SPECIFICATION TESTS IN PARAMETRIC VALUE-AT-RISK MODELS

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

Jose Olmo
City University, London, UK

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 and of out-of-sample backtesting for banking risk monitoring. We stress in this article that the results derived from this exercise can be spurious if one does not carry out a previous in-sample specification test to determine the adequacy of the VaR model. We study in this paper specification tests that, unlike the existing ones, are able to control the type-I error probability. More concretely, we show that not taking into account the effect of estimating the parameters of the VaR model in the in-sample specification tests can lead to invalid inferences, which in turn may imply wrong conclusions about the out-of-sample backtesting procedures. The first aim of this article is to quantify the effect of estimating the parameters of the model and to stress its impact in specification tests, and the second is then to propose a corrected method taking into account such risk, and thereby to provide a valid econometric framework for measuring and evaluating market risk. The results are given for general dynamic parametric models and illustrated with a Monte-Carlo simulation for location-scale models and with an empirical application for $S\&P500$ Index.

Keywords and Phrases: Backtesting; Basel Accord; Model Risk; Risk management; Specification tests; Value at Risk.

1 Introduction

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 and of out-of-sample backtesting for banking risk monitoring. As a result of this agreement financial institutions have to report their VaR, defined as a conditional quantile with coverage probability α of the distribution of returns on their trading portfolio. To test the performance of this and alternative VaR measures the Basel Accord (1996) set a statistical device denoted backtesting that consisted of out-of-sample comparisons between the actual trading results with internally model-generated risk measures. The magnitude and sign of the difference between the model-generated measure and actual returns indicate whether the VaR model reported by an institution is correct for forecasting the underlying market risk and if this is not so, whether the departures are due to over- or under-risk exposure of the institution. The implications of over- or under- risk exposure being diametrically different: either extra penalties on the level of capital requirements or bad management of the outstanding equity by the institution. These backtesting techniques are usually interpreted as statistical parametric tests for the coverage probability α defining the conditional quantile VaR measure.

This out-of-sample problem has been thoroughly studied by many authors; see Kupiec (1995), Christoffersen (1998), Berkowitz and O'Brien (2002), Engle and Manganelli (2004) and Escanciano and Olmo (2008), to name but a few. These references have assumed the correct specification of the VaR model in forecast evaluations. Therefore, prior to the forecasting stage, the risk manager has to decide, using the available information, i.e. all the sample, which econometric model is most adequate for the conditional VaR process. This preliminary stage involves model selection and validation, hence the importance of quantile specification tests associated to VaR. This step, and in particular the study of estimation risk for these specification tests has been overlooked in the risk management literature.

We stress in this article that in-sample standard specification procedures for VaR quantile processes can lead to invalid inferences, which in turn may imply wrong conclusions about the correct specification of the conditional VaR model, and about the subsequent out-of-sample backtesting procedures. We shall show that the introduction of uncertainty about the true parameter coming from the data, adds an additional term (*estimation risk*) in the standard specification procedures that must be taken into account to construct valid inferences on VaR in-sample diagnostics. Otherwise these techniques used for model checking can be completely misleading due to the choice of wrong critical values. Therefore, the first aim of this article is to quantify the effect of estimating the parameters of the model and to stress its impact in specification tests. The second aim of the paper is then to propose a corrected method taking into account such risk, and thereby to provide a valid statistical framework for measuring and evaluating market risk. The results are given for general dynamic parametric models (Section 2) although to illustrate the new statistical methodology we derive the analytic expressions of the estimation error for location-scale models and perform some experiments via Monte-Carlo simulation (Section 3). Section 4 presents an empirical application and Section 5 concludes.

2 Specification testing techniques for Quantile Models

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. Assuming that the conditional distribution of Y_t given W_{t-1} is continuous, we define the α-th conditional VaR (i.e. quantile) of Y_t given W_{t-1} as the 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 $\mathcal{M}$. From (1), the parametric VaR model $m_\alpha(W_{t-1}, \theta_0)$ is well specified if and only if

$$H_0 : E[I_{t,\alpha}(\theta_0) \mid W_{t-1}] = \alpha \text{ a.s. for some } \theta_0 \in \Theta, \tag{2}$$

where $I_{t,\alpha}(\theta_0) := 1(Y_t \leq m_\alpha(W_{t-1}, \theta_0))$. In particular, under (2), the sequence $\{I_{t,\alpha}(\theta_0)\}$ are *iid*, and moreover, the following unconditional hypothesis holds

$$H_{0u} : E[I_{t,\alpha}(\theta_0)] = \alpha. \tag{3}$$

Kupiec (1995), first proposed tests for H_{0u} based on the absolute value of the standardized sample mean $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (I_{t,\alpha}(\theta_0) - \alpha)$, which converges to zero-mean normal variable with variance $\alpha(1-\alpha)$ under (2). This test is optimal if θ_0 is known; see Chirstoffersen (1998). More modern tests in a related risk measurement testing framework are, for example, Giacomini and White (2006) for out-of-sample predictive ability, and Cotter and Dowd (2007) for exponential spectral risk measures.

An important limitation of the aforementioned techniques for model specification is the assumption of the parameter θ_0 being known. In practice however, the parameter θ_0 is unknown and must be estimated from a sample $\{Y_t, W'_{t-1}\}_{t=1}^{n}$ by a $\sqrt{n}-$consistent estimator, say θ_n. We show in the next section that the estimation of parameters in the VaR model has a nonnegligible effect in the asymptotic distribution of the specification test and leads to a different source of risk, called *estimation risk.* For the save of space and a better comparison with the existing literature we focus on tests for (2) in the direction of H_{0u}, but our theory can be applied to more general situations.

3 Specification tests free of estimation risk

Following Kupiec (1995), we consider tests based on

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

with θ_n a $\sqrt{n}-$consistent estimator replacing the parameter θ_0, that satisfies the following assumptions.

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

Assumption A2: The family of distributions functions $F_{W_{t-1}}(y) := P(Y_t \leq y \mid W_{t-1} = x)$, has

Lebesgue densities $f_{W_{t-1}}(y)$ 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 θ_n satisfies the asymptotic Bahadur expansion

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

where $l(\cdot)$ is such 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.

Under these regularity conditions we are in position to establish the first important result of the paper.

Theorem 1: *Under Assumptions A1 to A4,*

$$S_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}[I_{t,\alpha}(\theta_n)-\alpha]=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[I_{t,\alpha}(\theta_0)-F_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\right] \tag{4}$$

$$+\underbrace{\sqrt{n}(\theta_n-\theta_0)'E\left[g_\alpha(W_{t-1},\theta_0)f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\right]}_{\text{Estimation Risk}} \tag{5}$$

$$+\underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[F_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))-\alpha\right]}_{\text{Model Risk}}+o_P(1). \tag{6}$$

The proof of this result in an out-of-sample context can be found in Escanciano and Olmo (2008). Theorem 1 quantifies both *estimation risk* and *model risk* in the unconditional specification test introduced before. A corollary from our Theorem 1 is that as long as $E\left[F_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\right]\neq\alpha$, tests based on S_n will be consistent. In an out-of-sample framework Giacomini and White (2006) also account for the presence of model risk and propose tests of conditional predictive ability that are robust to possible misspecifications of the model, in our case misspecifications of VaR, and that provide more reliable forecasts due to the choice of suitable estimation methods. In what follows, however, we will assume no model risk. Next corollary provides the necessary corrections to carry out valid asymptotic inference for the specification test free of estimation risk.

Corollary 1: *Under Assumptions A1 to A4, and (2)*

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

where

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

with $\rho = E[(I_{t,\alpha}(\theta_0) - \alpha)\, l(Y_t, W_{t-1}, \theta_0)]$, and $A := E\left[g_\alpha(W_{t-1}, \theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))\right]$.

We further assume that the estimator θ_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 $\widehat{A}_\tau = -\frac{1}{n}\sum_{t=1}^{n} \frac{1}{\tau} \exp\left[(Y_t - m_\alpha(W_{t-1}, \theta_n))/\tau\right] I_{t,\alpha}(\theta_n) g_\alpha(W_{t-1}, \theta_n)$, with $\tau \to 0$ as $n \to \infty$; see Giacomini and Komunjer (2005). Hence, provided that $\tau \to 0$ the asymptotic variance can be consistently estimated by $\widehat{\sigma}_c^2 := \alpha(1-\alpha) + 2\widehat{A}_\tau \widehat{\rho} + \widehat{A}_\tau \widehat{V} \widehat{A}_\tau'$ *with* $\widehat{\rho} = \frac{1}{n}\sum_{t=1}^{n} (I_{t,\alpha}(\theta_n) - \alpha)\, l(Y_t, W_{t-1}, \theta_n)$ *and* $\widehat{V} = \frac{1}{n}\sum_{t=1}^{n} l(Y_t, W_{t-1}, \theta_n) l'(Y_t, W_{t-1}, \theta_n)$.

Corollary 2: *Under Assumptions A1 to A4, and (2)*

$$\widetilde{S}_n \xrightarrow{d} N(0,1),$$

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

3.1 Examples:

3.1.1 Historical Simulation

The historical simulation VaR is simply the unconditional quantile of Y_t. Hence the postulated model is $m_\alpha(W_{t-1}, \theta_0) = \theta_0 \equiv F_Y^{-1}(\alpha)$, where $F_Y^{-1}(\alpha)$ denotes the unconditional quantile function of Y_t evaluated at α. Let $F_Y(x)$ be the *cdf* of Y_t. The estimator of θ_0 is usually $\theta_n = F_{t,n}^{-1}(\alpha)$, where $F_{n,Y}^{-1}(\alpha)$ is the empirical quantile function of $\{Y_t\}_{t=1}^n$. Under some mild assumptions,

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

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

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

It is worth mentioning that the unconditional backtesting procedure based on historical simulation VaR will be inconsistent. This is so because it is always true that $E\left[F_{W_{t-1}}(\theta_0)\right] = \alpha$, since $\alpha = F_Y(F_Y^{-1}(\alpha))$ regardless if 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.

3.1.2 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

$$Y_t = \mu(W_{t-1}, \beta_0) + \sigma(W_{t-1}, \beta_0)\varepsilon_t, \tag{7}$$

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

$$m_\alpha(W_{t-1}, \theta_0) = \mu(W_{t-1}, \beta_0) + \sigma(W_{t-1}, \beta_0)F_\varepsilon^{-1}(\alpha), \tag{8}$$

where $F_\varepsilon^{-1}(\alpha)$ denotes a univariate quantile function of ε_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 β_0 is usually estimated by the Quasi-Maximum Likelihood Estimator (*QMLE*). 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. 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

$$\sqrt{n}(F_{\varepsilon,n}^{-1}(\alpha) - F_\varepsilon^{-1}(\alpha))f_\varepsilon(F_\varepsilon^{-1}(\alpha)) + \sqrt{n}(\beta_n - \beta_0)'b(\alpha, \beta_0), \tag{9}$$

where

$$b(\alpha, \beta_0) := f_\varepsilon(F_\varepsilon^{-1}(\alpha))E\left[a_{1,t}(\beta_0)\right] + f_\varepsilon(F_\varepsilon^{-1}(\alpha))F_\varepsilon^{-1}(\alpha)E\left[a_{2,t}(\beta_0)\right], \tag{10}$$

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

with $\dot{\mu}_t(\beta) = \partial\mu(W_{t-1}, \beta)/\partial\beta$ 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 resultant from estimating β_0. Thus, for example, for a homoscedastic model with zero conditional mean process the only estimation effect affecting the process comes from estimating the quantile of the error distribution, the other estimation effects vanish because $a_{1,t}(\beta) = 0$ and $a_{2,t}(\beta) = 0$. On the other hand if the process is an ARMA-GARCH process with error distribution following a Student-t with an unknown discrete number of degrees of freedom the estimation error of the VaR model stems from the uncertainty of estimating the location-scale model since the estimators of the degrees of freedom converge at a faster rate (n) to the parameter than the estimators of the location-scale model ($\sqrt{n}$), see Hannan and Quinn, 1979, p. 191, for general results on estimation of discrete-valued parameters. We will assume in the following simulation section models where the error distribution is either known (Gaussian) or unknown up to a finite number of parameters (Student-t). In both cases then the estimation error is given by expression (10).

4 Simulation Exercise

This section examines the performance through some Monte Carlo experiments of the test devised in Kupiec (1995) and the corrected test developed in this paper. We consider a realistic model for financial data: an ARMA(1,1)-GARCH(1,1) process represented by

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

with the true parameters given by $\beta_0 = (a, b, \eta_{00}, \eta_{10}, \eta_{20})' = (0.1, 0.1, 0.05, 0.1, 0.85)$. We assume two two different innovation processes $\{\varepsilon_t\}$ defined by $\varepsilon_t = (\sqrt{(\nu-2)/\nu})v_t$, with v_t distributed as a Student-t with $\nu = 30$ and $\nu = 10$ degrees of freedom. A t_{30} distribution is considered as an approximation to the Gaussian distribution and a t_{10} to illustrate the impact of heavier than normal tails in the different specification tests. The Value at Risk of these models is calculated at $\alpha = 1\%$ as recommended by Basel Committee.

This process is intended to describe actual dynamics followed by financial returns. Table 4.1 reports the simulated sizes corresponding to this ARMA(1,1)-GARCH(1,1) model with known Student-t error term.

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 4.1. *Size of specification tests S_n and $\widetilde{S}_n$ for $\alpha = 0.01$, and ε_t following a family of t_ν with $\nu = 30, 10$ 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 small simulation experiment are illuminating. $\widetilde{S}_n$ outperforms S_n in terms of size, and the simulated size reported for $\widetilde{S}_n$ is close to the actual nominal values. On the other hand the test statistic S_n is very unreliable in both cases. Also, the size distortions in S_n do not vanish as the sample size increases, confirming our findings in Theorem 1 claiming that estimation risk appears even asymptotically (*i.e.* with infinite sample sizes.) Other conclusions from this table 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).

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. These findings, though interesting, may be spurious if the specification test statistic employed to validate the model in-sample does not take into account estimation risk effects. In fact, we have uncovered in this paper that these standard testing procedures produce wrong critical values to assess in-sample VaR estimates from parametric models. In particular we show in this application that Kupiec's test statistic understates risk exposure when implemented for estimated parameters. This effect is gauged in this application for daily log-returns on the $S\&P500$ market-valued equity Index over the period 02/2000 - 11/2006 ($n = 1706$ observations), and obtained from *Freelunch.com*.

In order to detect if the VaR measure understates or overstates risk exposure we report the test statistics S_n and $\widetilde{S}_n$ derived from 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 estimating the degrees of freedom of a Student-t distribution for the residuals of the estimated ARMA(1,1)-GARCH(1,1) model. The experiment is designed as follows.

We consider a five-day rolling window of 250 observations each covering the period of data available (n=1706 observations). For each period of data in the window (250 observations) we estimate an ARMA(1,1)-GARCH(1,1) model and compute the corresponding time series of VaR estimates. The in-sample validation experiment consists on comparing the actual observations from the window with the corresponding VaR estimates, if the number of exceedances is significantly far from its expected value under "normal "circumstances the risk model specification is rejected for the relevant sample. According to Basel Committee there are three different tolerance levels: green, yellow and red. These correspond to values of the specification test statistics below the Gaussian 95% confidence level, to values between 95% and 99.99% and to values exceeding 99.99% confidence level respectively. In terms of capital requirements (CR_t) required by Basel Accord (1996) these regions imply different multiplication factors mf_t in the formula $CR_t = mf_t \times VaR_{0.01,t}$, being the lower bound $mf_t = 3$ for a VaR measure reporting a green zone and the upper bound $mf_t = 4$ for a red zone VaR.

The following plot reports the values of S_n and $\widetilde{S}_n$ corresponding to each rolling window. 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.

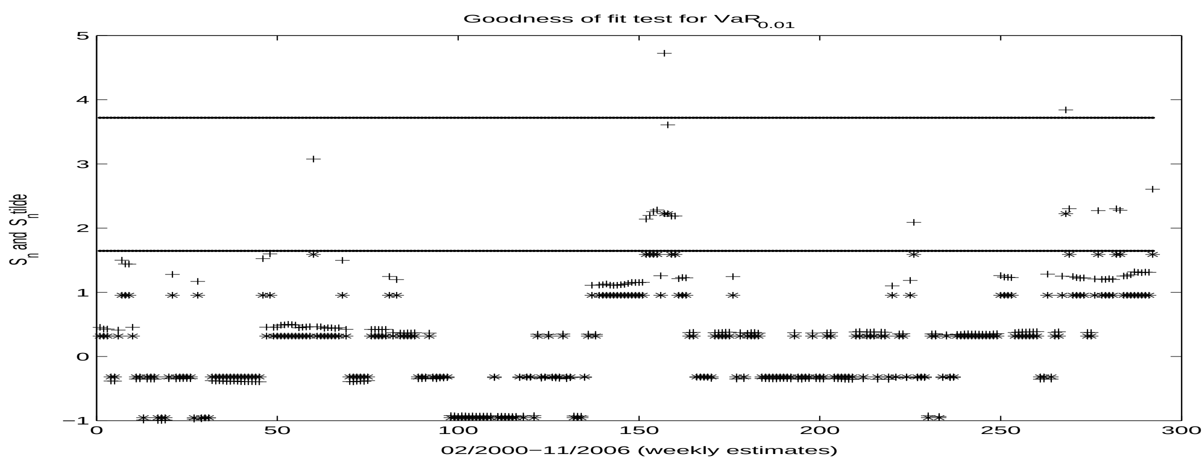

Figure 5.1. *S_n and $\widetilde{S}_n$ test statistics for $VaR_{0.01}$ estimates over 5-day rolling windows of 250 daily observations from an ARMA(1,1)-GARCH(1,1) model with Gaussian error ($N(0,1)$) that is re-estimated every five days. The data are returns on S&P500 Index over the period 02/2000-11/2006. 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 test reports massive warnings of model failure during the same time intervals. More importantly, the red zone (upper threshold level) defined by Basel Accord as 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 rules of Basel Accord (1996) would be indicating. These findings point towards either rejection of the dynamic parametric model or rejection of the Gaussian distribution. To see if the latter is the reason to reject the risk model we plot in the next figure the values of S_n and $\widetilde{S}_n$ when the degrees of freedom of the Student-t distribution are estimated.

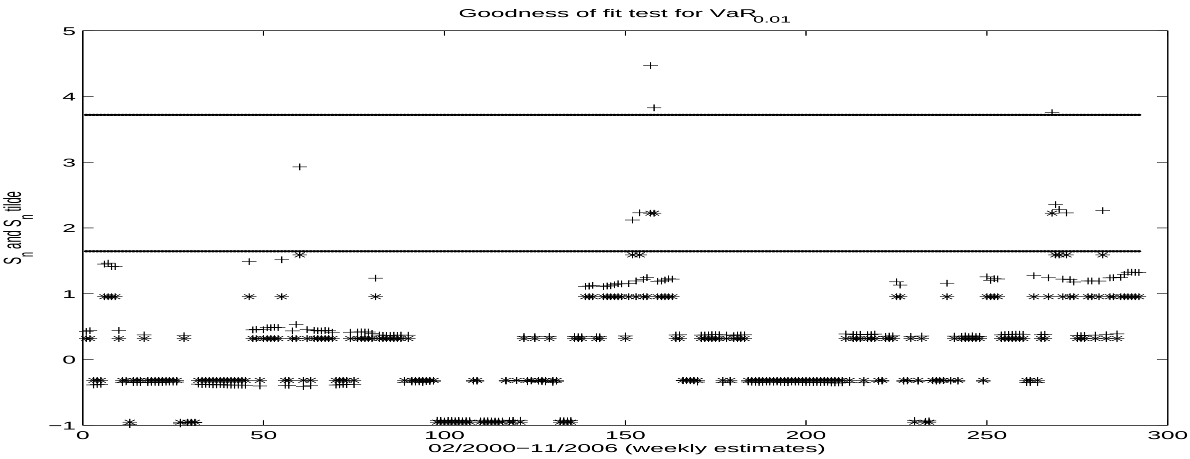

Figure 5.2. *S_n and $\widetilde{S}_n$ test statistics for $VaR_{0.01}$ estimates over 5-day rolling windows of 250 daily observations from an ARMA(1,1)-GARCH(1,1) model with Student-t_ν distribution error. The parameters and degrees of freedom of the distribution are re-estimated every five days. The 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.*

It is clear from figure 5.3 that the error distribution is not heavy-tailed, the estimates of the degrees of freedom throughout the sample indicate that the Gaussian distribution is not a bad approximation, hence the similarities between figures 5.1 and 5.2. Given that $\widetilde{S}_n$ corrects for the presence of estimation risk there are statistical grounds to conclude that the ARMA(1,1)-GARCH(1,1) family of models is not appropriate for certain periods of the analysis. It is worth noting that these findings are hindered by the presence of estimation risk when using standard specification tests.

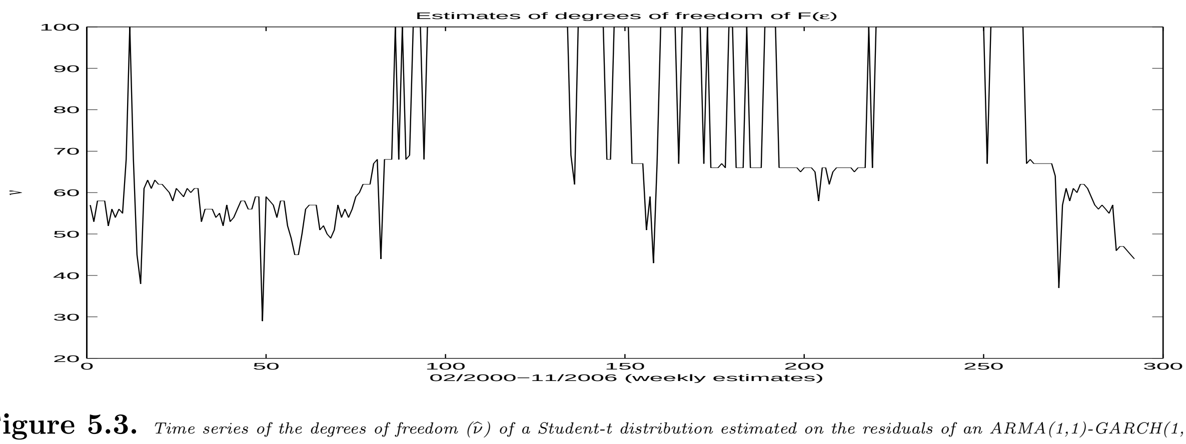

Figure 5.3. *Time series of the degrees of freedom ($\widehat{\nu}$) of a Student-t distribution estimated on the residuals of an ARMA(1,1)-GARCH(1,1) model with parameters estimated over 5-day rolling windows of 250 daily observations.*

6 Conclusion

Basel and Basel II Accords propose the use of specification tests and backtesting techniques to assess the accuracy and reliability of banks 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 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 testing procedures used by banks and regulators to assess dynamic parametric VaR estimates can be misleading when used for in-sample model validation. This implies that any conclusion regarding the validity of these risk models 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.

In an application to financial returns on $S\&P500$ Index we show that this correction is able to uncover the wrong specification of location-scale models of ARMA(1,1)-GARCH(1,1) type for some periods. These findings can lead to very heavy fines in the form of higher capital requirements, due to risk missmeasurement. However if the standard specification tests based on Kupiec and Christoffersen techniques are used this evidence vanishes due to the presence of estimation risk effects. The phenomenon of estimation risk effects on quantile models for risk management is of very practical importance and we believe it deserves further investigation. Its effects on an out-of-sample backtesting environment are studied in Escanciano and Olmo (2008).

REFERENCES

Basel Committee on Banking Supervision (1996): "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.

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

Cotter, J., Dowd, K. (2007): "Exponential Spectral Risk Measures. " *The Icfai Journal of Financial Economics,* 4, 57-66.

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

Escanciano, J. and Olmo, J., (2008): "Backtesting Value-at-Risk with Estimation Risk." *Forthcoming in Journal of Business and Economic Statistics.*

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

Giacomini, R. and White, H. (2006): "Tests of Conditional Predictive Ability." *Econometrica*, 74 (6), 1545-1578.

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.

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