**Department of Economics**
**School of Social Sciences**

# Estimation risk effects on backtesting for parametric value-at-risk models

**J. Carlos Escanciano**[1]
**Indiana University**

**Jose Olmo**[2]
**City University**

**Department of Economics**

**Discussion Paper Series**

**No. 07/11**

1 Indiana University, Bloomington, IN, USA.

2 Department of Economics, City University, Northampton Square, London, EC1V 0HB, UK. Email:author@city.ac.uk

# ESTIMATION RISK EFFECTS ON BACKTESTING FOR PARAMETRIC VALUE-AT-RISK MODELS

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

Jose Olmo†
*City University, London, UK*

This draft, March 2007

**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. Thereby the correct specification of parametric VaR models became of crucial importance in order to provide accurate and reliable risk measures. If the underlying risk model is not correctly specified, VaR estimates understate/overstate risk exposure. This can have dramatic consequences on stability and reputation of financial institutions or lead to sub-optimal capital allocation. We show that the use of the standard unconditional backtesting procedures to assess VaR models is completely misleading. These tests do not consider the impact of estimation risk and therefore 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 specification analyses. A Monte Carlo study illustrates our theoretical findings in finite-samples. Finally, an application to $S\&P500$ Index shows the importance of this correction and its impact on capital requirements as imposed by Basel Accord, and on the choice of dynamic parametric models for risk management.

**Keywords and Phrases**: Backtesting; Basel Accord; Model Risk; Risk management; Value at Risk; Conditional Quantile.

*Research funded by the Spanish Ministerio de Educación y Ciencia, reference numbers SEJ2004-04583/ECON and SEJ2005-07657/ECON.

†Research funded by the Spanish Ministerio de Educación y Ciencia, reference number SEJ2004-04101/ECON.

# 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 conditional quantile of the conditional distribution of returns on the portfolio given agent's information set. For banks with sufficiently high developed risk management systems this was a priori the only restriction set by the Basel Accord (1996a) for computing capital reserves. Thus, large financial institutions gained the possibility of computing their own risk measures and hence to err on the side of using models infra-estimating risk. To monitor these models the Basel Accord (1996a), and the Ammendment of Basel Accord (1996b) developed a backtesting procedure to assess the accuracy and quality of different risk measurement techniques. This process is used both by banks' internal units to measure the accuracy of their models and by external regulators that on the basis of backtesting performance set appropriate punishments reflected on additional capital requirements applicable in case of failure.

The essence of backtesting is the comparison of actual trading results with model-generated risk measures. If the comparison uncovers sufficient differences between both figures either the risk model, the data or the assumptions on the backtesting technique should be subject to revision by the corresponding regulatory body. From Basel Committee's perspective backtesting consists on statistically testing whether the observed percentage of outcomes covered by the risk measure is consistent with a 99% VaR level. In other words, under "normal" conditions the return on the market portfolio should be below the VaR reported by the institution once every one hundred times. The Basel Accord defines three zones: green, yellow and red dependent on the number of times returns fall below VaR. These zones signal the accuracy of the model and determine the penalty on financial institutions failing to report a green zone. The first method suggested by Basel Accord to assess the accuracy of relevant VaR measures was the use of one-sided confidence intervals derived from a binomial distribution. In this model then observations are assumed to be independent and identically distributed (*iid*).

The first attempt to improve this technique is Christoffersen (1998) that considers serial dependence of order one in the occurrence of exceedances. This author develops hypothesis tests to assess the unconditional and conditional coverage error. In these and other existing backtesting techniques the VaR measures are assumed to be known. The knowledge of VaR however is a strong an assumption that is rarely satisfied in practice. Other more realistic alternatives given by parametric specifications of the returns dynamics assume that VaR is known up-to a finite-dimensional parameter. Then, it is common practice in the literature to estimate the parameters in the VaR specification and proceed with standard backtesting procedures, including Christoffersen's (1998) conditional and unconditional tests. See *e.g.* Berkowitz and O'Brien (2002). We shall show that the estimation of parameters in the VaR model has a nonnegligible effect in the asymptotic distribution of unconditional backtesting tests and leads to a different source of risk, called *estimation risk*. This estimation risk invalidates inference results developed in Christoffersen (1998) and related unconditional backtesting inferences for the case when parameters are unknown. We shall show that, on the contrary, estimation of parameters in *conditional* backtesting procedures does not lead to estimation risk failures.

The study of estimation risk is not new in this literature but it has not been studied in detail yet. Christoffersen, Hahn and Inoue (2001) propose a method to compare non-nested VaR estimates in location-scale models. Although they take into account estimation risk, the parameters in these models have to be estimated by the sophisticated information theoretic alternative to GMM due to Kitamura and Stutzer (1997). Chirstoffersen and Gonçalves (2005) measure estimation uncertainty in this framework by constructing bootstrap predictive confidence intervals for risk measures, in particular VaR. The use of bootstrap techniques can be computationally time demanding, since usually re-estimation of parameters in each bootstrap replication is necessary. In a similar context Figlewski (2003) finds by simulation techniques that the estimation error can increase the probability of extreme events producing overconservative risk measures. Kerkhof and Melenberg (2002) apply the functional delta method to propose a general backtesting methodology. Note however that to perform the quantile transform proposed by these authors one needs to assume that the conditional distribution of the data is known. Hansen (2006) constructs asymptotic forecast intervals which incorporate the uncertainty due to parameter estimation, but this author does not consider model evaluation (see Hansen, 2006, p. 379).

To the best of our knowledge no paper has quantified the estimation risk in a general parametric VaR dynamic framework. Therefore, the first aim of this article is to quantify it and to stress its impact in the standard backtesting procedures. We show that in the unconditional framework these techniques used for model checking are completely misleading. The second aim of the paper is then to propose a corrected backtesting method taking into account such risk, and thereby providing a valid statistical framework for measuring and evaluating market risk. The results are given for general dynamic parametric models and two particular examples, historical simulation and location-scale models, will be worked out to illustrate the general methodology. In accordance with most of the financial and econometrics literature we deal with in-sample-tests for unconditional coverage and serial independence. Out-of-sample analysis can be analyzed similarly but involves cumbersome notation, see *e.g.* Giacomini and Komunjer (2005).

The paper is structured as follows. Section 2 formulates the statistical testing problem and reviews the most common backtesting techniques. Section 3 studies the estimation risk in unconditional and conditional tests and proposes a new corrected backtesting procedure for the unconditional case taking into account such risk. As particular examples, we consider the most common VaR specifications, namely, historical simulation and location-scale models with known standardized error distribution. Section 4 illustrates via Monte Carlo experiments with different data generating processes our theoretical findings in finite samples. Section 5 introduces an application of our procedures to quantify the implications on capital requirements of correcting the critical values of the standard backtesting test for $S\&P500$ Index tracking the $US$ equity market. Finally, Section 6 concludes. Mathematical proofs are gathered into an appendix.

## 2 Backtesting techniques

Under the Market Risk Amendment to the Basel Accord effective in 1996, qualifying financial institutions have the freedom to specify their own model to compute their Value-at-Risk. It thus becomes crucially important for regulators to assess the quality of the models employed by assessing the forecast accuracy - a procedure known as "backtesting".

More concretely, 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 $I_{t-1}$, which may contain past values of $Y_t$ and other relevant economic and financial variables, *i.e.*, $I_{t-1} = (Y_{t-1}, Z'_{t-1}, Y_{t-2}, Z'_{t-2}...)'$. Henceforth, $A'$ denotes the transpose matrix of $A$. Assuming that the conditional distribution of $Y_t$ given $I_{t-1}$ is continuous, we define the $\alpha$-th conditional VaR (i.e. quantile) of $Y_t$ given $I_{t-1}$ as the measurable function $q_\alpha(I_{t-1})$ satisfying the equation

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

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 $\mathcal{M}$. Inferences within the model, including forecasts analysis, depend crucially on the hypothesis that $q_\alpha \in \mathcal{M}$, i.e., if there exists some $\theta_0 \in \Theta$ such that $m_\alpha(I_{t-1}, \theta_0) = q_\alpha(I_{t-1})$ a.s. Hence the importance of model evaluation techniques. In parametric models the nuisance parameter $\theta_0$ belongs to $\Theta$, with $\Theta$ 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), where $\theta_0$ belongs to an infinite-dimensional space. This paper will focus on parametric[1] VaR models where $\theta_0$ is finite-dimensional and can be estimated by a $\sqrt{n}$-consistent estimator (cf. A4 below.) Parametric models are popular since the functional form $m_\alpha(I_{t-1}, \theta_0)$, jointly with the parameter $\theta_0$, describe in a very precise way the impact of agent's information set on VaR. See Section 3.3 for an analysis of the most common parametric VaR models, historical simulation and the 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).

From (1), the parametric VaR model $m_\alpha(I_{t-1}, \theta_0)$ is well specified if and only if

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

where $\Psi_{t,\alpha}(\theta_0) := 1(Y_t \leq m_\alpha(I_{t-1}, \theta_0))$. Condition (2) can be equivalently expressed as $\Psi_{t,\alpha}(\theta_0) - \alpha$ being a martingale difference sequence (*mds*) with respect to the sigma field generated by the agent's information set $I_{t-1}$.

[1]The reader should not be confused with the fully parametric approach in which the whole conditional distribution of $Y_t$ given $I_{t-1}$ is fully specified. We do not need such fully specification since our concern is just the conditional quantile. In this sense and in the statistical jargon our specified model is semiparametric.

The existing backtesting procedures are all 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 $\Psi_{t,\alpha}(\theta_0) - \alpha$ against its lagged values and other variables included in $I_{t-1}$, and testing whether these variables are significant in the regression. But the most popular explored implication is

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

where $\widetilde{I}_{t-1}(\theta_0) = (\Psi_{t-1,\alpha}(\theta_0), \Psi_{t-2,\alpha}(\theta_0)...)'$. 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 the backtesting procedures.

The popularity of condition (3) is mostly due to the discrete character and ease of interpretation of the variables $\{\Psi_{t,\alpha}(\theta_0)\}$, which are the so-called *hits* or *exceedances*. In particular, the discreteness of the exceedances implies that condition (3) is equivalent to

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

where Ber($\alpha$) stands for a Bernoulli *r.v.* with parameter $\alpha$.

The first tests were proposed for the unconditional version of (3) where $\{\Psi_{t,\alpha}(\theta_0)\}$ are assumed *iid* and the hypothesis of interest becomes

$$H_{0u} : E[\Psi_{t,\alpha}(\theta_0)] = \alpha.$$

Kupiec (1995) first proposed tests for $H_{0u}$ based on the absolute value of the standardized sample mean

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\Psi_{t,\alpha}(\theta_0) - \alpha). \tag{5}$$

This test is optimal if $\theta_0$ is known. Alternatively, Chirstoffersen (1998) proposed a likelihood ratio ($LR$) test for unconditional coverage based on

$$LR_u = -2log\frac{L(\alpha; \{\Psi_{t,\alpha}(\theta_0)\}_{t=1}^{n})}{L(\hat{\pi}; \{\Psi_{t,\alpha}(\theta_0)\}_{t=1}^{n})},$$

with $L(\hat{\pi}; \{\Psi_{t,\alpha}(\theta_0)\}_{t=1}^n) = (1-\hat{\pi})^{n_0}\hat{\pi}^{n_1}$, $\hat{\pi} = n_1/(n_0+n_1)$, with $n_1$ denoting the number of VaR exceedances and $n_0 = n - n_1$. The $LR$ test statistic $LR_u$ is asymptotically distributed as a $\chi^2$ distribution with one degree of freedom under the null hypothesis. Note that $LR_u$ is equivalent to Kupiec's (1995) test[2]. Hereafter when mentioning the *unconditional test* we will refer to Kupiec's test.

Cristoffersen (1998) also proposed joint $LR$ tests for

$$H_{0c} : \{\Psi_{t,\alpha}(\theta_0)\}_{t=1}^n \text{ are } iid.$$

This author embedded the sequence of hits $\{\Psi_{t,\alpha}(\theta_0)\}_{t=1}^n$ in a first-order Markov model with transition probability matrix

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}, \qquad \pi_{ij} = P\left(\Psi_{t,\alpha}(\theta_0) = j \,|\, \Psi_{t-1,\alpha}(\theta_0) = i\right), \quad i,j = 0,1.$$

The approximate joint likelihood for $\{\Psi_{t,\alpha}(\theta_0)\}_{t=2}^n$, conditional on the first observation, is

$$L(\Pi; \{\Psi_{t,\alpha}(\theta_0)\}_{t=2}^n) = (1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}},$$

where $n_{ij} = \sum_{t=2}^{n} 1\left(\Psi_{t,\alpha}(\theta_0) = j \,|\, \Psi_{t-1,\alpha}(\theta_0) = i\right)$. The maximum likelihood under the Markov model is

$$L(\hat{\Pi}; \{\Psi_{t,\alpha}(\theta_0)\}_{t=2}^n) = (1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}},$$

with

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}} \qquad and \qquad \hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}}.$$

Under $H_{0c}$ the maximum likelihood is given by $L(\hat{\pi}; \{\Psi_{t,\alpha}(\theta_0)\}_{t=2}^n)$ as in the unconditional test. The conditional version of Chirstoffersen's (1998) $LR$ test is then based on

$$LR_c = -2log\frac{L(\widehat{\pi}; \{\Psi_{t,\alpha}(\theta_0)\}_{t=2}^n)}{L(\widehat{\Pi}; \{\Psi_{t,\alpha}(\theta_0)\}_{t=2}^n)}.$$

This statistic under the null hypothesis $H_{0c}$ is asymptotically distributed as a $\chi^2$ distribution with

[2] The $LR_u$ in Chirstoffersen's (1998) is monotone in the sufficient statistic $n^{-1/2}\sum_{t=1}^n(\Psi_{t,\alpha}(\theta_0)-\alpha)$. This implies that the null hypothesis $H_{0u}$ is rejected at $\tau$% nominal level when $n^{-1}|\sum_{t=1}^n(\Psi_{t,\alpha}(\theta_0)-\alpha)/\sqrt{\alpha(1-\alpha)}|^2 > \chi^2_{1,\tau}$, where $\chi^2_{1,\tau}$ is the $\tau-$critical value of the $\chi^2_1$ distribution.

one degree of freedom.

More generally, tests for $H_{0c}$ can be based on the autocovariances

$$\gamma_j = Cov(\Psi_{t,\alpha}(\theta_0), \Psi_{t-j,\alpha}(\theta_0)) \qquad j \geq 1, \tag{6}$$

at different lags $j$, which can be consistently estimated under $H_{0c}$ (and $H_{0u}$) by

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

In fact, Berkowitz, Christoffersen and Pelletier (2006) discuss tests based on the sequence of sample autocovariances $\{\gamma_{n,j}\}$. In particular these authors discuss Portmanteau tests in the spirit of those proposed by Box and Pierce (1970) and Ljung and Box (1978), given by

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

It can be proved that $LB(m)$ is asymptotically distributed as a $\chi^2_m$ (cf. Theorem 2.) Note however that $LB(m)$ was proposed by Ljung and Box (1978) for ARMA models, for which the distribution of $LB(m)$ can be *approximated* by that of $\chi^2_m$ assuming the $m$ is large enough, and it is not clear then if this approximation is appropriate in the present framework. Section 3 proposes an alternative Portmanteau test in the spirit of $LB(m)$. In all of these cases, the choice of $m$ affects the performance of the test statistics. Berkowitz, Christoffersen and Pelletier (2006) choose $m = 5$ in their simulations for $LB(m)$.

To take into account all the lags in the sequence $\{\gamma_{n,j}\}$, Berkowitz, Christoffersen and Pelletier (2006) recommend the use of *mds* tests of Durlauf (1991). This author proposed a spectral distribution based test using the fact that the standardized spectral distribution function of a *mds* is a straight line. Durlauf's (1991) Cramer-von Mises test in the present context is

$$DUR_n = \sum_{j=1}^{n-1} n\left(\frac{\gamma_{n,j}}{\alpha(1-\alpha)}\right)^2\left(\frac{1}{j\pi}\right)^2.$$

The 10%, 5% and 1% asymptotic critical values for $DUR_n$ are obtained from Shorack and Wellner (1986, p.147) and are 0.347, 0.461 and 0.743 respectively. Tests for (2) rather than for $H_{0c}$ can be also based on spectral *mds* tests, see Escanciano and Velasco (2006) for such generalizations in the context of dynamic mean models.

Other backtesting procedures have been proposed. For instance, Christoffersen and Pelletier (2004) and Haas (2005) apply duration-based tests to the problem of assessing VaR forecast accuracy, see also Danielsson and Morimoto (2000). Berkowitz, Christoffersen and Pelletier (2006) compare the aforementioned and other backtesting procedures via some Monte Carlo experiments. Their conclusions are that the test of Engle and Manganelli (2004) performs best overall but duration-based tests also perform well in many cases. However, these conclusions have to be considered with caution given the unaccounted presence of estimation risk in such analyses that invalidates inferences based on unconditional backtesting procedures. In other words, in these Monte Carlo experiments one cannot distinguish a good power performance from a large power distorsion due to the estimation effect, *i.e.* real vs spurious power.

An important limitation of all of the aforementioned backtesting techniques is the assumption of the parameter $\theta_0$ being known. In practice however, the parameter $\theta_0$ is unknown and must be estimated from a sample $\{Y_t, I'_{t-1}\}_{t=1}^n$ by a $\sqrt{n}-$consistent estimator, say $\theta_n$. The standard approach in the literature consists on performing relevant inferences replacing $\theta_0$ by the estimator $\theta_n$. We stress in this article that this method of testing leads to invalid inferences in unconditional backtesting procedures, which in turn may imply higher levels of idle capital on the bank than required by the Basel Accord, or lower levels leading to understating risk exposure. We shall show that the introduction of $\theta_n$, *i.e.* uncertainty about $\theta_0$ coming from the data, adds an additional term in the unconditional backtesting procedures that must be taken into account to construct valid inferences on VaR diagnostics. This term is the so-called *estimation risk*. The purposes of the following sections are: first to quantify the estimation risk in the most popular backtesting procedures, and second, to propose a correction of this method free of estimation risk. This is detailed in the next section.

# 3 Backtesting procedures free of estimation risk

## 3.1 The unconditional composite hypothesis

In this section we study the effect of estimation risk when testing the unconditional *composite* hypothesis

$$H_{0u} : E[\Psi_{t,\alpha}(\theta_0)] = \alpha \qquad \text{for some } \theta_0 \in \Theta.$$

A natural two-sided test for $H_{0u}$ is based on rejecting for large values of $|S_n|$, where

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\Psi_{t,\alpha}(\theta_n) - \alpha),$$

and $\theta_n$ is a $\sqrt{n}-$consistent estimator of $\theta_0$. One-sided tests can also be considered and are based on rejecting for large positive (negative) values of $S_n$.

The next theorem quantifies the effect of the estimation risk in $S_n$. In order to see this we need some notation and assumptions. Define the family of conditional distributions

$$F_x(y) := P(Y_t \leq y \mid I_{t-1} = x), \tag{7}$$

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 $m_\alpha(I_{t-1}, \theta)$ is continously differentiable in $\theta$ (a.s.) with derivative $g_\alpha(I_{t-1}, \theta)$ such that $E\left[\sup_{\theta \in \Theta_0} |g_\alpha(I_{t-1}, \theta)|^2\right] < C$, for a neighborhood $\Theta_0$ of $\theta_0$.

**Assumption A4**: The parametric space $\Theta$ is compact in $\mathbb{R}^p$. The true parameter $\theta_0$ belongs to the interior of $\Theta$. The estimator satisfies $\sqrt{n}(\theta_n - \theta_0) = O_P(1)$.

Assumption A1 is made for simplicity in the exposition. Our results are also valid for some non-stationary and non-ergodic sequences, see Escanciano (2006) for details. See Hamilton (1994, p. 46) for a reference on ergodicity. A2 is required as in Koul and Stute (1999). Assumption A3 is classical in inference on nonlinear models, see Koul (2002) monograph. A3 is satisfied for most of the models considered in the literature under mild moment assumptions. A4 has been established in the literature under a variety of conditions and different models and data generating processes (DGP). See references below. With these assumptions in place we are in position to establish the first important result of the paper.

Theorem 1: *Under Assumptions A1-A4,*

$$\begin{aligned} S_n &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\Psi_{t,\alpha}(\theta_n) - \alpha] = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\Psi_{t,\alpha}(\theta_0) - F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right] && (8)\\ &+\underbrace{\sqrt{n}(\theta_n - \theta_0)' E\left[g_\alpha(I_{t-1},\theta_0) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right]}_{\text{Estimation Risk}} && (9)\\ &+\underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0)) - \alpha\right]}_{\text{Model Risk}} + o_P(1). && (10) \end{aligned}$$

Theorem 1 quantifies both *estimation risk* and *model risk* in the unconditional coverage test 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 all the aforementioned papers dealing with estimation risk, see e.g. Christoffersen et al (2001), Chirstoffersen and Gonçalves (2004) or Giacomini and Komunjer (2005). These mixing assumptions are difficult to verify in practice and are not satisfied for some simple models.

Under correct specification of the parametric VaR model, *i.e.* $F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0)) = \alpha$ *a.s.*, expression (10) cancels out and model risk vanishes. In contrast, under misspecification, even if $H_{0u}$ holds, *i.e.* $E\left[F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right] = \alpha$, model risk does not vanish and has a non-negligible effect on the unconditional test. This test in turn, as an specification test of the VaR model, is inconsistent given it has no power against such misspecifications. On the other hand if $E\left[F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right] \neq \alpha$, under some regularity conditions, Theorem 1 yields that

$$\frac{1}{n}\sum_{t=1}^{n}[\Psi_{t,\alpha}(\theta_n)-\alpha] \xrightarrow{P} E[F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))-\alpha] \neq 0.$$

In this case the unconditional test is consistent as an 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, but see our findings with the historical simulation method in Section 3.3. Thus we will assume hereafter that $F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0)) = \alpha$ *a.s.* under any of the null hypotheses considered.

The first term in the expansion of Theorem 1 has *mds* summands, so applying a Martingale Central Limit Theorem, see e.g. Hall and Heyde (1980), this term converges to a Gaussian distribution under mild conditions on the DGP. 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 $\theta_n$, model and true DGP. Section 3.3 below considers the two most popular cases, namely, historical simulation and location-scale models.

To simplify notation we write $A := E\left[g_\alpha(I_{t-1},\theta_0)f_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right]$ in the expression for the estimation risk. We further assume that the estimator $\theta_n$ is asymptotically normal (AN) with variance-covariance matrix $V$. Hence, the estimation risk will be AN with covariance $AVA'$. The vector $A$ can be consistently estimated by

$$A_{n,\tau} = -\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\tau}\exp\left[(Y_t - m_\alpha(I_{t-1},\theta_n))/\tau\right]\Psi_{t,\alpha}(\theta_n)g_\alpha(I_{t-1},\theta_n),$$

with $\tau \to 0$ as $n \to \infty$; see Giacomini and Komunjer (2005). Methods for estimating the variance-covariance matrix $V$ are abundant in the literature, including bootstrap techniques. Next corollary summarizes our previous discussion and provides the necessary corrections to carry out valid asymptotic inference for the unconditional test free of estimation risk. But first we need a stronger version of Assumption 4.

**Assumption A4'**: The parametric space $\Theta$ is compact in $\mathbb{R}^p$. The true parameter $\theta_0$ belongs to the interior of $\Theta$. The estimator $\theta_n$ satisfies the asymptotic Bahadur expansion

$$\sqrt{n}(\theta_n-\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}l(Y_t,I_{t-1},\theta_0)+o_P(1),$$

where $l(\cdot)$ is known up to the finite dimensional parameter $\theta_0$, and is such that $E[l(Y_t, I_{t-1}, \theta_0) \mid I_{t-1}] = 0$ a.s. and $V = E[l(Y_t, I_{t-1}, \theta_0) l'(Y_t, I_{t-1}, \theta_0)]$ exists and is positive definite. Moreover, $l(Y_t, I_{t-1}, \theta)$ is continuous (a.s.) in $\theta$ in $\Theta_0$ and $E\left[\sup_{\theta \in \Theta_0} |l(Y_t, I_{t-1}, \theta)|^2\right] \leq C$, where $\Theta_0$ is a small neighborhood around $\theta_0$.

COROLLARY 1: *Under Assumptions A1-A3, A4' and that $F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) = \alpha$ a.s.*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} [\Psi_{t,\alpha}(\theta_n) - \alpha] \xrightarrow{d} N(0, \sigma_c^2),$$

*where*

$$\sigma_c^2 = \alpha(1-\alpha) + 2A\rho + AVA'$$

*with $\rho = E[(\Psi_{t,\alpha}(\theta_0) - \alpha)\, l(Y_t, I_{t-1}, \theta_0)]$. Moreover, $\widehat{\sigma}_c^2 := \alpha(1-\alpha) + 2A_{n,\tau}\rho_n + A_{n,\tau} V_n A'_{n,\tau}$ is a consistent estimator for $\sigma_c^2$ provided $\tau \to 0$ as $n \to \infty$, with*

$$\rho_n = \frac{1}{n} \sum_{t=1}^{n} (\Psi_{t,\alpha}(\theta_n) - \alpha)\, l(Y_t, I_{t-1}, \theta_n)$$

*and*

$$V_n = \frac{1}{n} \sum_{t=1}^{n} l(Y_t, I_{t-1}, \theta_n) l'(Y_t, I_{t-1}, \theta_n).$$

Then valid inference can be accomplished by the *corrected unconditional test* statistic

$$\widetilde{S}_n = \frac{1}{\widehat{\sigma}_c \sqrt{n}} \sum_{t=1}^{n} [\Psi_{t,\alpha}(\theta_n) - \alpha],$$

which converges to a standard normal *r.v.* Alternatively, one can compute Chirstoffersen´s (1998) test $LR_{uc}$ and compute its critical value as $(\widehat{\sigma}_c^2 / \alpha(1-\alpha))\chi^2_{1,1-\tau}$, where $\chi^2_{1,1-\tau}$ is the $1-\tau$ quantile of the $\chi^2_1$ distribution.

## 3.2 The conditional composite hypothesis: Testing serial independence of the exceedances

Now we focus on

$$H_{0c} : \{\Psi_{t,\alpha}(\theta_0)\}_{t=1}^{n} \text{ are } iid \text{ for some } \theta_0 \in \Theta.$$

Several tests for independence have been proposed in the literature as discussed in the previous section. A large class of tests are based on the autocovariances defined in (6). These autocovariances are now estimated by

$$\widetilde{\gamma}_{n,j} = \frac{1}{n-j} \sum_{t=j+1}^{n} (\Psi_{t,\alpha}(\theta_n) - \alpha)(\Psi_{t-j,\alpha}(\theta_n) - \alpha).$$

In the next theorem we shall show that, contrary to $S_n$, the estimation of $\theta_0$ in $\widetilde{\gamma}_{n,j}$ has no assymptotic effect in the asymptotic null distribution of the test statistics. In other words, there is no estimation risk in conditional backtesting tests making use of the sample autocovariances $\widetilde{\gamma}_{n,j}$[3]. To the best of our knowledge this is the first paper in showing such robustness of the conditional backtesting procedures based on $\widetilde{\gamma}_{n,j}$. Intuitively, if we define

$$\widehat{\gamma}_{n,j} = \frac{1}{n-j} \sum_{t=1+j}^{n} \Psi_{t,\alpha}(\theta_n)\Psi_{t-j,\alpha}(\theta_n) - \frac{1}{(n-j)^2} \left\{ \sum_{t=1+j}^{n} \Psi_{t,\alpha}(\theta_n) \right\} \left\{ \sum_{t=1+j}^{n} \Psi_{t-j,\alpha}(\theta_n) \right\},$$

one can prove (see the proof of Theorem 2) that for all fixed $j \geq 1$,

$$\sqrt{n-j}\, |\widehat{\gamma}_{n,j} - \widetilde{\gamma}_{n,j}| = o_P(1).$$

Hence, the estimation risk, if any, of $\widetilde{\gamma}_{n,j}$ will be the same as that of $\widehat{\gamma}_{n,j}$. Now, the estimation risk in $\widehat{\gamma}_{n,j}$ will appear in both of its summands in such way that both contributions will cancel out and the total estimation effect will be vanished asymptotically.

As an alternative to $LB(m)$, we define the test statistic $C_{n,m}$ as

$$C_{n,m} = \sum_{j=1}^{m} \left( \frac{\widetilde{\gamma}_{n,j}}{\alpha(1-\alpha)} \right)^2,$$

to test the null hypothesis of serial independence.

[3]However, one can show that if instead of $\widetilde{\gamma}_{n,j}$ one uses

$$\widetilde{\gamma}^{(2)}_{n,j} = \frac{1}{n-j} \sum_{t=j+1}^{n} \Psi_{t,\alpha}(\theta_n)\Psi_{t-j,\alpha}(\theta_n) - \alpha^2,$$

then the estimation risk will appear as twice the estimation risk of $S_n$. See the proof of Theorem 2 for details. Therefore the choice between $\widetilde{\gamma}_{n,j}$ and $\widetilde{\gamma}^{(2)}_{n,j}$ for conditional backtesting is not without importance.

Theorem 2: *Under Assumptions A1-A4, for all fixed $j \geq 1$,*

$$\sqrt{n-j}\,|\widetilde{\gamma}_{n,j} - \gamma_{n,j}| = o_P(1).$$

*As a consequence under $H_{0c}$, $C_{n,m} \to_d \chi^2_m$.*

Similar results hold for the composite version of $DUR_n$. We omit such extensions for the sake of space. The next subsection discusses the existence of estimation risk and the corrections introduced here for two well known examples in the VaR literature. These are historical simulation and location-scale models.

### 3.3 Examples

#### 3.3.1 Historical Simulation

The historical simulation VaR is simply the unconditional quantile of $Y_t$. Hence the postulated model is $m_\alpha(I_{t-1}, \theta_0) = \theta_0 \equiv F_Y^{-1}(\alpha)$, where $F_Y^{-1}(\alpha)$ denotes the unconditional quantile function of $Y_t$ evaluated at $\alpha$. Let $F_Y(x)$ be the *cdf* of $Y_t$. In a forecast framework the estimator of $\theta_0$ is usually a rolling estimator based on the last $m$ observations as $\theta_{t,m} = F_{t,m,Y}^{-1}(\alpha)$, where $F_{t,m,Y}^{-1}(\alpha)$ is the empirical quantile function of $\{Y_s\}_{s=t-m}^{t-1}$. For simplicity of exposition, we assume $m = n$ and a fixed estimator $\theta_{t,n} \equiv \theta_n = F_{n,Y}^{-1}(\alpha)$ for all $1 \leq t \leq n$, where $F_{n,Y}^{-1}(\alpha)$ is the empirical quantile function of $\{Y_t\}_{t=1}^n$. The arguments for the general case are similar but with a cumbersome notation. Under some mild assumptions (see e.g. Wu, 2005)

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{-1}{f_Y(\theta_0)} \right) (\Psi_{t,\alpha}(\theta_0) - \alpha) + o_P(1),$$

where $f_Y$ is the density function of $Y_t$. For this example, $g_\alpha(I_{t-1}, \theta_0) \equiv 1$, and the quantities in Corollary 1 reduce to

$$A = E\left[f_{I_{t-1}}(\theta_0)\right], \qquad \rho = \frac{-\alpha(1-\alpha)}{f_Y(\theta_0)}, \qquad V = \frac{\alpha(1-\alpha)}{f_Y^2(\theta_0)}.$$

In this example $l(Y_t, I_{t-1}, \theta_0)$ in A4' is unknown (i.e. involves the unknown density $f_Y(\theta_0)$), but it can be estimated using smoothed nonparametric estimation. See *e.g.* Koul (2002).

It is worth mentioning that the unconditional backtesting procedure based on historical simulation VaR will be inconsistent in all directions. This is so because it is always true that $E\left[F_{I_{t-1}}(\theta_0)\right] = \alpha$ with $\alpha = F_Y(F_Y^{-1}(\alpha))$ regardless the model is correctly specified or not. In other words, under the alternative hypothesis of model misspecification the summands in the model risk term of the expansion in Theorem 1 are always centered and hence, its contribution to the power of the test is always bounded (in probability) under certain weak dependence assumptions in the data. As a by-product of this analysis we claim that the unconditional backtesting test is not appropriate for testing the correct specification of the historical simulation VaR. We stress that the problem is not of the historical simulation, which may or may not be correctly specified, but of the use of the unconditional test as a diagnostic test for the historical simulation model.

#### 3.3.2 Location-scale models.

Now we confine ourselves to consider the VaR parametric model derived from the location-scale model. This parametric approach has been the most popular in attempting to describe the dynamics of the VaR measure. These models are defined as

$$Y_t = \mu(I_{t-1}, \beta_0) + \sigma(I_{t-1}, \beta_0)\varepsilon_t, \tag{11}$$

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

$$m_\alpha(I_{t-1}, \theta_0) = \mu(I_{t-1}, \beta_0) + \sigma(I_{t-1}, \beta_0)F_\varepsilon^{-1}(\alpha), \tag{12}$$

where $F_\varepsilon^{-1}(\alpha)$ denotes a univariate quantile function of $\varepsilon_t$ and the nuisance parameter is $\theta_0 = (\beta_0, F_\varepsilon^{-1}(\alpha))$. Among 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 $\beta_0$ is usually estimated by the Quasi-Maximum Likelihood Estimator (*QMLE*). See Li, Ling and McAleer (2002) for a review of estimators for $\beta_0$. The second component of $\theta_0$, $F_\varepsilon^{-1}(\alpha)$, 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. All these methods are reviewed in Kuester et al. (2006). In any case, $\sqrt{n}$-consistent estimation of $F_\varepsilon^{-1}(\alpha)$ is usually achieved.

For these models our Theorem 1 allows us to quantify estimation risk. It takes this form

$$\sqrt{n}(F_{\varepsilon,n}^{-1}(\alpha) - F_\varepsilon^{-1}(\alpha))f_\varepsilon(F_\varepsilon^{-1}(\alpha)) + \sqrt{n}(\beta_n - \beta_0)'b(\alpha, \beta_0), \tag{13}$$

where

$$b(\alpha, \beta_0) := f_\varepsilon(F_\varepsilon^{-1}(\alpha))E\left[a_{1,t}(\beta_0)\right] + f_\varepsilon(F_\varepsilon^{-1}(\alpha))F_\varepsilon^{-1}(\alpha)E\left[a_{2,t}(\beta_0)\right], \tag{14}$$

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

with $\dot{\mu}_t(\beta) = \partial\mu(I_{t-1}, \beta)/\partial\beta$ and $\dot{\sigma}_t(\beta) = \partial\sigma(I_{t-1}, \beta)/\partial\beta$.

There are two sources of estimation risk in this model, one from estimating $F_\varepsilon^{-1}(\alpha)$ and other resultant from estimating $\beta_0$[4]. If the cumulative distribution function, $F_\varepsilon(x)$, is (strictly) convex in the left tail, as is the case for most assumed distributions, the density function $f_\varepsilon(F_\varepsilon^{-1}(\alpha))$, is (strictly) increasing in $\alpha$, and it is clear from (14) that the estimation risk from this term is therefore increasing in $\alpha$. For the term $b(\alpha, \beta_0)$ the effect of estimation risk as a function of $\alpha$ is more involved. It depends both on the density function of the error distribution and also on the VaR parameter of the error term itself, that is negative in the left tail. Thus, estimation risk will be increasing in $\alpha$ if the following condition holds:

$$f_\varepsilon'(F_\varepsilon^{-1}(\alpha))E[a_{1,t}(\beta_0)] + \left[f_\varepsilon'(F_\varepsilon^{-1}(\alpha))F_\varepsilon^{-1}(\alpha) + f_\varepsilon(F_\varepsilon^{-1}(\alpha))\right]E[a_{2,t}(\beta_0)] > 0,$$

where $f_\varepsilon'(x)$ denotes the first derivative of $f_\varepsilon(x)$. This result is obtained by a simple application of the Chain Rule,

$$\frac{\partial f_\varepsilon(F_\varepsilon^{-1}(\alpha))}{\partial\alpha} = \frac{f_\varepsilon'(F_\varepsilon^{-1}(\alpha))}{f_\varepsilon(F_\varepsilon^{-1}(\alpha))},$$

and assuming that the density function $f_\varepsilon(x)$ is strictly positive at $x = F_\varepsilon^{-1}(\alpha)$.

[4]For zero-mean ARMA homokedastic models only the estimation effect coming from the innovation's quantile estimator remains, i.e. the estimation effect from the mean parameters estimates vanishes.

For illustration purposes we plot $f_\varepsilon(F_\varepsilon^{-1}(\alpha))$ as a function of $\alpha$ and of the degrees of freedom of a Student-t distribution driving the error term.

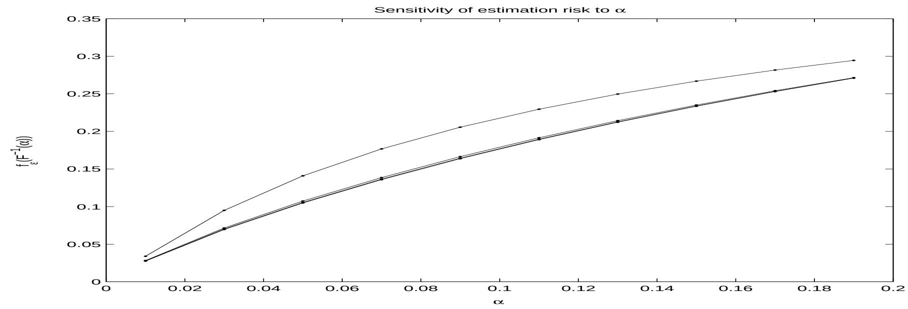


**Figure 3.1.** *Sensitivity of the estimation risk to coverage probability $\alpha$. Different error distributions within the Student-t family are generated. The only significant difference is for $t_5$ that exhibits higher estimation risk than the other distributions studied: $t_{10}, t_{20}, t_{30}$.*

The graphical analysis of the Student-t family in figure 3.1 seems to indicate that the degree of heaviness of the tail of the error distribution does not seem to have a significant effect on the estimation risk unless the distribution is very heavy tailed.

Now, we proceed to analyze one of the most used processes for modelling financial returns: ARMA(1,1)-GARCH(1,1) model with Student-t distributed innovations. This model is defined as

$$Y_t = aY_{t-1} + bu_{t-1} + u_t, \qquad u_t = \sigma(I_{t-1}, \beta_0)\varepsilon_t, \qquad \sigma^2(I_{t-1}, \beta_0) = \eta_{00} + \eta_{10}u_{t-1}^2 + \eta_{20}\sigma^2(I_{t-1}, \beta_0),$$

where $\{\varepsilon_t\}$ are *iid* $t_\nu$ standardized disturbances (i.e. $\varepsilon_t = (\sqrt{(\nu-2)/\nu})v_t$, with $v_t$ distributed as a Student-t with $\nu$ degrees of freedom), the true parameters are $\beta_0 = (a, b, \eta_{00}, \eta_{10}, \eta_{20}) \in \Theta$, with

$$\Theta \subset \{(a, b, \eta_0, \eta_1, \eta_2) \in \mathbb{R}^5 : |a| < 1, \eta_0 > 0, \eta_j \geq 0, j = 1 \text{ and } 2, \eta_1 + \eta_2 < 1\}.$$

The formulas developed here will be useful for the simulation experiments in next section. Assuming that the quantile of $\varepsilon_t$ is not estimated, the first term in (13) does not appear, and the estimation risk boils down to

$$\sqrt{n}(\beta_n - \beta_0)'A,$$

where $A$ is defined as $b(\alpha, \beta_0)$ in (14) but with the Student-$t_\nu$ standardized disturbances density and distribution, denoted by $\varphi_\varepsilon$ and $\Phi_\varepsilon$, replacing $f_\varepsilon$ and $F_\varepsilon$. Denote by $\gamma = (a, b)'$, $\eta_0 = (\eta_{00}, \eta_{10}, \eta_{20})'$, $u_t(\gamma) = Y_t - aY_{t-1} - bu_{t-1}(\gamma)$ and $\mu(I_{t-1}, \beta_0) = Y_t - u_t(\gamma)$. To simplify notation write $\mu_t(\beta) = \mu(I_{t-1}, \beta)$ and $\sigma_t(\beta) = \sigma(I_{t-1}, \beta)$.

Here we consider the *QMLE* as the estimator $\beta_n$ of $\beta_0$. As shown in *e.g.* Francq and Zakoïan (2004), the QMLE of an ARMA(1,1)-GARCH(1,1) model satisfies A4' undel mild conditions and $H_0$, with

$$\sqrt{n}(\beta_n - \beta_0) = \frac{1}{\sqrt{n}} J^{-1} \sum_{t=1}^{n} \frac{\partial \ell_t(\beta_0)}{\partial \beta} + o_P(1),$$

where $J \equiv J(\theta_0) = E[\partial^2 \ell_t(\beta_0)/\partial\beta\partial\beta']$ and $\ell_t(\beta_0) = -\frac{1}{2}\log 2\pi - \frac{1}{2}u_t^2(\gamma)/\sigma_t^2(\beta_0) - \frac{1}{2}\log(\sigma_t^2(\beta_0)).$ When the innovation distribution is symmetric, as in the Student-t or Gaussian cases, the matrices $J$ and $V$ in A4' are block diagonal. We partition $J$ and $V$ according to $\beta_0 = (\gamma', \eta_0')'$ and denote

$$J = \begin{bmatrix} J_\gamma & 0 \\ 0 & J_\eta \end{bmatrix}, \qquad V = \begin{bmatrix} V_\gamma & 0 \\ 0 & V_\eta \end{bmatrix}.$$

Note that the derivatives of the score $\partial \ell_t(\beta_0)/\partial\beta$ are

$$\begin{aligned} \frac{\partial \ell_t(\beta_0)}{\partial \gamma} &= -\frac{1}{2}\left(1 - \frac{u_t^2(\gamma)}{\sigma_t^2(\beta_0)}\right) \frac{\partial \sigma_t^2(\beta_0)}{\partial \gamma} \frac{1}{\sigma_t^2(\beta_0)} - \frac{u_t(\gamma)}{\sigma_t^2(\beta_0)} \frac{\partial u_t(\gamma)}{\partial \gamma} \quad \text{and} \\ \frac{\partial \ell_t(\beta_0)}{\partial \eta} &= -\frac{1}{2}\left(1 - \frac{u_t^2(\gamma)}{\sigma_t^2(\beta_0)}\right) \frac{\partial \sigma_t^2(\beta_0)}{\partial \eta} \frac{1}{\sigma_t^2(\beta_0)}. \end{aligned}$$

Define

$$A_\mu = -\frac{1}{2} E[\frac{\partial u_t(\gamma_0)}{\partial \gamma} \frac{1}{\sigma_t(\beta_0)}], \qquad A_\sigma = -\frac{1}{2} E[\frac{\partial \sigma_t^2(\beta_0)}{\partial \beta} \frac{1}{\sigma_t^2(\beta_0)}],$$

and write accordingly $A_\sigma = (A'_{\sigma\gamma}, A'_{\sigma\eta})'$.

Analogously, we write $\rho = (\rho'_\gamma, \rho'_\eta)'$, where

$$\begin{aligned} \rho &= \begin{bmatrix} J_\gamma^{-1} \{\rho_0 A_{\sigma\gamma} + \rho_1 A_\mu\} \\ J_\eta^{-1} A_{\sigma\eta} \rho_0 \end{bmatrix}, \text{ with} \\ \rho_0 &= E[(\Psi_{t,\alpha}(\theta_0) - \alpha)\left(1 - \frac{u_t^2(\gamma)}{\sigma_t^2(\beta_0)}\right)], \qquad \rho_1 = E[(\Psi_{t,\alpha}(\theta_0) - \alpha)\frac{u_t(\gamma)}{\sigma_t(\beta_0)}]. \end{aligned}$$

The derivatives in $A$ and other quantities can be computed as

$$
\begin{aligned}
\frac{\partial \mu_t(\beta_0)}{\partial a} &= -\frac{\partial u_t(\gamma)}{\partial a} = Y_{t-1} + b\frac{\partial u_{t-1}(\gamma)}{\partial a} = \sum_{j=1}^{\infty}(-1)^{j+1}b^{j-1}Y_{t-j},\\
\frac{\partial \mu_t(\beta_0)}{\partial b} &= -\frac{\partial u_t(\gamma)}{\partial b} = u_{t-1}(\gamma) + b\frac{\partial u_{t-1}(\gamma)}{\partial b} = \sum_{j=1}^{\infty}(-1)^{j+1}b^{j-1}u_{t-j}(\gamma),\\
\frac{\partial \sigma_t^2(\beta_0)}{\partial a} &= 2\eta_{10}u_{t-1}(\gamma)\frac{\partial u_{t-1}(\gamma)}{\partial a} + \eta_{20}\frac{\partial \sigma_{t-1}^2(\beta_0)}{\partial a} = 2\eta_{10}\sum_{j=1}^{\infty}\eta_{20}^{j-1}u_{t-j}(\gamma)\frac{\partial u_{t-j}(\gamma)}{\partial a},\\
\frac{\partial \sigma_t^2(\beta_0)}{\partial b} &= 2\eta_{10}u_{t-1}(\gamma)\frac{\partial u_{t-1}(\gamma)}{\partial b} + \eta_{20}\frac{\partial \sigma_{t-1}^2(\beta_0)}{\partial b} = 2\eta_{10}\sum_{j=1}^{\infty}\eta_{20}^{j-1}u_{t-j}(\gamma)\frac{\partial u_{t-j}(\gamma)}{\partial b},\\
\frac{\partial \sigma_t^2(\beta_0)}{\partial \eta_{00}} &= (1-\eta_{20})^{-1}, \qquad \frac{\partial \sigma_t^2(\beta_0)}{\partial \eta_{10}} = u_{t-1}^2(\gamma) + \eta_{20}\frac{\partial \sigma_{t-1}^2(\beta_0)}{\partial \eta_{10}} = \sum_{j=1}^{\infty}\eta_{20}^{j-1}u_{t-j}^2(\gamma),\\
\frac{\partial \sigma_t^2(\beta_0)}{\partial \eta_{20}} &= \sigma_{t-1}^2(\beta_0) + \eta_{20}\frac{\partial \sigma_{t-1}^2(\beta_0)}{\partial \eta_{20}} = \sum_{j=1}^{\infty}\eta_{20}^{j-1}\sigma_{t-j}^2(\beta_0).
\end{aligned}
$$

With these quantities, and noting that

$$
\frac{\partial \sigma_t(\beta_0)}{\partial \beta} = \frac{1}{2\sigma_t(\beta_0)}\frac{\partial \sigma_t^2(\beta_0)}{\partial \beta},
$$

we compute $A = (A_\gamma', A_\eta')'$ as

$$
A = \begin{bmatrix} -\varphi_\varepsilon(\Phi_\varepsilon^{-1}(\alpha))A_\mu - \varphi_\varepsilon(\Phi_\varepsilon^{-1}(\alpha))\Phi_\varepsilon^{-1}(\alpha)A_{\sigma\gamma} \\ -\varphi_\varepsilon(\Phi_\varepsilon^{-1}(\alpha))\Phi_\varepsilon^{-1}(\alpha)A_{\sigma\eta} \end{bmatrix}.
$$

Expressions for $J_\gamma$, $J_\eta$, $V_\gamma$ and $V_\eta$ are computed as $V_\gamma = J_\gamma^{-1}I_\gamma J_\gamma^{-1}$ and $V_\eta = J_\eta^{-1}I_\eta J_\eta^{-1}$, and

$$
\begin{aligned}
I_\gamma &= \frac{1}{4}(E[\varepsilon_t^4]-1)E\left[\frac{1}{\sigma_t^4(\beta_0)}\frac{\partial \sigma_t^2(\beta_0)}{\partial \gamma}\frac{\partial \sigma_t^2(\beta_0)}{\partial \gamma'}\right] + E\left[\frac{u_t^2(\gamma)}{\sigma_t^4(\beta_0)}\frac{\partial u_t(\gamma)}{\partial \gamma}\frac{\partial u_t(\gamma)}{\partial \gamma'}\right],\\
I_\eta &= \frac{1}{4}(E[\varepsilon_t^4]-1)E\left[\frac{1}{\sigma_t^4(\beta_0)}\frac{\partial \sigma_t^2(\beta_0)}{\partial \eta}\frac{\partial \sigma_t^2(\beta_0)}{\partial \eta'}\right],\\
J_\gamma &= \frac{1}{2}E\left[\frac{1}{\sigma_t^4(\beta_0)}\frac{\partial \sigma_t^2(\beta_0)}{\partial \gamma}\frac{\partial \sigma_t^2(\beta_0)}{\partial \gamma'}\right] + E\left[\frac{u_t^2(\gamma)}{\sigma_t^4(\beta_0)}\frac{\partial u_t(\gamma)}{\partial \gamma}\frac{\partial u_t(\gamma)}{\partial \gamma'}\right],\\
J_\eta &= \frac{1}{2}E\left[\frac{1}{\sigma_t^4(\beta_0)}\frac{\partial \sigma_t^2(\beta_0)}{\partial \eta}\frac{\partial \sigma_t^2(\beta_0)}{\partial \eta'}\right].
\end{aligned}
$$

Note that for $t_\nu$ standardized disturbances

$$
E[\varepsilon_t^4] = \left(\frac{3\nu-6}{\nu-4}\right)\left(\frac{\nu-2}{\nu}\right)^2 \equiv \kappa_\nu.
$$

Hence $I_\eta = ((\kappa_\nu - 1)/2)J_\eta$, so $V_\eta = ((\kappa_\nu - 1)/2)J_\eta^{-1}$. If the distribution is Gaussian we further have that $V_\eta = J_\eta^{-1}$ and $I_\gamma = J_\gamma$ and $V_\gamma = J_\gamma^{-1}$.

The corrected unconditional test uses the variance $\sigma_c^2$ that can be consistently estimated by

$$\widehat{\sigma}_c^2 = \alpha(1-\alpha) + 2\widehat{A}_\gamma \widehat{\rho}_\gamma' + 2\widehat{A}_\eta \widehat{\rho}_\eta + \widehat{A}_\gamma \widehat{V}_\gamma \widehat{A}_\gamma' + \widehat{A}_\eta \widehat{V}_\eta \widehat{A}_\eta',$$

where $\widehat{A}_\gamma$, $\widehat{\rho}_\gamma$, $\widehat{A}_\eta$, $\widehat{\rho}_\eta$, $\widehat{V}_\gamma$, and $\widehat{V}_\eta$ estimate consistently $A_\gamma$, $\rho_\gamma$, $A_\eta$, $\rho_\eta$, $V_\gamma$, and $V_\eta$, respectively. The estimators for $\widehat{V}_\gamma$, and $\widehat{V}_\eta$ are often computed by many statistical packages, and $\widehat{A}_\gamma$, $\widehat{\rho}_\gamma$, $\widehat{A}_\eta$, $\widehat{\rho}_\eta$ are easily obtained by replacing population expectations and $\beta_0$ by sample expectations and $\beta_n$, respectively.

# 4 Simulation Exercise

This section examines the performance through some Monte Carlo experiments of the unconditional test devised in Kupiec (1995) and the unconditional corrected test developed in this paper. We consider different innovations processes $\{\varepsilon_t\}$ within the Student-t family. More concretely, $\varepsilon_t = (\sqrt{(\nu-2)/\nu})v_t$, with $v_t$ distributed as a Student-t with $\nu$ degrees of freedom. In the simulations we first consider a $t_{30}$ distribution as an approximation to the Gaussian distribution and a $t_{10}$ to illustrate the impact of heavier than normal tails in the different backtesting tests.

We consider two blocks of models in the simulations. The first block corresponds to a simple pure location model:

$$Y_t = a + \varepsilon_t, \qquad a \in \mathbb{R},$$

where $a$ is unknown and estimated by the sample mean $a_n = n^{-1}\sum_{t=1}^n Y_t$. The true parameter is $a = 5$. For this model

$$A = \varphi_\varepsilon(\Phi_\varepsilon^{-1}(\alpha)), \qquad \rho = E[(\Psi_{t,\alpha}(\theta_0) - \alpha)\,(Y_t - a)], \qquad V = 1.$$

Hence we estimate $\sigma_c^2$ by $\widehat{\sigma}_c^2 = \alpha(1-\alpha) + 2\varphi_\varepsilon(\Phi_\varepsilon^{-1}(\alpha))\rho_n + \varphi_\varepsilon^2(\Phi_\varepsilon^{-1}(\alpha))$ where

$$\rho_n = \frac{1}{n}\sum_{t=1}^n (\Psi_{t,\alpha}(\theta_n) - \alpha)\,(Y_t - a_n).$$

For this model the estimation risk is distributed as a normal distribution with zero mean and standard deviation 0.0265 ($\alpha = 0.01$) and 0.1011 ($\alpha = 0.05$) for the $t_{30}$ distribution, and 0.0253 ($\alpha = 0.01$) and 0.096 ($\alpha = 0.05$) for the $t_{10}$ case. Hence, we do not expect a large estimation effect in the unconditional backtesting tests for this model. This is confirmed in our simulations below.

A more realistic model for financial data is provided by our second block of simulations; an ARMA(1,1)-GARCH(1,1) process of the form:

$$Y_t = aY_{t-1} + bu_{t-1} + u_t, \qquad u_t = \sigma(I_{t-1}, \beta_0)\varepsilon_t, \qquad \sigma^2(I_{t-1}, \beta_0) = \eta_{00} + \eta_{10}u_{t-1}^2 + \eta_{20}\sigma^2(I_{t-1}, \beta_0),$$

with the true parameters given by $\beta_0 = (a, b, \eta_{00}, \eta_{10}, \eta_{20})' = (0.1, 0.1, 0.05, 0.1, 0.85)$. This process is intended to emulate actual processes describing financial returns.

The Value at Risk of these models is calculated at 1% as recommended by Basel Committee and at 5% to see the effect of increasing the coverage probability. Finally the effect of the error distribution is reflected by simulating a $t_{30}$ and a $t_{10}$ distribution.

For the first model the simulated size of backtesting using $S_n$ and $\widetilde{S}_n$ is reported in Tables 4.1 and 4.2.

| $S_n$ | n=500 | | | n=1000 | | | n=2000 | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.01$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.098 | 0.028 | 0.012 | 0.065 | 0.031 | 0.004 | 0.068 | 0.039 | 0.004 |
| $t_{10}$ | 0.092 | 0.028 | 0.006 | 0.074 | 0.035 | 0.014 | 0.074 | 0.043 | 0.004 |
| $\widetilde{S}_n$ | n=500 | | | n=1000 | | | n=2000 | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.098 | 0.062 | 0.012 | 0.097 | 0.051 | 0.010 | 0.092 | 0.039 | 0.010 |
| $t_{10}$ | 0.092 | 0.062 | 0.016 | 0.115 | 0.054 | 0.017 | 0.092 | 0.046 | 0.007 |

**Table 4.1.** *Size of unconditional tests $S_n$ and $\widetilde{S}_n$ for $\alpha = 0.01$, and $\varepsilon_t$ following a family of $t_\nu$ with $\nu = 30, 10$ for model (1) with $a = 5$. 1000 Monte-Carlo replications.*

| $S_n$ | n=500 | | | n=1000 | | | n=2000 | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.05$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.039 | 0.021 | 0.004 | 0.053 | 0.021 | 0.002 | 0.052 | 0.016 | 0.003 |
| $t_{10}$ | 0.048 | 0.033 | 0.005 | 0.059 | 0.021 | 0.002 | 0.046 | 0.017 | 0.000 |
| $\widetilde{S}_n$ | n=500 | | | n=1000 | | | n=5000 | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.089 | 0.043 | 0.014 | 0.106 | 0.051 | 0.008 | 0.096 | 0.046 | 0.011 |
| $t_{10}$ | 0.092 | 0.054 | 0.021 | 0.100 | 0.057 | 0.013 | 0.090 | 0.048 | 0.007 |

**Table 4.2.** *Size of unconditional tests $S_n$ and $\widetilde{S}_n$ for $\alpha = 0.05$, and $\varepsilon_t$ following a family of $t_\nu$ with $\nu = 30, 10$ for model (1) with $a = 5$. 1000 Monte-Carlo replications.*

The following two tables (4.3) and (4.4) report the simulated sizes corresponding to the ARMA(1,1)-GARCH(1,1) model.

| $S_n$ | n=500 | | | n=1000 | | | n=2000 | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.01$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.07 | 0.012 | 0.004 | 0.041 | 0.016 | 0.002 | 0.035 | 0.018 | 0.001 |
| $t_{10}$ | 0.053 | 0.014 | 0.002 | 0.039 | 0.014 | 0.001 | 0.046 | 0.018 | 0.000 |
| $\widetilde{S}_n$ | n=500 | | | n=1000 | | | n=2000 | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.122 | 0.052 | 0.028 | 0.101 | 0.058 | 0.013 | 0.097 | 0.038 | 0.010 |
| $t_{10}$ | 0.11 | 0.075 | 0.041 | 0.135 | 0.073 | 0.045 | 0.123 | 0.061 | 0.023 |

**Table 4.3.** *Size of unconditional tests $S_n$ and $\widetilde{S}_n$ for $\alpha = 0.01$, and $\varepsilon_t$ following a family of $t_\nu$ with $\nu = 30, 10$ for model (2) with $\beta_0 = (a, b, \eta_{00}, \eta_{10}, \eta_{20})' = (0.1, 0.1, 0.05, 0.1, 0.85)$. 1000 Monte-Carlo replications.*

| $S_n$ | n=500 | | | n=1000 | | | n=2000 | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.05$ | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.011 | 0.003 | 0.001 | 0.016 | 0.005 | 0.000 | 0.013 | 0.001 | 0.000 |
| $t_{10}$ | 0.012 | 0.004 | 0.0 | 0.016 | 0.007 | 0.001 | 0.017 | 0.002 | 0.0 |
| $\widetilde{S}_n$ | n=500 | | | n=1000 | | | n=2000 | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| $t_{30}$ | 0.055 | 0.028 | 0.006 | 0.054 | 0.021 | 0.006 | 0.067 | 0.028 | 0.002 |
| $t_{10}$ | 0.102 | 0.062 | 0.011 | 0.089 | 0.044 | 0.011 | 0.077 | 0.038 | 0.003 |

**Table 4.4.** *Size of unconditional tests $S_n$ and $\widetilde{S}_n$ for $\alpha = 0.05$, and $\varepsilon_t$ following a family of $t_\nu$ with $\nu = 30, 10$ for model (2) with $\beta_0 = (a, b, \eta_{00}, \eta_{10}, \eta_{20})' = (0.1, 0.1, 0.05, 0.1, 0.85)$. 1000 Monte-Carlo replications.*

The conclusions from this simulation experiment are illuminating. For both blocks of simulations, $\widetilde{S}_n$ outperforms $S_n$ in terms of size performance, and the simulated size reported for $\widetilde{S}_n$ is close to the actual nominal values. Although for the model with a mean (model (1)) $S_n$ provides reliable results for $\alpha = 0.01$, for the ARMA(1,1)-GARCH(1,1) model the test statistic is very unreliable in every case. The size distortions in $S_n$ increase with $\alpha$. As the sample size increases the distortions do not vanish, confirming our findings in Theorem 1 claiming that estimation risk appears even asymptotically (*i.e.* with infinite sample sizes.) Other conclusions from the tables are that the distribution error seems to have an effect (the observed sizes for $t_{10}$ double their counterparts for $t_{30}$ in many cases) and the choice of the coverage probability also has an effect, in this case augmenting the estimation risk effect for the Student-$t_{30}$ distribution.

# 5 Application to financial data

In a recent important paper Berkowitz and O'Brien (2002) compared the VaR forecasts obtained from the location-scale family (ARMA-GARCH model) with the internal structural models used by banks. Their conclusion is that the GARCH model generally provides for lower VaRs and is better at predicting changes in volatility, thereby permitting comparable risk coverage with less regulatory capital. The other advantage of this approach is that risk managers do not need to be specialists in each business line and know of the best techniques that suit to the risk profile of each specific business line.

One implication of the results in Berkowitz and O'Brien (2002) would be that risk managers would only need to model the serial dependence found on each business line returns using ARMA-GARCH volatility filters and decide on the distribution fitting the residuals to determine the VaR measure. Moreover, managing the overall risk of the institution would boil down to model multivariate as well as serial dependence between business lines. *BEKK* and *VEC* models of Engle and Kroner (1995), and *CCC* or *DCC* models introduced by Bollerslev (1990) and Engle (2002) respectively are successful at extending the univariate GARCH methodology to a multivariate setting and would be natural candidates to assess the overall amount of risk accounting for business lines dependencies.

These findings, though interesting, may be spurious if the backtesting procedure employed does not take into account estimation risk effects. In fact, we have uncovered in this paper that the standard backtesting techniques produce wrong critical values to assess VaR estimates from parametric models as those used in Berkowitz and O'Brien (2002). In particular we show in this application that Kupiec's method understates risk exposure. This is tested for daily returns on $S\&P500$ market-valued equity Index over the period 02/2000 - 11/2006 (n=1706 observations), and are obtained from *Freelunch.com*. In order to detect if the VaR measure understates or overstates risk exposure we report a version of the test statistics $S_n$ and $\widetilde{S}_n$ without absolute values. The VaR measures correspond to daily data fitted to an ARMA(1,1)-GARCH(1,1) model estimated by *QMLE*. VaR measures are calculated assuming first the error term being Gaussian and then being distributed as a Student-t with 10 degrees of freedom. The results for $LR_u$ are not presented for sake of space but are consistent with those for $S_n$.

The following plot reports daily $VaR_{0.01}$ estimates over the relevant period. The procedure followed is as suggested by Basel Accord (1996b) but using in-sample and not out-of-sample backtesting; thus we consider windows of 250 daily observations to compute daily $VaR_{0.01}$ from an ARMA(1,1)-GARCH(1,1) model and count the number of in-sample exceedances. For ease of exposition we re-estimate the parameters and test the validity of the model every five days (and not daily) during the period of interest.

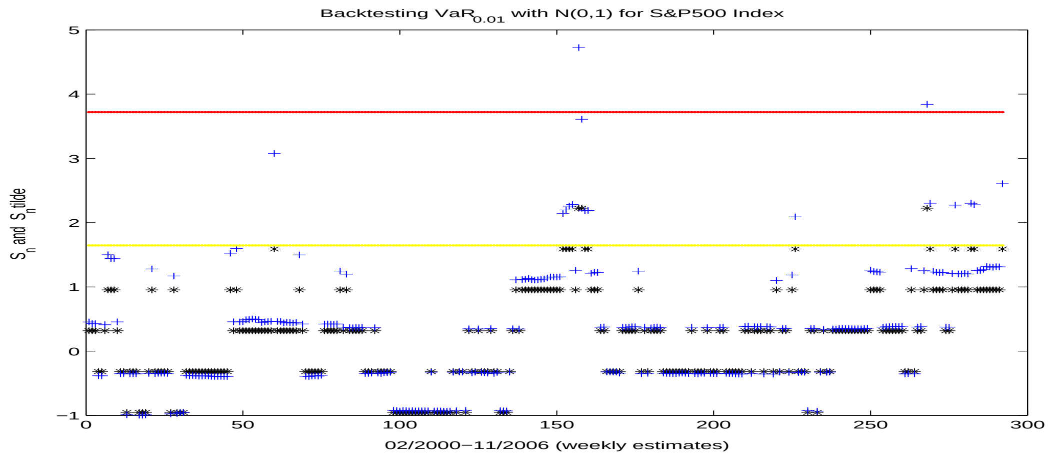


**Figure 5.1.** *$VaR_{0.01}$ estimates over 5-day rolling windows of 250 daily observations from an ARMA(1,1)-GARCH(1,1) model with Gaussian error ($N(0,1)$) that is re-estimated every five days. The data are returns on S&P500 Index over the period 02/2000-11/2006. The yellow (inferior) straight line defines the lower limit of the yellow zone. The red (superior) line denotes the lower limit of the red zone. (+) is used to denote $\widetilde{S}_n$ test statistic and (*) for $S_n$.*

The plot is really conclusive. While the standard procedure reports two periods where the ARMA(1,1)-GARCH(1,1) model lies in the yellow zone the corrected backtesting during the same time intervals reports massive warnings of model failure. More importantly, the red zone, defined by Basel Accord in values exceeding 99.99% coverage probability, is exceeded two times by $\widetilde{S}_n$. In terms of capital requirements this would imply a multiplication factor of 4 rather than a value of 3 or 3.40 (in the worst case) as the standard backtesting would be indicating. [5]

These findings point towards either rejection of the dynamic parametric model or rejection of the Gaussian distribution. The latter would imply that the square root of time used to estimate $VaR_{0.01}$ for a 10-day holding period is not correctly specified. The following plot reports the corresponding estimates assuming a Student-$t_{10}$ distribution.

[5] The capital requirements $CR_t$ required by Basel Accord (1996b) is calculated as

$$CR_t = mf_t \times VaR_{0.01,t},$$

with $mf_t$ a multiplication factor of 3 if backtesting reports a green zone (area below yellow zone) and 4 if it reports a red zone.

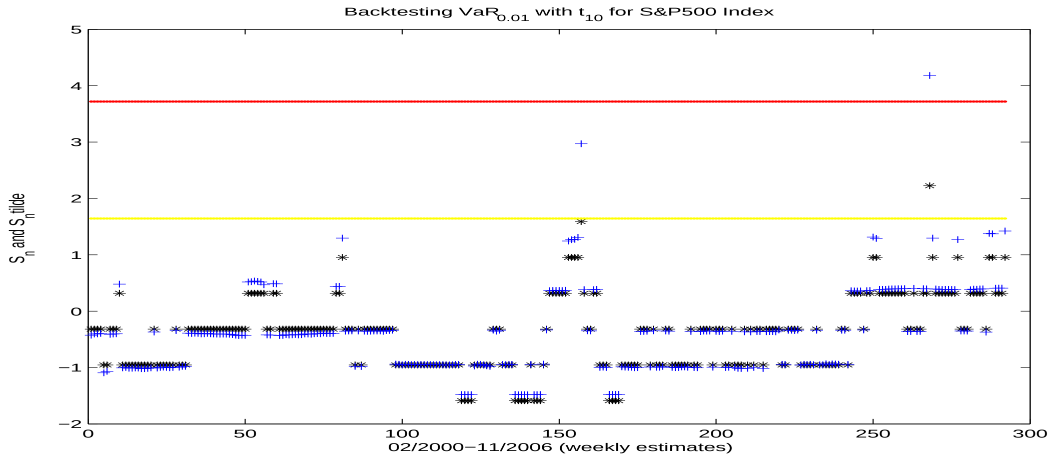


**Figure 5.2.** *$VaR_{0.01}$ estimates over 5-day rolling windows of 250 daily observations from an ARMA(1,1)-GARCH(1,1) model with Student-$t_{10}$ distribution error that is re-estimated every five days. The data are returns on S&P500 Index over the period 02/2000-11/2006. The yellow (inferior) straight line defines the lower limit of the yellow zone. The red (superior) line denotes the lower limit of the red zone. (+) is used to denote $\widetilde{S}_n$ test statistic and (*) for $S_n$.*

The choice of an error distribution with heavier tails responds better to backtesting risk monitoring. The number of yellow zones warnings is dramatically inferior and there is only one red zone warning. This finding is consistent with current literature on financial time series modelling where returns are assumed to exhibit conditional heavier tails than Gaussian along with conditional heteroscedasticity.

# 6 Conclusion

The implementation of the risk management techniques derived from the Basel Accord are at the center of current discussion between European banks and regulators and their American counterparts. While European institutions welcome this new framework American regulators are more cautious about the success of these risk measures. They argue that the Accord relies too heavily on banks' internal models and intend to propose a number of extra safeguards to keep capital requirements higher.

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 correct specification of these backtesting procedures is of paramount importance for the reliability of the whole internal and external monitoring process. However we have shown in this paper that the standard unconditional backtesting used by banks and regulators to assess dynamic parametric VaR estimates is misleading. 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 statistic.

The importance of this correction is shown in an empirical application for financial returns on $S\&P500$ Index. We find that the standard backtesting procedures failed to report red zones warnings that imply dramatic implications on extra capital requirements for financial institutions. We also find that the ARMA(1,1)-GARCH(1,1) model with $t_{10}$ distribution error performs much better in terms of dynamic risk management than the normal distribution, and as a result the square root of time to compute risk measures for longer than daily time intervals should be applied with extra caution if not refused.

These findings 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.

Finally we observe from our analysis of the effects of estimation risk on backtesting that the standard methods for conditional backtesting (testing the presence of serial dependence in the sequence of VaR exceedances) are free from this effect. We also note that for the widely used historical simulation VaR the use of unconditional backtesting is not appropriate to discriminate between models correctly specified from those that are not, for the method has no statistical power against this VaR estimation technique.

## Appendix: Mathematical Proofs

We prove Theorem 1 using empirical processes theory. First, for the shake of exposition we shall state a weak convergence theorem which appears in Delgado and Escanciano (2006), and which is crucial for the subsequent asymptotic results. Let for each $n \geq 1$, $I'_{n,0}, ..., I'_{n,n-1}$, be an array of random vectors in $\mathbb{R}^p$, $p \in \mathbb{N}$, and $Y_{n,1}, ..., Y_{n,n}$, be an array of real random variables (r.v.'s). Denote by $(\Omega_n, \mathcal{A}_n, P_n)$, $n \geq 1$, the probability space in which all the r.v.'s $\{Y_{n,t}, I'_{n,t-1}\}_{t=1}^n$ are defined. Let $\mathcal{F}_{n,t}$, $0 \leq t \leq n$, be a double array of sub $\sigma$-fields of $\mathcal{A}_n$ such that $\mathcal{F}_{n,t-1} \subset \mathcal{F}_{n,t}$, $t = 1, ..., n$, and such that for each $n \geq 1$ and each $\gamma \in \mathcal{H}$,

$$E[w(Y_{n,t}, I_{n,t-1}, \gamma) \mid \mathcal{F}_{n,t-1}] = 0 \text{ a.s.} \qquad 1 \leq t \leq n, \ \forall n \geq 1. \tag{15}$$

Moreover, we shall assume that $\{w(Y_{n,t}, I_{n,t-1}, \gamma), \mathcal{F}_{n,t}, 0 \leq t \leq n\}$ is a square-integrable real-valued martingale difference sequence for each $\gamma \in \mathcal{H}$, that is, (15) holds, $Ew^2(Y_{n,t}, I_{n,t-1}, \gamma) < \infty$ and $w(Y_{n,t}, I_{n,t-1}, \gamma)$ is $\mathcal{F}_{n,t}$-measurable for each $\gamma \in \mathcal{H}$ and $\forall t, 1 \leq t \leq n, \forall n \in \mathbb{N}$. The following result gives sufficient conditions for the weak convergence of the empirical process

$$\alpha_{n,w}(\gamma) = n^{-1/2} \sum_{t=1}^{n} w(Y_{n,t}, I_{n,t-1}, \gamma) \qquad \gamma \in \mathcal{H}.$$

Under mild conditions the empirical process $\alpha_{n,w}$ can be viewed as a mapping from $\Omega_n$ to $\ell^\infty(\mathcal{H})$, the space of all real-valued functions that are uniformly bounded on $\mathcal{H}$, with $\mathcal{H}$ a generic metric space.

An important role in the weak convergence theorem is played by the conditional quadratic variation (CV) of the empirical process $\alpha_{n,w}$ on a finite partition $\mathcal{B} = \{H_k; 1 \leq k \leq N\}$ of $\mathcal{H}$, which is defined as

$$CV_{n,w}(\mathcal{B}) = \max_{1 \leq k \leq N} n^{-1} \sum_{t=1}^{n} E\left[ \sup_{\gamma_1, \gamma_2 \in H_k} |w(Y_{n,t}, I_{n,t-1}, \gamma_1) - w(Y_{n,t}, I_{n,t-1}, \gamma_2)|^2 \mid \mathcal{F}_{n,t-1} \right]. \tag{16}$$

Then, for the weak convergence theorem we need the following assumptions.

**W1:** For each $n \geq 1$, $\{(Y_{n,t}, I_{n,t-1})' : 1 \leq t \leq n\}$ is a strictly stationary and ergodic process. The sequence $\{w(Y_{n,t}, I_{n,t-1}, \gamma), \mathcal{F}_{n,t}, 1 \leq t \leq n\}$ is a square-integrable martingale difference sequence for each $\gamma \in \mathcal{H}$. Also, there exists a function $C_w(\gamma_1, \gamma_2)$ on $\mathcal{H} \times \mathcal{H}$ to $\mathbb{R}$ such that uniformly in $(\gamma_1, \gamma_2) \in \mathcal{H} \times \mathcal{H}$

$$n^{-1} \sum_{t=1}^{n} w(Y_{n,t}, I_{n,t-1}, \gamma_1) w(Y_{n,t}, I_{n,t-1}, \gamma_2) = C_w(\gamma_1, \gamma_2) + o_{P_n}(1).$$

**W2:** The family $w(Y_{n,t}, I_{n,t-1}, \gamma)$ is such that $\alpha_{n,w}$ is a mapping from $\Omega_n$ to $\ell^\infty(\mathcal{H})$ and for every $\delta > 0$ there exists a finite partition $\mathcal{B}_\delta = \{H_k; 1 \leq k \leq N_\delta\}$ of $\mathcal{H}$, with $N_\delta$ being the elements of such partition, such that

$$\int_0^\infty \sqrt{\log(N_\delta)} d\delta < \infty \tag{17}$$

and

$$\sup_{\delta \in (0,1) \cap \mathbb{Q}} \frac{CV_{n,w}(\mathcal{B}_\delta)}{\delta^2} = O_{P_n}(1). \tag{18}$$

Let $\alpha_{\infty,w}(\cdot)$ be a Gaussian process with zero mean and covariance function given by $C_w(\gamma_1, \gamma_2)$. We are now in position to state the following

Theorem A1: *If Assumptions W1 and W2 hold, then it follows that*

$$\alpha_{n,w} \Longrightarrow \alpha_{\infty,w} \text{ in } \ell^\infty(\mathcal{H}).$$

Proof of Theorem A1: Theorem A1 in Delgado and Escanciano (2006).

To prove Theorem 1 we need a useful lemma and further definitions. Let $(\mathcal{G}, \|\cdot\|_{\mathcal{G}})$ be a subset of a metric space of real-valued functions $g$. The *covering number* $N(\varepsilon, \mathcal{G}, \|\cdot\|_{\mathcal{G}})$ is the minimal number of $N$ for which there exist $\varepsilon$-balls $\{f : \|f - g_j\|_{\mathcal{G}} \leq \varepsilon,\ \|g_j\|_{\mathcal{G}} < \infty,\ j =, 1..., N\}$ to cover $\mathcal{G}$. For the definition of assymptotic tightness see van der Vaart and Wellner (1996). Define the process

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

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-A4, the process $K_n(c)$ is asymptotically tight with respect to $c \in C_K$.*

Proof of Lemma A1: Let us define the class of functions $\mathcal{K} = \{\Psi_{t,\alpha}(\theta_0 + cn^{-1/2}) - F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0 + cn^{-1/2})) : c \in \mathcal{C}_K\}$. Write $w_{t-1}(c) = \Psi_{t,\alpha}(\theta_0 + cn^{-1/2}) - F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0 + cn^{-1/2}))$. From the compactness of $\mathcal{C}_K$ we have

$$\int_0^\infty \sqrt{\log(N(\delta, \mathcal{C}_K, d))}d\delta < \infty,$$

where $d(c_1, c_2) = |c_1 - c_2|^{1/2}$. Let $\mathcal{B}_\delta = \{B_k; 1 \leq k \leq N_\delta \equiv N(\delta, \mathcal{C}_K, d)\}$ be a partition of $\mathcal{C}_K$ in $\delta$-balls with respect to $d$. Thus, (17) holds for such partition. Now we shall prove that also (18) follows for such partition. By A1 and A3, $CV_{n,w}(\mathcal{B}_\delta)$ in (16) is bounded by

$$\begin{aligned}
& \max_{1 \leq k \leq N_\delta} n^{-1} \sum_{t=1}^n E\left[\sup_{c_1, c_2 \in \mathcal{C}_K : d(c_1, c_2) = \delta} |w_{t-1}(c_1) - w_{t-1}(c_2)|^2 \mid \mathcal{F}_{t-1}\right] \\
\leq\ & C \max_{1 \leq k \leq N_\delta} n^{-1} \sum_{t=1}^n \sup_{c_1, c_2 \in \mathcal{C}_K : d(c_1, c_2) = \delta} \left|F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0 + c_1 n^{-1/2})) - F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0 + c_2 n^{-1/2}))\right| \\
\leq\ & C\delta^2.
\end{aligned}$$

Hence, (18) holds for the partition $\mathcal{B}_\delta$. By Theorem A1 the asymptotically tightness of $K_n(c)$ is then proved. $\square$

Proof of Theorem 1: A consequence of the asymptotically tightness of $K_n(c)$ is that if $\widehat{c}$ converges in distribution to $c_0$, then

$$|K_n(\widehat{c}) - K_n(c_0)| = o_P(1).$$

Apply this argument with $\widehat{c} = \sqrt{n}(\theta_n - \theta_0)'$ to prove that under Assumptions A1-A4,

$$\begin{aligned}
n^{-1/2} \sum_{t=1}^n (\Psi_{t,\alpha}(\theta_n) - \alpha) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\Psi_{t,\alpha}(\theta_0) - F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0))\right] \\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_n)) - F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0))\right] \\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) - \alpha\right] + o_P(1).
\end{aligned}$$

Now, by the Mean Value Theorem,

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_n))-F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right] \\
= \quad & \sqrt{n}(\theta_n-\theta_0)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_\alpha(I_{t-1},\theta_0) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))(\theta_n-\theta_0) \\
& +\sqrt{n}(\theta_n-\theta_0)\frac{1}{n}\sum_{t=1}^{n}\left[g_\alpha(I_{t-1},\widetilde{\theta}_n) f_{I_{t-1}}(m_\alpha(I_{t-1},\widetilde{\theta}_n))-g_\alpha(I_{t-1},\theta_0) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right] \\
= \quad & A_{1n}+A_{2n},
\end{aligned}$$

where $\widetilde{\theta}_n$ is between $\theta_n$ and $\theta_0$. The Ergodic Theorem and the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2) implies for any compact set $\Theta_c \subset \Theta_0$,

$$\sup_{\theta\in\Theta_c}\left|\frac{1}{n}\sum_{t=1}^{n} g_\alpha(I_{t-1},\theta) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta))-E[g_\alpha(I_{t-1},\theta) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta))]\right| = o_P(1).$$

Hence, the last convergence and A4 imply $A_{2n}=o_P(1)$, and hence

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_n))-F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right] \\
= \quad & \sqrt{n}(\theta_n-\theta_0)' E\left[g_\alpha(I_{t-1},\theta_0) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0))\right]+o_P(1).
\end{aligned}$$

This proves the theorem. □

Proof of Corollary 1: Define the process

$$S_n(c) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[(\Psi_{t,\alpha}(\theta_0+cn^{-1/2})-\alpha) l(Y_t,I_{t-1},\theta_0+cn^{-1/2})-G_{I_{t-1}}(\theta_0+cn^{-1/2})\right],$$

where

$$G_{I_{t-1}}(\theta)=E\left[(\Psi_{t,\alpha}(\theta)-\alpha) l(Y_t,I_{t-1},\theta)\mid \mathcal{F}_{t-1}\right].$$

The process $S_n(c)$ is 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. Similar arguments to those of the proof of our Lemma A1 show that $S_n(c)$ is asymptotically tight. Hence,

$$\rho_n=\frac{1}{n}\sum_{t=1}^{n}\left(\Psi_{t,\alpha}(\theta_n)-\alpha\right) l(Y_t,I_{t-1},\theta_n)=\frac{1}{n}\sum_{t=1}^{n} G_{I_{t-1}}(\theta_n)+o_P(1).$$

Now, the ULLN of Jennrich (1969, Theorem 2) implies that

$$\left| \frac{1}{n} \sum_{t=1}^{n} G_{I_{t-1}}(\theta_n) - E[(\Psi_{t,\alpha}(\theta_0) - \alpha) l(Y_t, I_{t-1}, \theta_0)] \right| = o_P(1),$$

and that

$$V_n = V + o_P(1).$$

On the other hand, Giacomini and Komunjer (2005) proved that $A_{n,\tau} = A + o_P(1)$. The proof follows then from Slutsky's Lemma. □

Proof of Theorem 2: Define the following quantities

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

and similarly, define $\xi_{n,j}$, $\xi_{1n,j}$ and $\xi_{2n,j}$ with $\theta_0$ replacing $\theta_n$. Now, simple algebra shows that

$$(n-j)^{1/2} \widetilde{\gamma}_{n,j} = \widehat{\xi}_{n,j} - \alpha \widehat{\xi}_{1n,j} - \alpha \widehat{\xi}_{2n,j}.$$

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

$$\begin{aligned}
(n-j)^{1/2} \left( \widetilde{\gamma}_{n,j} - \gamma_{n,j} \right) &= \left( \widehat{\xi}_{n,j} - \xi_{n,j} \right) - \alpha \left( \widehat{\xi}_{1n,j} - \xi_{1n,j} \right) - \alpha \left( \widehat{\xi}_{2n,j} - \xi_{2n,j} \right) \\
: &= A_{1n} - A_{2n} - A_{3n}.
\end{aligned}$$

By the Ergodic Theorem

$$(n-j)^{-1/2} \xi_{1n,j} = (n-j)^{-1/2} \xi_{2n,j} = o_P(1).$$

From Theorem 1, for $h = 1, 2$,

$$\widehat{\xi}_{hn,j} - \xi_{hn,j} = \sqrt{n}(\theta_n - \theta_0)' E \left[ g_\alpha(I_{t-1}, \theta_0) f_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) \right] + o_P(1).$$

And by similar arguments to those of Theorem 1,

$$\begin{aligned}
\widehat{\xi}_{n,j} - \xi_{n,j} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[ F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_n)) F_{I_{t-1-j}}(m_\alpha(I_{t-1-j},\theta_n)) - F_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0)) F_{I_{t-1-j}}(m_\alpha(I_{t-1-j},\theta_0 \right. \\
&= 2\alpha\sqrt{n}(\theta_n - \theta_0)' E\left[ g_\alpha(I_{t-1},\theta_0) f_{I_{t-1}}(m_\alpha(I_{t-1},\theta_0)) \right] + o_P(1).
\end{aligned}$$

From this, it follows that $A_{1n} - A_{2n} - A_{3n} = o_P(1)$.

Let $\eta'_{n,m} = (\gamma_{n,1}, ..., \gamma_{n,m})'$. Simple but tedious algebra shows that

$$Cov(\gamma_{n,i}, \gamma_{n,j}) = \alpha(1-\alpha)\delta_{ij} + o_P(1) \qquad i,j = 1, ..., m,$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. By a Martingale Central Limit Theorem

$$n^{1/2}\eta_{n,m} \to_d N(0, \alpha(1-\alpha)I_m),$$

where $I_m$ is the identity matrix of order $m$. Theorem 2 then follows from Slutsky's Lemma. $\square$

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

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.

Bollerslev, T. (1990): "Modelling the Coherence in Sort-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model."*Review of Economics and Statistics*, 72, 498-505.

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.* Forthcoming.

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

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

Danielsson, J., Morimoto, Y. (2003): “Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market. ”Monetary and Economic Studies, 18,2, 25-48.

Delgado, M.A. and Escanciano, J.C. (2006): “Nonparametric tests for conditional symmetry in dynamic models”, forthcoming in *Journal of Econometrics.*

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

Engle, R., Kroner, K. (1995): “Multivariate Simulataneous GARCH. ”*Econometric Theory*, 11, 122-150.

Engle, R., Manganelli, S. (2004): “CAViaR: Conditional autoregressive Value-at-Risk by Regression quantiles. ”*Journal of Business and Economic Statistics,* 22, 367-381.

Engle, R. (2002): “Dynamic Conditional Correlation- A Simple Class of Multivariate GARCH Models. ”*Journal of Business and Economic Statistics,* 20, 339-350.

Escanciano, J. C. (2006): “Weak Convergence of Function-Parametric Marked Empirical Processes with Applications to Efficient Model Checks. ”. Manuscript.

Escanciano, J. C. and Velasco, C., (2006): “ Generalized spectral tests for the martingale difference hypothesis” *Journal of Econometrics,* 134, 151-185.

Fan, J. and Gu, J. (2003): “Semiparametric estimation of Value at Risk.” *Econometrics Journal*, 6, 261-290.

Figlewski, S. (2003): “Estimation Error in the Assessment of Financial Risk Exposure. ”Available at SSRN: http://ssrn.com/abstract=424363

Giacomini and Komunjer (2005): “Evaluation and Combination of Conditional Quantile Forecasts. ”*Journal of Business & Economic Statistics*, 23, 4, 416-431(16).

Goureioux, C., Jasiak, J. (2006): “Dynamic Quantile Models. ”Working Paper 2006-4, Department of Economics, York University.

Haas, M. (2005): “Improved Duration-based Backtesting of Value-at-risk,” *Journal of Risk*, 8, 17–38.

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

HAMILTON, J. (1994): *Time Series Analysis.* Princeton University Press, Princeton, New Jersey.

HANSEN, B.E. (2006): "Interval Forecast and Parameter Uncertainty. "*Journal of Econometrics*, 135, 377-398.

KERKHOF, J., MELENBERG, B. (2002): "Backtesting for Risk-Based Regulatory Capital. "*Discussion Paper* 110, Tilburg University, Center for Economic Research.

KITAMURA, Y., STUTZER, M. (1997): "Information theoretic alternative to generalized method of moments estimation."*Econometrica*, 65, 861-874.

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. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* 2nd ed, Lecture Notes in Statistics, Vol. 166, Springer.

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.

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.

LJUNG, G. M. AND BOX, G. E. P. (1978): "A measure of lack of fit in time series models." *Biometrika,* 65, 297-303.

Shorack, G.R., Wellner, J.A. (1986): *Empirical Processes with Applications to Statistics.* New York: Wiley.

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

Wu,W.B. (2005): "On the Bahadur representation of sample quantiles for dependent sequences." *Annals of Statistics*, 33, 1934–1963