t-j,\alpha}(\theta_0)) \qquad j \geq 1, \tag{11}$$

at different lags j, which can be consistently estimated (under $E[I_{t,\alpha}(\theta_0)] = \alpha$) by

$$\xi_{P,j} = \frac{1}{P-j} \sum_{t=R+j+1}^{n} (I_{t,\alpha}(\theta_0) I_{t-j,\alpha}(\theta_0) - \alpha^2) \quad \text{for} \quad j \geq 1.$$

Other estimators for the autocovariance in (11) are also possible. Indeed, Berkowitz, Christoffersen and Pelletier (2006) discuss Portmanteau tests in the spirit of those proposed by Box and Pierce (1970) and Ljung and Box (1978) that make use of the sequence of sample autocovariances $\{\gamma_{P,j}\}$, where

$$\gamma_{P,j} = \frac{1}{P-j} \sum_{t=R+j+1}^{n} (I_{t,\alpha}(\theta_0) - \alpha)(I_{t-j,\alpha}(\theta_0) - \alpha) \qquad j \geq 1.$$

Berkowitz et al. (2006) propose the test statistic

$$LB(m) = P(P+2) \sum_{j=1}^{m} (P-j)^{-1} \left(\frac{\gamma_{P,j}}{\alpha(1-\alpha)} \right)^2. \tag{12}$$

These authors also explore spectral-based tests along the lines suggested in Durlauf (1991), taking into account all possible lags $j \geq 1$.

Notice that tests based on either $\{\xi_{P,j}\}$ or $\{\gamma_{P,j}\}$ are actually joint tests of the *iid* and the unconditional hypothesis, since they explicitly used the fact that $E[I_{t,\alpha}(\theta_0)] = \alpha$. A proper test of independence should be based instead on

$$\zeta_{P,j} = \frac{1}{P-j}\sum_{t=R+j+1}^{n} I_{t,\alpha}(\theta_0)I_{t-j,\alpha}(\theta_0) - \frac{1}{(P-j)^2}\left\{\sum_{t=R+j+1}^{n} I_{t,\alpha}(\theta_0)\right\}\left\{\sum_{t=R+j+1}^{n} I_{t-j,\alpha}(\theta_0)\right\}.$$

In practice, however, tests for (10) need to be based on estimates of the relevant parameters, such as

$$\widehat{\xi}_{P,j} = \frac{1}{P-j}\sum_{t=R+j+1}^{n}(I_{t,\alpha}(\widehat{\theta}_{t-1})I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - \alpha^2),$$

$$\widehat{\gamma}_{P,j} = \frac{1}{P-j}\sum_{t=R+j+1}^{n}(I_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha)(I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - \alpha)$$

or

$$\widehat{\zeta}_{P,j} = \frac{1}{P-j}\sum_{t=R+j+1}^{n} I_{t,\alpha}(\widehat{\theta}_{t-1})I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - \frac{1}{(P-j)^2}\left\{\sum_{t=R+j+1}^{n} I_{t,\alpha}(\widehat{\theta}_{t-1})\right\}\left\{\sum_{t=R+j+1}^{n} I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})\right\}.$$

Next theorem is the equivalent to Theorem 1 for the joint and independence backtesting tests. Define $B \equiv B_j := E[g'_\alpha(W_{t-1},\theta_0)f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\{I_{t-j,\alpha}(\theta_0)+\alpha\}]$ and $\eta \equiv \eta_j =: E[(I_{t,\alpha}(\theta_0)I_{t-j,\alpha}(\theta_0) - \alpha^2)l(Y_t, W_{t-1}, \theta_0)]$.

THEOREM 2: *Under Assumptions A1-A5 and (2), for any* $j \geq 1$,

(i) $\sqrt{P-j}(\widehat{\xi}_{P,j} - \xi_{P,j}) = B\frac{1}{\sqrt{P-j}}\sum_{t=R+j+1}^{n} H(t-j-1) + o_P(1)$.

(ii) $\sqrt{P-j}(\widehat{\gamma}_{P,j} - \gamma_{P,j}) = \{B - 2\alpha A\}\frac{1}{\sqrt{P-j}}\sum_{t=R+j+1}^{n} H(t-j-1) + o_P(1)$.

(iii) $\sqrt{P-j}(\widehat{\zeta}_{P,j} - \zeta_{P,j}) = \{B - 2\alpha A\}\frac{1}{\sqrt{P-j}}\sum_{t=R+j+1}^{n} H(t-j-1) + 2\alpha K_P + o_P(1)$.

A direct consequence of Theorem 2 is that tests based on $LB(m)$ with estimated parameters will be invalid. The necessary corrections can be straightforwardly obtained from a multivariate extension of our Theorem 2. Details are omitted to save space. In what follows we provide a correction for the joint test based on $\widehat{\xi}_{P,j}$ similar to that carried out for S_P in the unconditional case. Corrections for $\widehat{\gamma}_{P,j}$ and $\widehat{\zeta}_{P,j}$ are analogous and hence omitted. Note that if we want to construct an independence test robust to the unconditional assumption, α should be replaced by $E[I_{t,\alpha}(\theta_0)]$ in the limit distribution of $\widehat{\zeta}_{P,j}$. We remark that our Theorem 2 generalizes some results in Linton and Whang (2004). These authors established the in-sample asymptotic expansion of

$\widehat{\gamma}_{P,j}$ when no covariates are present in the VaR model, that is, when $m_\alpha(W_{t-1},\theta) \equiv \theta_\alpha$. Also, we note that Engle and Manganelli (2004) also studied a dynamic out-of-sample test that can be considered a joint test for (2). They only analyzed, however, the fixed forecasting scheme with $\pi = 0$, and found that there is not estimation risk. Their result is consistent with our findings in Theorem 2.

As with A, the vector B can be consistently estimated by $\widehat{B}_\tau$, where

$$\widehat{B}_\tau = -\frac{1}{P-j}\sum_{t=R+j+1}^{n}\frac{1}{\tau}\exp\left[\left(Y_t - m_\alpha(W_{t-1},\widehat{\theta}_{t-1})\right)/\tau\right] I_{t,\alpha}(\widehat{\theta}_{t-1})\{I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})+\alpha\}g'_\alpha(W_{t-1},\widehat{\theta}_{t-1}),$$

with $\tau \to 0$ as $n \to \infty$. Again, simpler estimators for B are available in some popular models, e.g. ARMA-GARCH models. Using the sample estimators as in Section 2 for $\lambda_{hl} = \lambda_{hl}(\pi)$, $\lambda_{ll} = \lambda_{ll}(\pi)$ and V, we propose the estimator

$$\widehat{\sigma}_c^2 := \alpha^2(1-\alpha)^2 + 2\widehat{\lambda}_{hl}\widehat{B}_\tau\widehat{\eta} + \widehat{\lambda}_{ll}\widehat{B}_\tau\widehat{V}\widehat{B}'_\tau,$$

where

$$\widehat{\eta} = \frac{1}{P-j}\sum_{t=R+j+1}^{n}(I_{t,\alpha}(\widehat{\theta}_{t-1})I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - \alpha^2)l(Y_t, W_{t-1}, \widehat{\theta}_{t-1}).$$

Then, valid inference can be accomplished by the *corrected joint backtesting test* statistic

$$\widetilde{\xi}_{P,j} \equiv \widetilde{\xi}(P,R,\widehat{\theta}_{t-1}) = \frac{1}{\widehat{\sigma}_c\sqrt{P-j}}\sum_{t=R+j+1}^{n}(I_{t,\alpha}(\widehat{\theta}_{t-1})I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - \alpha^2),$$

which converges to a standard normal $r.v$ as shown in the next corollary.

Corollary 3: *Under Assumptions A1-A5, (2) and that $\tau \to 0$ as $n \to \infty$, for each $j \geq 1$*

$$\sqrt{P-j}\widehat{\xi}_{P,j} \xrightarrow{d} N(0,\sigma_c^2),$$

where $\sigma_c^2 = \alpha^2(1-\alpha)^2 + 2\lambda_{hl}B\eta + \lambda_{ll}BVB'$, with λ_{hl} and λ_{ll} as in (6). Therefore,

$$\widetilde{\xi}_{P,j} \xrightarrow{d} N(0,1).$$

3 Simulation exercise on location-scale models

In this section we confine ourselves to consider the parametric VaR model derived from a location-scale model. This parametric approach has been the most popular in attempting to describe the dynamics of the VaR measure (cf. Berkowitz and O'Brien, 2002). These models are defined as

$$Y_t = \mu(W_{t-1}, \beta_0) + \sigma(W_{t-1}, \beta_0)\varepsilon_t, \tag{13}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are specifications for the conditional mean and standard deviation of Y_t given W_{t-1}, respectively, and ε_t are the standardized innovations which are usually assumed to be *iid,* and independent of W_{t-1}. Under such assumptions the α-th conditional VaR is given by

$$m_\alpha(W_{t-1}, \theta_0) = \mu(W_{t-1}, \beta_0) + \sigma(W_{t-1}, \beta_0)F_\varepsilon^{-1}(\alpha), \tag{14}$$

where $F_\varepsilon^{-1}(\alpha)$ denotes the univariate quantile function of ε_t and the nuisance parameter is $\theta_0 = (\beta_0, F_\varepsilon^{-1}(\alpha))$. Amongst the most common models for $\mu(\cdot)$ and $\sigma(\cdot)$ are the ARMA and GARCH models, respectively, under different distributional assumptions on the error term. The vector of parameters β_0 is usually estimated by the Quasi-Maximum Likelihood Estimator (*QMLE*). See Li, Ling and McAleer (2002) for a review of estimators for β_0. The second component of θ_0, $F_\varepsilon^{-1}(\alpha)$, is assumed to be either known (*e.g.* Gaussian), unknown up to a finite-dimensional unknown parameter (*e.g.* Student-t distributed with unknown degrees of freedom), or unknown up to an infinite-dimensional unknown parameter (for instance, semiparametric estimators based on extreme value theory. These have been extensively used, see *e.g.* Chan, Deng, Peng and Xia (2006) for a recent reference.) See Koenker and Zhao (1996) for alternative quantile estimators in ARCH models. Some of these methods are reviewed in Kuester et al. (2006).

For these models our Theorem 1 allows us to quantify estimation risk for the unconditional test. For simplicity in the exposition and to save space we only consider throughout this section the fixed forecasting scheme. The aim of this section is not to make a thorough finite-sample study of the estimation risk in location-scale models but just to illustrate our findings in a realistic situation. Thus, for model (14) with a fixed forecasting scheme the estimation risk term for the unconditional test takes the form

$$\sqrt{\pi}\sqrt{R}(F_{\varepsilon,R}^{-1}(\alpha) - F_{\varepsilon}^{-1}(\alpha))f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha)) + \sqrt{\pi}\sqrt{R}(\widehat{\beta}_R - \beta_0)'A, \tag{15}$$

where $F_{\varepsilon,R}^{-1}(\alpha)$ is an α-quantile estimator of the innovation distribution, and

$$A = A(\alpha, \beta_0) := f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha))E\left[a_{1,t}(\beta_0)\right] + f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha))F_{\varepsilon}^{-1}(\alpha)E\left[a_{2,t}(\beta_0)\right], \tag{16}$$

with

$$a_{1,t}(\beta) = \dot{\mu}_t(\beta)/\sigma(W_{t-1}, \beta), \qquad a_{2,t}(\beta) = \dot{\sigma}_t(\beta)/\sigma(W_{t-1}, \beta),$$

and where $\dot{\mu}_t(\beta) = \partial\mu(W_{t-1}, \beta)/\partial\beta$ and $\dot{\sigma}_t(\beta) = \partial\sigma(W_{t-1}, \beta)/\partial\beta$. There are two sources of estimation risk in this model, one from estimating $F_{\varepsilon}^{-1}(\alpha)$ and other from estimating β_0.

To illustrate the effect of estimation risk in backtesting procedures we proceed to analyze one of the most common processes for modelling financial returns: the GARCH(1,1) model with Student-t innovations. This model is defined as

$$Y_t = \sigma(W_{t-1}, \beta_0)\varepsilon_t, \qquad \sigma^2(W_{t-1}, \beta_0) = \eta_{00} + \eta_{10}Y_{t-1}^2 + \eta_{20}\sigma^2(W_{t-2}, \beta_0),$$

where $\{\varepsilon_t\}$ are *iid* t_ν standardized disturbances (i.e. $\varepsilon_t = (\sqrt{(\nu-2)/\nu})v_t$, with v_t distributed as a Student-t with ν degrees of freedom), the true parameters are $\beta_0 = (\eta_{00}, \eta_{10}, \eta_{20}) \in \Theta$, with $\Theta \subset \{(\eta_0, \eta_1, \eta_2) \in \mathbb{R}^3 : \eta_0 > 0, \eta_1 \geq 0, \eta_2 \geq 0, \eta_1 + \eta_2 < 1\}$.

The estimation risk in this example in which the error distribution is Student-t with a discrete number of degrees of freedom only depends on the estimation error stemming from the uncertainty of estimating the scale model (see Hannan and Quinn, 1979, p. 191, for general results on estimation of discrete-valued parameters). Thus, given that there is no estimation risk coming from the error the first term in (15) disappears and the estimation risk for the unconditional test boils down to $\sqrt{\pi}\sqrt{R}(\widehat{\beta}_R - \beta_0)'A$, where $A = f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha))F_{\varepsilon}^{-1}(\alpha)E\left[a_{2,t}(\beta_0)\right]$.

In order to shed some light on the relation between α and the magnitude of the estimation risk we plot in figure 6.1 an average of the 5% size-distortion effect (cf. 7)

$$d\left(\alpha\right) = 2\left(1 - \Phi\left(\frac{\alpha^{1/2}(1-\alpha)^{1/2}}{\widehat{\sigma}_u(\alpha)}z_{0.025}\right)\right) - 0.05$$

for 500 Monte-Carlo simulations and the corresponding confidence interval at 5% as a function of α, and where $\widehat{\sigma}_u^2(\alpha) = \alpha(1-\alpha) + \widehat{A}'\widehat{V}\widehat{A}$, $\widehat{A} = f_\varepsilon(F_\varepsilon^{-1}(\alpha))F_\varepsilon^{-1}(\alpha)\sum_{t=R+1}^{n} a_{2,t}(\widehat{\beta}_R)$, $\widehat{V}$ given by

$$\widehat{V} = 2\left(\kappa_{\widehat{\nu}} - 1\right)\left[P^{-1}\sum_{t=R+1}^{n}\frac{1}{\sigma_t^4(\widehat{\beta}_R)}\frac{\partial\sigma_t^2(\widehat{\beta}_R)}{\partial\beta}\frac{\partial\sigma_t^2(\widehat{\beta}_R)}{\partial\beta'}\right]^{-1}, \tag{17}$$

with $\kappa_{\widehat{\nu}}$ the kurtosis of the standardized Student-$t_{\widehat{\nu}}$ error, and $\widehat{\beta}_R$ the *QMLE* of β_0 using the first R observations. The true *DGP* uses parameter values $\beta_0' = (\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$ and innovations distributed as Student-t with $\nu = 30$ degrees of freedom.

We observe from these plots that the size distortion calculated in (8) for the GARCH(1,1) process increases with α up to a 0.05 coverage probability, attaining the maximum distortion of 9% for $R = 250$ and $P = 500$, and then slightly decreases again. This size distortion is more pronounced for values of the in-sample size R small compared to the out-of-sample size P. This is confirmed in the Monte Carlo simulation experiment.

Similarly, the estimation risk for the joint test for the GARCH(1,1) model is $\sqrt{\pi}\sqrt{R}(\widehat{\beta}_R - \beta_0)'B$, where $B = f_\varepsilon(F_\varepsilon^{-1}(\alpha))F_\varepsilon^{-1}(\alpha)E\left[a_{2,t}(\beta_0)\{I_{t-j,\alpha}(\theta_0) + \alpha\}\right]$. Further details for these expressions are found in the working paper version of this article. Figure 6.2 illustrates the estimation risk effect for the independence test by plotting the size distortion in (8) where $\widehat{\xi}_{P,1}$ replaces S_P in (7), and with $\widehat{\sigma}_c^2 = (\alpha(1-\alpha))^2 + \widehat{B}'\widehat{V}\widehat{B}$ and $\widehat{B} = f_\varepsilon(F_\varepsilon^{-1}(\alpha))F_\varepsilon^{-1}(\alpha)\left\{P^{-1}\sum_{t=R+1}^{n}(a_{2,t}(\widehat{\beta}_R)\{I_{t-j,\alpha}(\widehat{\beta}_R) + \alpha\}\right\}$. The plot also reports the corresponding confidence intervals at 5% for 500 Monte-Carlo simulations.

In contrast to the unconditional test the distortion in size between the uncorrected and corrected method strictly increases with α, being this effect more important for values of R small compared to P. Overall, we observe a larger size distortions for the unconditional test than for the joint. This is further discussed in the following section.

3.1 Monte Carlo Simulation Experiment

The asymptotic results of preceding sections need only be appropriate for large in-sample and out-of-sample sizes. It is not clear how well the asymptotic approximation will perform in small and moderate sample sizes. To examine this problem we carry out some Monte Carlo experiments. For the sake of space and simplicity of computation we just report results for the fixed forecasting scheme.

The aim of the first study is to compare the size performance of $\widetilde{S}_P$ and $\alpha^{-1/2}(1-\alpha)^{-1/2}S_P$, *i.e.*, analyze the impact of estimation risk in unconditional tests for GARCH(1,1) models. In order to do this we shall use for both test statistics the critical values of a standard Gaussian distribution. The innovation process ε_t, is assumed to be distributed as a Student-t distribution with $\nu = 30$ degrees of freedom (for $\nu = 5$ refer to the working paper version.) The parameters of the GARCH(1,1) process are chosen to reflect standard values found in real time series of financial returns. We consider $\beta_0' = (\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$. The Value at Risk of these models is calculated at VaR levels 1%, 5% and 10%.

Figure 6.3 in Appendix B describes the surfaces corresponding to the empirical 5% size for different in-sample and out-of-sample lengths (R, P) when ν is known. The simulation exercise consists on generating data from the GARCH process described above; in a second stage the parameters of the model are estimated by $QMLE$ using the first R observations and the corresponding VaR_α model is computed for the remaining P out-of-sample observations. Finally we compute S_P and $\widetilde{S}_P$ for each Monte Carlo iteration.

We draw four main conclusions from this battery of plots and other unreported simulations. First, the corrected estimator $\widetilde{S}_P$ clearly outperforms the uncorrected method based on S_P since the simulated sizes are much closer to the nominal value 5% across all experiments, specially for large values of the parameter $\widehat{\pi} = P/R$. Second, as expected, the sizes of the tests are very sensitive to the choice of in-sample and out-of-sample window lengths. Thus, as $\widehat{\pi}$ decreases the sizes of the uncorrected test are closer to the nominal size, as predicted by the theory. On the other hand, as $\widehat{\pi}$ increases these estimates worsen off. Third, unreported simulations for the Student-t with 5 degrees of freedom reveal the importance of the thickness of the tails in this framework since the size of both methods is distorted for VaR levels at $\alpha = 0.05$ and $\alpha = 0.1$. Again, for the uncorrected test statistic S_P the distortions are much more significant. Finally, we should also mention that the approximation by the asymptotic theory of the finite sample distributions of both test statistics, S_P and $\widetilde{S}_P$, for low α levels is not accurate, especially in the tails of the distribution which is the important part for testing. This problem is intrinsic to VaR inferences at low quantile levels and not to the existence of estimation risk, and raises an important issue for backtesting at small VaR levels such as $\alpha = 0.01$. Arguably, a different asymptotic theory based on $\alpha \to 0$ as $n \to \infty$ may help to this end, see Leadbetter, Lindgren and Rootzen (1983), chapter 2, for a study of the asymptotic distribution of the k^{th} largest maximum of a random sample when k is fixed. This important problem is beyond the scope of this paper and deserves further research.

Figure 6.4 illustrates the estimation risk effects on the joint test statistic defined by $\widetilde{C}_{P,j} = \widetilde{\xi}^2_{P,j}$, which by Corollary 3 is distributed as a χ^2_1 distribution, and on the uncorrected version of the test $C_{P,j} = (P-j)\left(\frac{\widehat{\xi}^2_{P,j}}{\alpha^2(1-\alpha)^2}\right)$. In the simulations we consider the case $j = 1$, which is the most used in empirical work (cf. Christoffersen (1998)). For $\alpha = 0.01$, the estimated sizes are not very close to the nominal values for both methods and large values of $\widehat{\pi}$. Further, it seems that there are no significant differences between uncorrected and corrected tests for small values of α. The sizes of both tests improve for $\alpha = 0.05$, more importantly, the correction given by $\widetilde{C}_{P,1}$ yields size values closer to the nominal level. These results are made clear for $\alpha = 0.1$; in this case the effect of estimating the parameters produces distortions in the uncorrected joint test. These effects are largely corrected by using $\widetilde{C}_{P,1}$. These effects are similar for the Student-t distribution error with $\nu = 5$. Note however that the plots corresponding to this scenario are not reported to save space. Finally, other unreported simulations for $LB(m)$ in (12) and Durlauf-type tests for testing serial dependence at 5 lags show a substantial impact of the estimation risk in the size of the uncorrected test.

Although the main aim of this paper is to show that the current backtesting techniques applied to composite hypotheses may be over- or undersized in general cases, we also present a simple Monte Carlo experiment to compare the empirical powers of $\widetilde{S}_P$ and S_P for the unconditional backtesting, and $C_{P,1}$ and $\widetilde{C}_{P,1}$ for the joint test, in rejecting the alternatives (to the null of GARCH(1,1) model) given by the following DGPs:

1. GARCH-M model: $Y_t = 2.5\sigma_t^2 + u_t$, $u_t = \sigma_t^2\varepsilon_t$, $\sigma_t^2 = 0.001 + 0.29u_{t-1}^2 + 0.70\sigma_{t-1}^2$.

2. TAR model: $Y_t = a_t Y_{t-1} + \varepsilon_t$, $a_t = 0.7 \cdot 1(\varepsilon_{t-1} < -0.5) - 0.7 \cdot 1(\varepsilon_{t-1} > 0.5)$.

3. EGARCH(1,1) model: $Y_t = h_t\varepsilon_t$, $\ln h_t^2 = 0.01 + 0.9 \ln h_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - (2/\pi)^{1/2}) - 0.8\varepsilon_{t-1}$.

4. Stochastic Volatility (SV) model: $Y_t = h_t\varepsilon_t$, $\ln h_t^2 = 0.1 + 0.78 \ln h_{t-1}^2 + v_t$, with $v_t \sim N(0,1)$.

5. Bilinear model (BIL): $Y_t = 0.5\varepsilon_{t-1}Y_{t-1} + \varepsilon_t$.

6. Non-Linear Moving Average model (NLMA): $Y_t = 0.5\varepsilon_{t-1}^2 + \varepsilon_t$.

For some discussion on these models and their parameter values see Martinez and Olmo (2007) and Escanciano (2008). The error ε_t in these models is assumed to follow a Student-t distribution with $\nu = 30$. In Tables 3.1 and 3.2 we report the rejection probabilities at 10%, 5% and 1% significance level for the different tests. The empirical power of these tests is size-corrected, using empirical critical values computed under the GARCH(1,1) model with parameter values $(\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$. The in-sample and out-of-sample sizes considered are $R = 500$ and $P = 250$ and 500.

$\alpha = 0.01$		GARCH-M			TAR			EGARCH		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$P = 250$	S_P	0.000	0.000	0.000	0.176	0.176	0.026	0.622	0.622	0.280
	$\widetilde{S}_P$	0.208	0.000	0.000	0.342	0.176	0.056	0.700	0.620	0.396
	$C_{P,1}$	1.000	1.000	0.000	1.000	1.000	0.001	1.000	1.000	0.100
	$\widetilde{C}_{P,1}$	0.539	0.519	0.001	0.881	0.854	0.056	0.582	0.525	0.261
$P = 500$	S_P	0.193	0.193	0.000	0.234	0.234	0.050	0.806	0.806	0.462
	$\widetilde{S}_P$	0.384	0.131	0.000	0.324	0.230	0.080	0.842	0.772	0.558
	$C_{P,1}$	1.000	0.006	0.006	1.000	0.148	0.148	1.000	0.587	0.587
	$\widetilde{C}_{P,1}$	0.548	0.006	0.003	0.924	0.147	0.117	0.729	0.586	0.411
$\alpha = 0.05$		GARCH-M			TAR			EGARCH		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$P = 250$	S_P	0.805	0.501	0.071	0.748	0.578	0.320	0.348	0.220	0.094
	$\widetilde{S}_P$	0.778	0.487	0.064	0.748	0.578	0.320	0.344	0.220	0.094
	$C_{P,1}$	0.007	0.007	0.000	0.831	0.831	0.499	0.660	0.660	0.234
	$\widetilde{C}_{P,1}$	0.004	0.003	0.000	0.831	0.822	0.498	0.612	0.497	0.214
$P = 500$	S_P	0.954	0.823	0.304	0.842	0.750	0.484	0.422	0.262	0.102
	$\widetilde{S}_P$	0.938	0.869	0.508	0.842	0.770	0.606	0.406	0.300	0.154
	$C_{P,1}$	0.006	0.002	0.000	0.926	0.837	0.619	0.786	0.658	0.365
	$\widetilde{C}_{P,1}$	0.005	0.001	0.000	0.926	0.837	0.730	0.756	0.658	0.490

Table 3.1. *Empirical power of unconditional S_P and $\widetilde{S}_P$ backtesting tests, and $\widehat{C}_{P,1}$ and $\widetilde{C}_{P,1}$ independence tests for the fixed forecasting scheme. The VaR is computed at $\alpha = 0.01$ and*

$\alpha = 0.05$. *The error term* ε_t *is assumed to follow a Student-t with* $\nu = 30$. 1000 *Monte Carlo simulations.* $R = 500$, $P = 250$, 500. *Models GARCH-M, TAR and EGARCH.*

$\alpha = 0.01$		SV			BIL			NLMA		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$P = 250$	S_P	0.320	0.320	0.066	0.004	0.004	0.000	0.002	0.002	0.000
	$\widetilde{S}_P$	0.420	0.320	0.116	0.290	0.004	0.002	0.392	0.002	0.000
	$C_{P,1}$	1.000	1.000	0.002	1.000	1.000	0.000	1.000	1.000	0.000
	$\widetilde{C}_{P,1}$	0.272	0.105	0.026	0.957	0.947	0.000	0.960	0.939	0.000
$P = 500$	S_P	0.510	0.510	0.144	0.084	0.084	0.000	0.164	0.164	0.000
	$\widetilde{S}_P$	0.594	0.508	0.220	0.296	0.084	0.000	0.432	0.164	0.000
	$C_{P,1}$	1.000	0.125	0.125	1.000	0.000	0.000	1.000	0.000	0.000
	$\widetilde{C}_{P,1}$	0.268	0.124	0.065	0.978	0.000	0.000	0.980	0.000	0.000
$\alpha = 0.05$		SV			BIL			NLMA		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$P = 250$	S_P	0.134	0.040	0.002	0.526	0.226	0.016	0.694	0.364	0.034
	$\widetilde{S}_P$	0.122	0.040	0.002	0.526	0.226	0.016	0.694	0.364	0.034
	$C_{P,1}$	0.067	0.067	0.002	0.000	0.000	0.000	0.000	0.000	0.000
	$\widetilde{C}_{P,1}$	0.046	0.010	0.002	0.000	0.000	0.000	0.000	0.000	0.000
$P = 500$	S_P	0.164	0.050	0.002	0.764	0.490	0.056	0.926	0.730	0.176
	$\widetilde{S}_P$	0.144	0.074	0.014	0.764	0.610	0.206	0.926	0.812	0.392
	$C_{P,1}$	0.061	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\widetilde{C}_{P,1}$	0.049	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3.2. *Empirical power of unconditional* S_P *and* $\widetilde{S}_P$ *backtesting tests, and* $\widehat{C}_{P,1}$ *and* $\widetilde{C}_{P,1}$ *independence tests for the fixed forecasting scheme. The* VaR *is computed at* $\alpha = 0.01$ *and* $\alpha = 0.05$. *The error term* ε_t *is assumed to follow a Student-t with* $\nu = 30$. 1000 *Monte Carlo simulations.* $R = 500$, $P = 250$, 500. *Models SV, BIL and NLM.*

The power of both test statistics S_P and $\widetilde{S}_P$ is generally increasing in the out-of-sample size P when $\alpha = 0.05$, but that is not generally the case for $\alpha = 0.01$ at significance levels 5% and 1%. Overall, the unconditional tests possess good finite-sample power properties, pointing out the consistency of the tests against these alternatives. No test uniformly dominates the other for all models, although in general, the power of $\widetilde{S}_P$ is either comparable or higher than that of S_P. The power function of both tests is very sensitive to the choice of the significance level of the test. There is also variability of the results in α. Thus, the power against the GARCH-M is moderate for $\alpha = 0.01$ and high for $\alpha = 0.05$. The power increases from $\alpha = 0.01$ to $\alpha = 0.05$ for the GARCH-M, TAR, BIL and NLMA models, yielding the opposite results for the EGARCH and SV models. Summarizing, the corrected test statistic $\widetilde{S}_P$ presents excellent power properties, being either comparable or much better than the uncorrected test S_P for all the alternatives considered, see for instance the cases with $\alpha = 0.01$ and 0.1 significance levels.

The joint tests possess a satisfactory power against the TAR and EGARCH models but low power against the GARCH-M, BIL and NLMA models. The uncorrected test has in general more power than the corrected one, although the difference is not substantial for $\alpha = 0.05$. Like for the unconditional case, the empirical power is quite sensitive to α and to the choice of significance level, showing one more time the inaccurate approximation by the asymptotic theory, especially when $\alpha = 0.01$. This can be observed, for example, for some estimates of the power that go from zero to one when the level of the test goes from 1% to 5%.

To illustrate this we report in figure 6.5 the (kernel) smoothed finite-sample density of the test statistic $\frac{\xi_{P,1}}{\alpha(1-\alpha)}$ for $\alpha = 0.01$ and $\alpha = 0.05$, as well as the standard normal density. We abstract from any estimation effect and use the uncorrected version with known parameters. Notice that we are not interested here in the estimation effects in the asymptotic distribution but in the approximation of the finite-sample distributions by the asymptotic ones. We are particularly interested in the case $\alpha = 0.01$. Thus, we assume the parameters of the GARCH(1,1) model to be known and estimate the corresponding finite-sample density functions nonparametrically using a standard normal kernel function with bandwith parameter $h = 0.50$. The experiment is carried out for $R = 500$ and $P = 250, 500, 750, 1000$. The results of the different panels in figure 6.5 are illuminating in showing the non-negligible probability mass in the far right tail of the finite-sample density for $\alpha = 0.01$ and therefore the stark differences between the finite-sample behaviour of the test statistic in this case and for $\alpha = 0.05$. The approximation is clearly better for $\alpha = 0.05$ than for $\alpha = 0.01$, and the addtional mass at large values in the distribution when $\alpha = 0.01$ may explain the decreasing power when one test at 5% or 1%, see Tables 3.1 and 3.2.

We summarize the findings of our Monte Carlo simulations as follows. The corrected unconditional and joint tests have, for the models considered and uniformly in all VaR levels, better size performance than the respective uncorrected tests. The larger the parameter $\widehat{\pi} = P/R$ the higher the distortions. These distortions are of positive sign, *i.e.* overrejection, which is consistent with our asymptotic theory. For the joint case, the greater the α the more important is the correction. For instance, for $\alpha = 0.1$ there is a clear improvement of the corrected test over the uncorrected one. The thickness of the tails does not play a significant role in distorting the size values for these joint tests. Thus, our paper contributes to the existent literature, *e.g.* West (1996) and McCracken (2000), documenting that in many circumstances it is inappropriate to ignore parameter uncertainty in forecast predictive ability tests. In addition, we have shown that the size improvement is without sacrificing power for the unconditional hypothesis. In fact, the unconditional corrected test is either comparable or better than the uncorrected test for the alternatives considered. For the joint test, the uncorrected version turns out to exhibit higher power than the corrected one, although the difference is not significant for $\alpha = 0.05$. Finally, in all our experiments, the larger the VaR level α the better is the approximation by the asymptotic theory of the finite sample distributions. The approximation for $\alpha = 0.01$ is poor and may lead to misleading conclusions for common in-sample and out-of-sample sizes.

4 Application to financial data

We have uncovered in this paper that the standard backtesting (unconditional and independence) techniques can produce wrong type-I error probabilities for assessing VaR estimates from parametric models with unknown parameters. This fact can have a significant impact on market risk management depending on the backtesting technique employed, the in-sample period used to estimate the parameters, the corresponding out-of-sample period, and/or the choice of parametric model. This effect is gauged in this application for daily log-returns on the $S\&P500$ market-valued equity Index obtained from *Freelunch.com* over the period 02/2000 - 02/2006 ($n = 1500$ observations).

We entertain a pure Gaussian GARCH(1,1) model for the log-returns Y_t, that yields the following VaR risk model,

$$m_\alpha(W_{t-1}, \theta_0) = \sigma_t \Phi_\varepsilon^{-1}(\alpha), \qquad \sigma_t^2 = \eta_{00} + \eta_{10} Y_{t-1}^2 + \eta_{20}\sigma_{t-1}^2,$$

where $\Phi_\varepsilon^{-1}(\alpha)$ is the α-quantile of the Gaussian error distribution. This specification is motivated from the application of some goodness-of-fit tests developed in Escanciano (2007b) for testing the correct specification of the variance model and the error distribution in location-scale models. These specifications are not rejected for the sample periods considered. Under the assumption that the true data generating processes belong to the location-scale family, the fact that the GARCH(1,1) model seems to be a good fit for these data helps to identify the differences between the corrected and uncorrected tests as being solely caused by the estimation risk effects. We have also entertained a GARCH(1,1) and an ARMA(1,1)-GARCH(1,1) models with Student-t distributions, with and without constants terms, leading to similar conclusions. Details are omitted to save space.

The parameters are estimated by $QMLE$ using $R = 250$ observations and their values, jointly with their standard errors, can be obtained from the authors upon request. The out-of-sample period is also $P = 250$ observations, thus, with the data set available we have repeated the backtesting experiment for five different periods starting in February 2000. That is, for the second period, observations from $t = 251$ to 500 form the in-sample period and from $t = 501$ to 750 the out-of-sample period, and so forth.

The aim of this application is to illustrate how the estimation risk may lead to different decisions in inferences with the corrected and uncorrected tests. We consider the fixed forecasting scheme studied in the simulations. Table 4.1 reports the different statistics of the unconditional backtesting and joint tests for $VaR_{0.01}$ and $VaR_{0.05}$ for the five periods under study. In this table we also report the number of exceedances (vio) for each period.

GARCH(1,1)	vio	S_P	$\widetilde{S}_P$	$C_{P,1}$	$\widetilde{C}_{P,1}$
$\alpha = 0.01$					
P1	4	0.953	0.869	39.11^{**}	37.39^{**}
P2	3	0.317	0.278	0.025	0.025
P3	0	-1.589	-1.228	0.025	0.025
P4	2	-0.317	-0.261	0.025	0.025
P5	1	-0.953	-0.870	0.025	0.025
$\alpha = 0.05$					
P1	14	0.435	0.380	0.254	0.223
P2	16	1.015	0.838	3.391	2.923
P3	4	-2.466^{**}	-1.733	0.692	0.643
P4	14	0.435	0.331	3.391	2.857
P5	8	-1.305	-1.143	0.692	0.652

Table 4.1. *Statistics corresponding to the unconditional backtesting and independence tests for VaR_α, with $\alpha = 0.01$ and 0.05 for five samples of 500 observations starting on February 2000. $R = 250$ and $P = 250$. Data are fitted to a GARCH(1,1) process with Gaussian innovations.* $^{(*)}$ *denotes statistical significance at 5% level and* $^{(**)}$ *at 1% level.*

We observe that the values of both uncorrected backtesting tests are larger (in absolute value) than those of the corrected tests $\widetilde{S}_P$ and $\widetilde{C}_{P,1}$. This can lead, and in fact does for the unconditional test in the third period, to an overrejection of the risk model. Likewise, given that $\chi^2_{1,0.07} = 3.283$, the uncorrected joint test $C_{P,1}$ would lead to spurious rejections at a 7% significance level in the second and fourth periods that $\widetilde{C}_{P,1}$ would not. In terms of capital requirements, these uncorrected tests would indicate that the GARCH(1,1) model is a conservative risk model that would be implying an extra allocation of idle capital. However, by correcting by estimation risk effects we observe that this is not the case and that the VaR obtained from the GARCH(1,1) family of models seems to be an appropriate risk model for these data sets. It is also worth observing the overwhelming rejection of $C_{P,1}$ and $\widetilde{C}_{P,1}$ for $VaR_{0.01}$ in the first period, these values are probably due to absence of data for the analysis for very low quantile levels, rather than to a truly rejection of the GARCH model.

5 Conclusion

Basel and Basel II Accords propose the use of backtesting techniques to assess the accuracy and reliability of these internal risk management models, usually encapsulated in Value at Risk measures, and set different failure areas for institutions failing to report valid risk models. Thereby the validity of these backtesting procedures is of paramount importance for the reliability of the whole internal and external monitoring process.

We have shown in this paper that the standard unconditional and independence backtesting used by banks and regulators to assess dynamic parametric VaR estimates may be very misleading in composite environments. This implies that any conclusion regarding the validity of these risk models based on standard backtesting procedures may be spurious. This is because the cut-off point determining the validity of the risk management model is wrong. We find the appropriate cut-off point by correcting the variance in the relevant test statistics corresponding to the recursive, rolling and fixed out-of-sample forecasting schemes. In fact, in the simulation exercises performed for the fixed forecasting scheme we find evidence of significant size distortions for the Kupiec uncorrected test. For joint tests, as predicted by our theory, the distortions are only significant for moderate and large values of α such as $\alpha = 0.1$. These distortions are remarkably important for backtesting exercises using large out-of-sample sizes and small in-sample sizes for estimating the parameters. The opposite case, on the other hand, yields negligible estimation risk effects. Finally, our simulations indicate that the approximation by the asymptotic theory is not accurate for small values of α such as $\alpha = 0.01$.

The importance of our corrections has been also studied in an empirical application with financial returns on $S\&P500$ Index. We find that the standard unconditional backtesting procedure with VaR calculated with the fixed forecasting scheme overstates risk exposure yielding in the third period under study to a spurious rejection of VaR at 5% for the GARCH(1,1) model.

These findings somehow support the scepticism of American regulators about the implementation of Basel II risk measurement and risk monitoring techniques, and should help to restore their confidence on internal risk management systems validated by this new corrected backtesting procedure.

Extensions of this study to analyze estimation risk effects on historical simulation and hybrid methods are ongoing research. Also, since our focus in this paper was on estimation risk, we have assumed herein a correctly specified underlying VaR model (with the exception of our Theorem 1). More general backtesting tests robust to both, estimation and model risks, are of paramount practical importance. For developing such robust backtests different alternatives are available. The extension of our martingale methods to such a general framework is difficult, but different theories based on HAC estimations using mixing conditions (see McCracken (2000)) or bootstrap methods for correct inference (see e.g. Corradi and Swanson (2007)) are attractive alternatives that deserve further research.

Acknowledgements

The authors thank Peter Burridge, Marcelo Fernandes, Michael Gordy, Jim O'Brien, Pei Pei, Peter C. B. Phillips, Alessio Sancetta, Peter Spencer, and the participants in the annual Workshop in Financial Econometrics organized by the Economics Department of York University for their valuable comments. They also thank the editor, the associate editor, and two anonymous referees for their useful and stimulating suggestions and comments that led to a considerably improved and more focused version of the article. Any remaining errors are the authors' own. Juan Carlos Escanciano acknowledges financial support from the Spanish Ministerio de Educación y Ciencia, reference numbers SEJ2004-04583/ECON and SEJ2005-07657/ECON, and Jose Olmo from the Spanish Ministerio de Educación y Ciencia, reference number SEJ2004-04101/ECON.

6 Appendices

6.1 Appendix A: Mathematical Proofs

We prove Theorem 1 using empirical processes theory and a small variation of a weak convergence theorem in Delgado and Escanciano (2006). The complete version of this proof is found in the working paper version of this article and available from the web pages of the authors.

Define the process

$$K_n(c) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[I_{t,\alpha}(\theta_0 + c(t-1)^{-1/2}) - F_{t-1}(\theta_0 + c(t-1)^{-1/2}) \right]$$

indexed by $c \in \mathcal{C}_K$, where $\mathcal{C}_K = \{c \in \mathbb{R}^p : |c| \leq K\}$, and $K > 0$ is an arbitrary but fixed constant.

Lemma A1: *Under Assumption A1-A5, the process $K_n(c)$ is asymptotically tight with respect to $c \in C_K$.*

The proof of Lemma A1 can be found in the working paper version.

Proof of Theorem 1: Simple but tedious algebra shows that for each $c \in \mathcal{C}_K$,

$$E\left[|K_n(c) - K_n(0)|^2 \right] = o(1).$$

The last display and the asymptotically tightness of $K_n(c)$ imply that if $\widehat{c}$ is bounded in probability, $\widehat{c} = O_P(1)$, then

$$|K_n(\widehat{c}) - K_n(0)| = o_P(1). \tag{18}$$

Now, we will apply this argument with $\widehat{c} := \max_{R \leq t \leq n} \sqrt{t}(\widehat{\theta}_t - \theta_0)$, with R denoting the in-sample sample size. Thus, we should prove that under our three forecasting schemes

$$\max_{R \leq t \leq n} \sqrt{t}(\widehat{\theta}_t - \theta_0) = O_P(1) \tag{19}$$

holds.

(i) Recursive: Our assumptions imply that $\sqrt{t}S_t = \sum_{s=1}^{t} l(Y_t, W_{t-1}, \theta_0)$ is a martingale with respect to $\mathcal{F}_{t-1}$, where S_t is implicitly defined. Hence, by Corollary 2.1 in Hall and Heyde

(1980) and A5

$$
\begin{aligned}
P\left(\left|\max_{R\le t\le n} S_t\right| > \varepsilon\right) &\le P\left(\left|\max_{R\le t\le n} \sqrt{t}S_t\right| > \sqrt{R}\varepsilon\right) \\
&\le \frac{1}{R\varepsilon^2} E\left[\left|\sqrt{n}S_n\right|^2\right] \\
&\le C\frac{n}{R\varepsilon^2},
\end{aligned}
$$

which can be made arbitrarily small by choosing ε sufficiently large, since $n/R \to (1+\pi)$ as $n \to \infty$.

(ii) Rolling: same proof as for the recursive. Details are omitted.

(iii) Fixed: $\left|\max_{R\le t\le n} (\sqrt{t}/R) \sum_{s=1}^{R} l(Y_t, W_{t-1}, \theta_0)\right| \le \left|(1/\sqrt{R}) \sum_{s=1}^{R} l(Y_t, W_{t-1}, \theta_0)\right| = O_P(1).$

Then, (18) holds for $\widehat{c} = \max_{R\le t\le n} \sqrt{t}(\widehat{\theta}_t - \theta_0)$, and hence

$$
\left|\frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[I_{t,\alpha}(\widehat{\theta}_{t-1}) - F_{t-1}(\widehat{\theta}_{t-1})\right] - \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[I_{t,\alpha}(\theta_0) - F_{t-1}(\theta_0)\right]\right| = o_P(1),
$$

which implies the decomposition

$$
\begin{aligned}
\frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}(I_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha) &= \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[I_{t,\alpha}(\theta_0) - F_{t-1}(\theta_0)\right] \qquad (20) \\
&+ \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[F_{t-1}(\widehat{\theta}_{t-1}) - F_{t-1}(\theta_0)\right] \\
&+ \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[F_{t-1}(\theta_0) - \alpha\right] + o_P(1).
\end{aligned}
$$

Now, by the Mean Value Theorem and since we can interchange expectation and differentiation,

$$
\begin{aligned}
A_{1n} &:= \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[F_{t-1}(\widehat{\theta}_{t-1}) - E[F_{t-1}(\widehat{\theta}_{t-1})] - F_{t-1}(\theta_0) + E[F_{t-1}(\theta_0)]\right] \\
&= \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left(g'_\alpha(W_{t-1}, \widetilde{\theta}_{t-1}) f_{t-1}(\widetilde{\theta}_{t-1}) - E\left[g'_\alpha(W_{t-1}, \widetilde{\theta}_{t-1}) f_{t-1}(\widetilde{\theta}_{t-1})\right]\right)(\widehat{\theta}_{t-1} - \theta_0),
\end{aligned}
$$

where $\widetilde{\theta}_{t-1}$ is between $\widehat{\theta}_{t-1}$ and θ_0. Note that A2 and A3 imply that $E\left[\sup_{\theta\in\Theta_0}\left|g_\alpha(W_{t-1}, \theta) f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta))\right|\right] < C$. Hence, by the uniform law of large numbers

(ULLN) of Jennrich (1969, Theorem 2) and (19), then $A_{1n} = o_P(1)$ holds. Similarly,

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[E[F_{t-1}(\widehat{\theta}_{t-1})] - E[F_{t-1}(\theta_0)]\right] \\
= \ & \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n} E[g'_\alpha(W_{t-1},\theta_0)f_{t-1}(\theta_0)](\widehat{\theta}_{t-1}-\theta_0) + \\
& + \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[E[g'_\alpha(W_{t-1},\widetilde{\theta}_{t-1})f_{t-1}(\widetilde{\theta}_{t-1})] - E[g'_\alpha(W_{t-1},\theta_0)f_{t-1}(\theta_0)]\right](\widehat{\theta}_{t-1}-\theta_0) \\
: \ & = B_{1n} + B_{2n}.
\end{aligned}$$

Now, by the ULLN and (19), then $B_{2n} = o_P(1)$ holds. Hence,

$$\left|\frac{1}{\sqrt{P}}\sum_{t=R+1}^{n}\left[F_{t-1}(\widehat{\theta}_{t-1}) - F_{t-1}(\theta_0)\right] - E[g'_\alpha(W_{t-1},\theta_0)f_{t-1}(\theta_0)]\frac{1}{\sqrt{P}}\sum_{t=R+1}^{n} H(t-1)\right| = o_P(1).$$

The theorem follows from (20) and the last display. □

Proof of Corollary 1: Once Theorem 1 has been established, the proof follows the same arguments as in McCracken (2000, Theorem 2.3.1). Details are omitted to save space. □

Proof of Corollary 2: The consistency of $\widehat{\rho}$ and $\widehat{V}$ follows from the ULLN of Jennrich (1969, Theorem 2) and (19). Giacomini and Komunjer (2005), on the other hand, proved the consistency of the out-of-sample version of A_τ. It also follows in this context that $\widehat{A}_\tau = A + o_P(1)$. Now, by Slutsky's Lemma the corollary is proved. □

Proof of Theorem 2: The proof is similar to that of Theorem 1. Define the process

$$K_{n,j}(c) := \frac{1}{\sqrt{P}}\sum_{t=R+j+1}^{n}\left[I_{t,\alpha}(\theta_0 + c(t-1)^{-1/2}) - F_{t-1}(\theta_0 + c(t-1)^{-1/2})\right] I_{t-j,\alpha}(\theta_0 + c(t-j-1)^{-1/2}),$$

indexed by $c \in \mathcal{C}_K$, where $\mathcal{C}_K = \{c \in \mathbb{R}^p : |c| \leq K\}$, $j \geq 1$, and $K > 0$ is an arbitrary but fixed constant. Applying Theorem A1 to $K_{n,j}(c)$, as in Lemma A1, and following the arguments in

Theorem 1, we obtain the decomposition

$$\begin{aligned}
\sqrt{P-j}\left(\widehat{\xi}_{P,j}-\xi_{P,j}\right) &= \frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}\left[F_{t-1}(\widehat{\theta}_{t-1})I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})-F_{t-1}(\theta_0)I_{t-j,\alpha}(\theta_0)\right]+o_P(1)\\
&= \frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}\left[F_{t-1}(\theta_0)I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})-F_{t-1}(\theta_0)I_{t-j,\alpha}(\theta_0)\right]\\
&\quad +\frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}\left[g_{\alpha}'(W_{t-1},\widetilde{\theta}_{t-1})f_{t-1}(\widetilde{\theta}_{t-1})I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})\right](\widehat{\theta}_{t-1}-\theta_0)+o_P(1)\\
&: \quad = C_{1n}+C_{2n}+o_P(1),
\end{aligned}$$

where $\widetilde{\theta}_{t-1}$ is between $\widehat{\theta}_{t-1}$ and θ_0. Since, $F_{t-1}(\theta_0)=\alpha$ a.s., Theorem 1 implies

$$C_{1n}=\alpha E\left[g_{\alpha}'(W_{t-j-1},\theta_0)f_{t-j-1}(\theta_0)\right]\frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}H(t-j-1)+o_P(1).$$

Whereas the arguments after (20) imply that

$$\left|C_{2n}-E[g_{\alpha}'(W_{t-1},\theta_0)f_{t-1}(\theta_0)I_{t-j,\alpha}(\theta_0)]\frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}H(t-j-1)\right|=o_P(1).$$

This proves condition i). As for condition ii), define the following quantities

$$\begin{aligned}
\widehat{\xi}_{1n,j} &= \frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}\left[I_{t,\alpha}(\widehat{\theta}_{t-1})-\alpha\right]\\
\widehat{\xi}_{2n,j} &= \frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n}\left[I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})-\alpha\right],
\end{aligned}$$

and similarly, define $\xi_{1n,j}$ and $\xi_{2n,j}$ with θ_0 replacing $\widehat{\theta}_{t-1}$. Now, simple algebra shows that

$$\sqrt{P-j}\widehat{\gamma}_{P,j}=\widehat{\xi}_{P,j}-\alpha\widehat{\xi}_{1n,j}-\alpha\widehat{\xi}_{2n,j}.$$

The same equality holds for $\gamma_{P,j}$, $\xi_{P,j}$, $\xi_{1n,j}$ and $\xi_{2n,j}$. Hence

$$\sqrt{P-j}\left(\widehat{\gamma}_{P,j}-\gamma_{P,j}\right)=\left(\widehat{\xi}_{P,j}-\xi_{P,j}\right)-\alpha\left(\widehat{\xi}_{1n,j}-\xi_{1n,j}\right)-\alpha\left(\widehat{\xi}_{2n,j}-\xi_{2n,j}\right). \tag{21}$$

Theorem 1 implies that, for $h = 1$ and 2,

$$\widehat{\xi}_{hn,j} - \xi_{hn,j} = E\left[g'_{\alpha}(W_{t-1},\theta_0)f_{t-1}(\theta_0)\right]\frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n} H(t-j-1) + o_P(1).$$

The latter display, part i) and (21) prove condition ii).

As for condition iii), it can be similarly shown that

$$\begin{aligned}
\sqrt{P-j}(\widehat{\zeta}_{P,j} - \zeta_{P,j}) &= \sqrt{P-j}(\widehat{\xi}_{P,j} - \zeta_{P,j}) - 2\alpha A \frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n} H(t-j-1) + o_P(1) \\
&= \sqrt{P-j}\left(\widehat{\xi}_{P,j} - \xi_{P,j}\right) - 2\alpha A\left(\frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n} H(t-j-1)\right) + 2\alpha\xi_{1n,j} + o_P(1) \\
&= \{B - 2\alpha A\}\left(\frac{1}{\sqrt{P-j}}\sum_{t=R+1+j}^{n} H(t-j-1)\right) + 2\alpha\xi_{1n,j} + o_P(1).
\end{aligned}$$

Details are omitted to save space. □

Proof of Corollary 3: The consistency of $\widehat{\eta}$ and $\widehat{V}$ follows from the ULLN of Jennrich (1969, Theorem 2) and (19). The consistency of $\widehat{B}_{\tau}$ follows from Giacomini and Komunjer (2005). Now, by Slutsky's Lemma the corollary follows. □

6.2 Appendix B: Figures

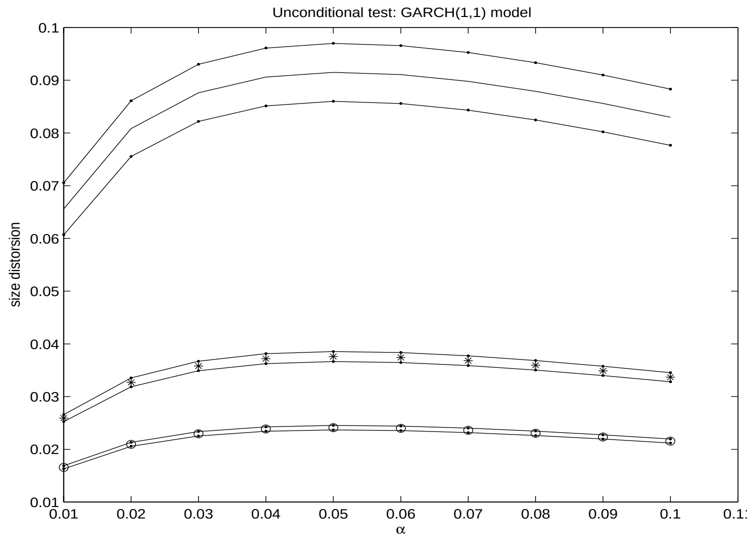

Figure 6.1. *Sample average and 95% empirical confidence intervals for $d(\alpha)$, with $\upsilon = 0.05$, in expression (8). The relevant process is a GARCH(1,1) with parameters $(\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$ and error distributed as a Student-t with $\nu = 30$. $(R = 250, P = 500)$ is plotted with $(.-)$, $(R = 500, P = 500)$ is plotted with $(*-)$ and $(R = 750, P = 500)$ with $(o-)$. $M = 500$ Monte Carlo replications.*

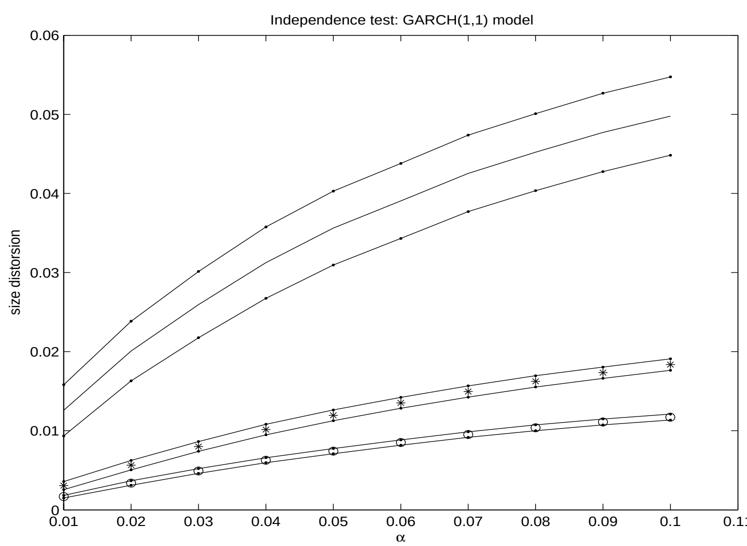

Figure 6.2. *Sample average and 95% empirical confidence intervals for the independence test version of $d(\alpha)$, with $\upsilon = 0.05$, in expression (8). $\widehat{\sigma}_u(\alpha)$ is replaced by $\widehat{\sigma}_c(\alpha)$. The relevant process is a GARCH(1,1) with parameters $(\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$ and error distributed as a Student-t with $\nu = 30$. $(R = 250, P = 500)$ is plotted with $(.-)$, $(R = 500, P = 500)$ is plotted with $(*-)$ and $(R = 750, P = 500)$ with $(o-)$. $M = 500$ Monte Carlo replications.*

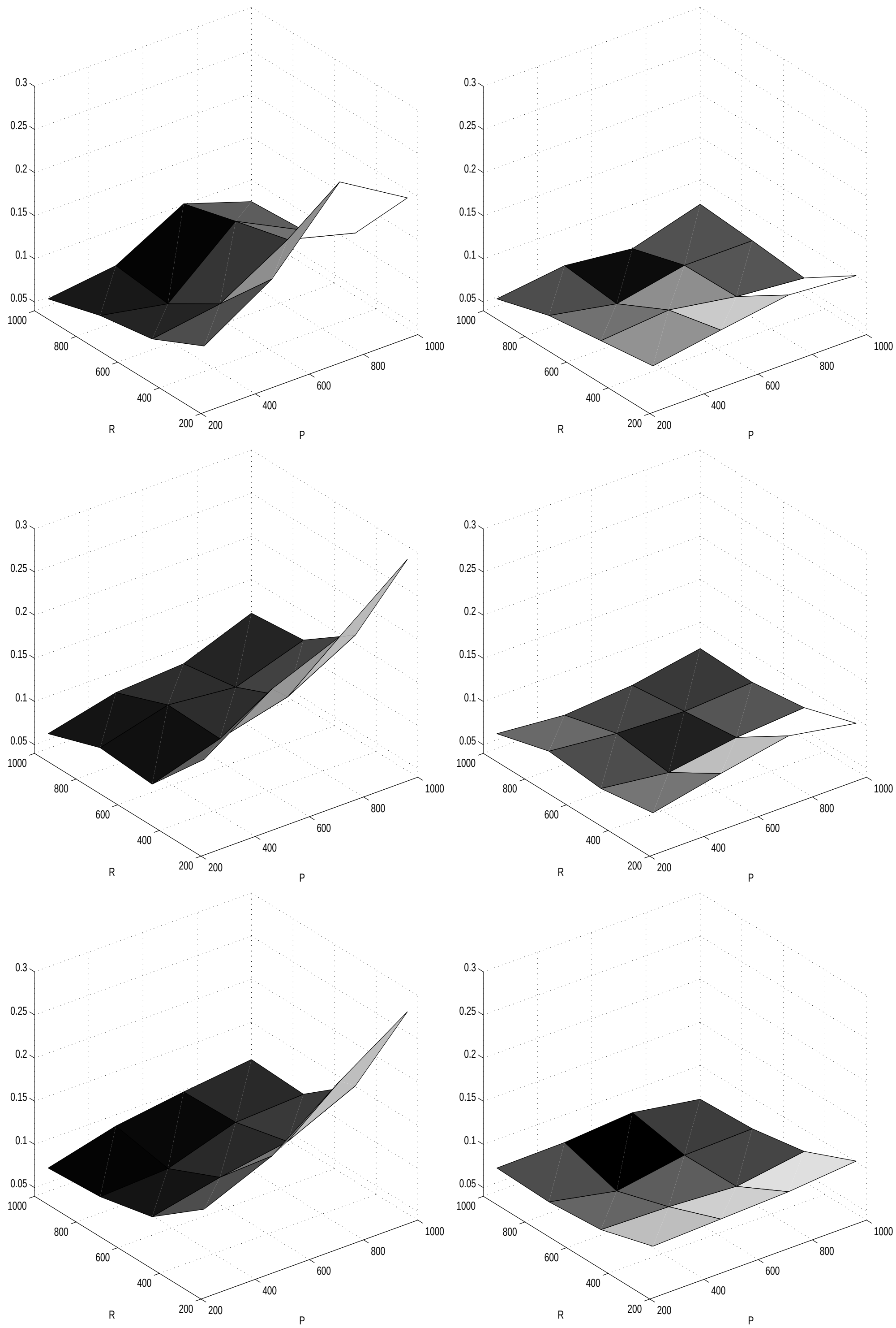

Figure 6.3. *Simulated 0.05 size for* S_P *and* $\widetilde{S}_P$ *tests for* VaR_α *of a GARCH(1,1) with* $(\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$*, for* $\alpha = 0.01$ *in the upper panels,* $\alpha = 0.05$ *in the middle and* $\alpha = 0.1$ *in the lower panels.* $\nu = 30$ *df for a Student-t.* S_P *is on the left and* $\widetilde{S}_P$ *on the right.* R *and* P *take the values* $[250, 500, 750, 1000]$*. 500 Monte Carlo simulations are carried out.*

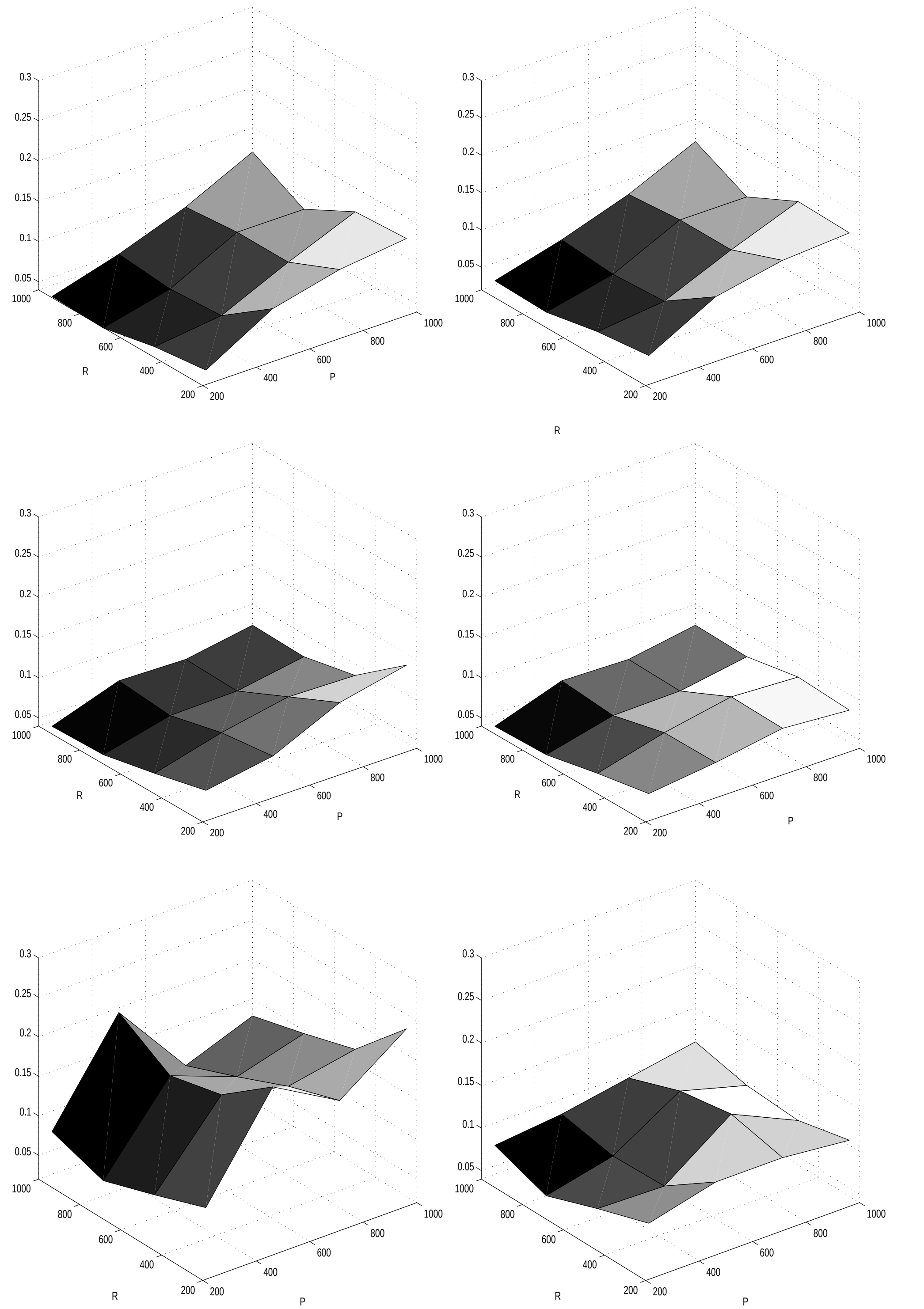

Figure 6.4. *Simulated 0.05 size for* $C_{P,1}$ *and* $\widetilde{C}_{P,1}$ *tests for* VaR_α *of a GARCH(1,1) with* $(\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$*, for* $\alpha = 0.01$ *in the upper panels,* $\alpha = 0.05$ *in the middle and* $\alpha = 0.1$ *in the lower panels.* $\nu = 30$ *df for a Student-t.* $C_{P,1}$ *is on the left and* $\widetilde{C}_{P,1}$ *on the right.* R *and* P *take the values* $[250, 500, 750, 1000]$*. 500 Monte Carlo simulations are carried out.*

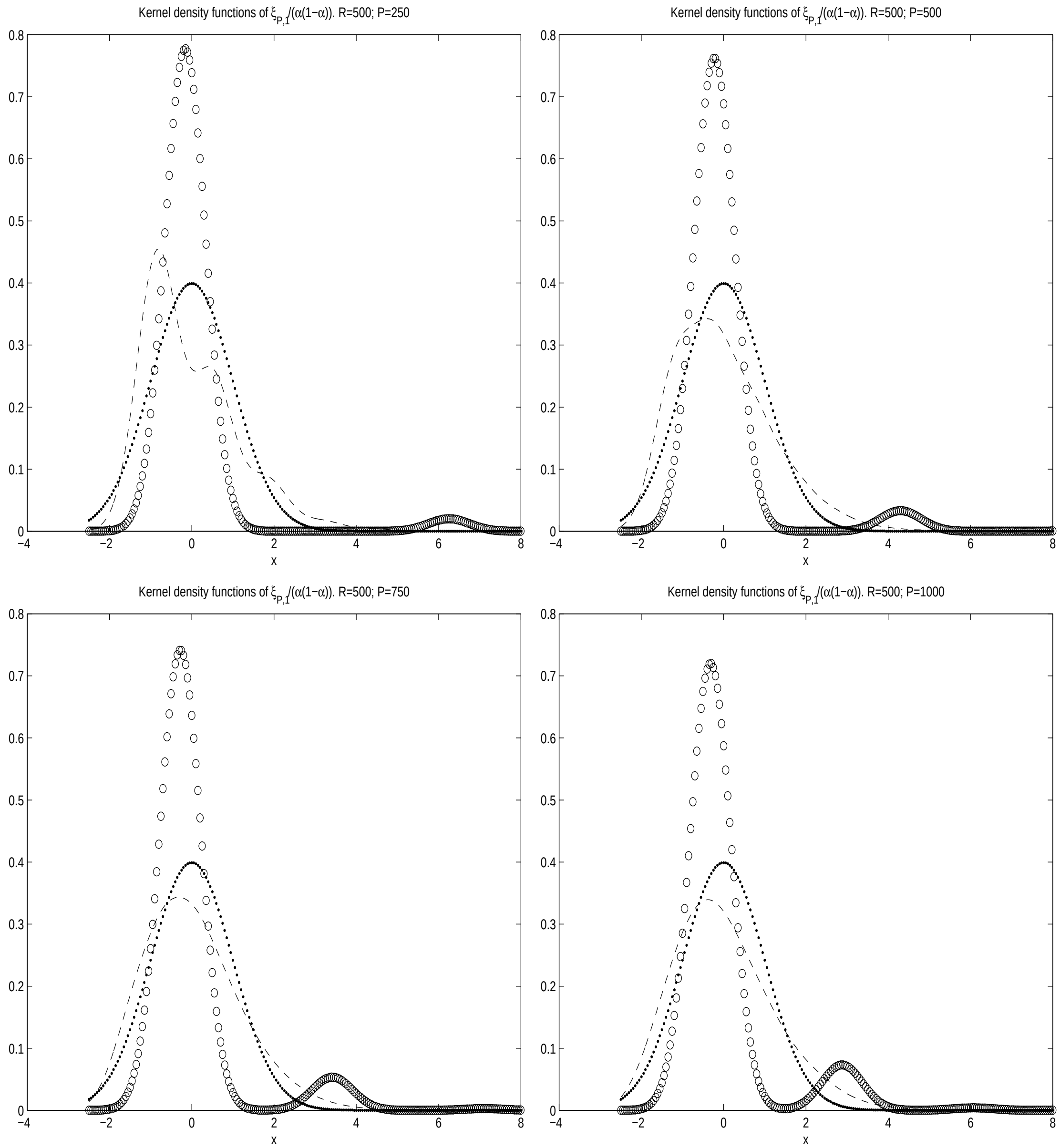

Figure 6.5. *Finite-sample density function of* $\frac{\xi_{P,1}}{\alpha(1-\alpha)}$ *plotted with a circle line for* $\alpha = 0.01$*, and with a dashed line for* $\alpha = 0.05$*. The asymptotic Gaussian density function is represented with a dotted line. The relevant process is a* VaR_α *of a GARCH(1,1) with* $(\eta_{00}, \eta_{10}, \eta_{20}) = (0.05, 0.1, 0.85)$ *and error term given by a Student-t distribution with* $\nu = 30$ *degrees of freedom. 5000 Monte Carlo simulations are carried out.*

REFERENCES

Basel Committee on Banking Supervision (1996a): "Amendment to the Capital Accord to Incorporate Market Risks, Bank for International Settlements. "

Basel Committee on Banking Supervision (1996b): "Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements, Bank for International Settlements. "

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2006): "Evaluating Value-at-Risk Models with Desk-level Data. " *Working Paper Series, 010* Department of Economics, North Carolina State University.

Berkowitz, J., and O'Brien, J. (2002): "How Accurate Are Value at Risk Models at Commercial Banks?" *Journal of Finance*, 57, 1093-1111.

Box, G. and Pierce, D. (1970): "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models." *Journal of American Statistical Association*, 65, 1509-1527.

Chan, N.H, Deng, S., Peng, L. and Xia, Z. (2006): "Interval Estimation for the Conditional Value-at-Risk based on GARCH Models with Heavy Tailed Innovations." *Journal of Econometrics,* 137, 556-576.

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

Christoffersen, P., Hahn, J., and Inoue, A. (2001): "Testing and Comparing Value at Risk Measures." *Journal of Empirical Finance*, 8, 325-342.

Christoffersen, P., Gonçalves, S. (2005): "Estimation Risk in Financial Risk Management. " *Journal of Risk*, 7, 1-28.

Corradi, V., and Swanson, N.R. (2007): "Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes." *International Economic Review,* 48, 67-109.

Delgado, M.A. and Escanciano, J.C. (2006): "Nonparametric Tests for Conditional Symmetry in Dynamic Models." *Journal of Econometrics*, 141, 652-682.

Durlauf, S.N. (1991): "Spectral Based Testing on the Martingale Hypothesis. " *Journal of Econometrics*, 50, 355-376.

ENGLE, R., AND MANGANELLI, S. (2004): "CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles. "*Journal of Business and Economic Statistics,* 22, 367-381.

ESCANCIANO, J. C. (2007a): "Weak Convergence of Non-stationary Multivariate Marked Processes with Applications to Martingale Testing", *Journal of Multivariate Analysis*, 98, 1321-1336.

ESCANCIANO, J. C. (2007b): "Weak Convergence of Weighted Conditional Empirical Processes of Residuals in Heteroscedastic Time Series Models", working paper.

ESCANCIANO, J. C. (2008): "Joint and Marginal Specification Tests for Conditional Mean and Variance Models", *Journal of Econometrics,* 143, 74-87.

ESCANCIANO, J. C. AND OLMO, J. (2007): "Estimation Risk Effects on Backtesting for Parametric Value-at-Risk Models". CAEPR Working Paper No. 2007-005.

FAN, J. AND GU, J. (2003): "Semiparametric Estimation of Value at Risk." *Econometrics Journal*, 6, 261-290.

GIACOMINI, R. AND KOMUNJER, I. (2005): "Evaluation and Combination of Conditional Quantile Forecasts. "*Journal of Business & Economic Statistics*, 23, 4, 416-431(16).

GOUREIOUX, C., AND JASIAK, J. (2006): "Dynamic Quantile Models. " Working Paper 2006-4, Department of Economics, York University.

HALL, P., AND HEYDE, C. (1980): *Martingale Limit Theory and Its Application. Probability and Mathematical Statistics.* Academic Press, New York.

HANNAN, E. J. AND QUINN, B. (1979): "The determination of the order of an autoregression", *Journal of the Royal Statistical Society, Series B,* 41, 190-191.

JENNRICH, R.I. (1969): "Asymptotic Properties of Nonlinear Least Square Estimators. "*Annals of Mathematical Statistics,* 40, 633-643.

JORION, P. (2000): "Value-at-Risk: The New Benchmark for Controlling Financial Risk. "Chicago: McGraw-Hill.

KOENKER, R. AND XIAO, Z. (2006): "Quantile autoregression." *Journal of the American Statistical Association*, 101, 980-990.

KOENKER, R., ZHAO, Q. (1996): "Conditional Quantile Estimation and Inference for ARCH Models."*Econometric Theory*, 12, 5, 793-814.

Koul, H.L., and Stute W. (1999): "Nonparametric Model Checks for Time Series." *Annals of Statistics*, 27, 204-236.

Kuester, K., Mittnik, S., and Paolella, M.S. (2006): "Value-at-Risk Prediction: A Comparison of Alternative Strategies. "*Journal of Financial Econometrics*, 4, 1, 53-89.

Kupiec, P. (1995): "Techniques for Verifying the Accuracy of Risk Measurement Models, "*Journal of Derivatives*, 3,73-84.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983): *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Li, Q. and Racine, J. S. (2006): *Nonparametric Econometrics: Theory and Practice.* Princeton University Press.

Li, W.K., Ling, S., and McAleer, M. (2002): "Recent Theoretical Results for Time Series Models with GARCH Errors. *Journal of Economic Surveys*, 16, 245-269. Reprinted in M. McAleer and L. Oxley (eds.), *Contributions to Financial Econometrics: Theoretical and Practical Issues* (Oxford: Blackwell, 2002), 9-33.

Linton, O. and Whang Y-J (2004): "A Quantilogram approach to testing for Directional Predictability." Forthcoming in *Journal of Econometrics*.

Ljung, G. M. and Box, G. E. P. (1978): "A Measure of Lack of Fit in Time Series Models."*Biometrika,* 65, 297-303.

Martinez, O., and Olmo, J. (2007): "A Nonlinear Threshold Model for Dependence of Extremes of Stationary Sequences. "Working Paper.

Martins-Filho, C., and Yao, F. (2006): "Estimation of Value-at-Risk and Expected Shortfall based on Nonlinear Models of Return Dynamics and Extreme Value Theory." *Studies in Nonlinear Dynamics & Econometrics*, 10, Article 4.

McCracken, (2000): "Robust Out-of-sample Inference." *Journal of Econometrics,* 99, 5, 195-223.

van der Vaart, A. W. (1998): *Asymptotic Statistics.* Cambridge Univ. Press.

van der Vaart, A. W., and Wellner, J.A.(1996): *Weak Convergence and Empirical Processes.* Springer, New York.

West, K.D. (1996): "Asymptotic Inference about Predictive Ability."*Econometrica,* 64, 5, 1067-1084.