Wharton

Financial Institutions Center

Testing, Comparing, and Combining Value at Risk Measures

by
Peter Christoffersen
Jinyong Hahn
Atsushi Inoue

99-44

The Wharton School
University of Pennsylvania

THE WHARTON FINANCIAL INSTITUTIONS CENTER

The Wharton Financial Institutions Center provides a multi-disciplinary research approach to the problems and opportunities facing the financial services industry in its search for competitive excellence. The Center's research focuses on the issues related to managing risk at the firm level as well as ways to improve productivity and performance.

The Center fosters the development of a community of faculty, visiting scholars and Ph.D. candidates whose research interests complement and support the mission of the Center. The Center works closely with industry executives and practitioners to ensure that its research is informed by the operating realities and competitive demands facing industry participants as they pursue competitive excellence.

A variety of working papers related to this subject are also available online at:

http://fic.wharton.upenn.edu/fic

If you would like to learn more about the Center or become a member of our research community, please let us know of your interest.

Anthony M. Santomero
Director

The Working Paper Series is made possible by a generous grant from the Alfred P. Sloan Foundation

Testing, Comparing, and Combining Value-at-Risk Measures[1]

Peter Christoffersen
McGill University[2]

Jinyong Hahn
University of Michigan[3]

Atsushi Inoue
North Carolina State University[4]

October 1, 1999

[1] We would like to thank Mike Chernov and Eric Ghysles for providing us with their data, and Yuichi Kitamura for sharing his Gauss code. We would also like to thank participants at the 1998 Triangle Econometics Conference and the 1999 ESEM Conference for helpful comments. The usual disclaimer applies.

[2] Corresponding author. Department of Finance, 1001 Sherbrooke Street W, Montreal, PQ H3A 1G5. Phone: 514-398-2869, fax: 514-398-3876, email: christop@management.mcgill.ca.

[3] Department of Economics. Phone: 734-764-2447, fax: 734-764-2769, email: jinhahn@econ.lsa.umich.edu.

[4] Department of Agricultural and Resource Economics. Phone: (919) 515-5969, fax: 919-515-1824, email: Atsushi_Inoue@ncsu.edu.

Abstract

Value-at-Risk (VaR) has emerged as the standard tool for measuring and reporting financial market risk. Currently, more than eighty commercial vendors offer enterprise or trading risk management systems which report VaR-like measures. Risk managers are therefore often left with the daunting task of having to choose from this plethora of risk measures. Accordingly, this paper develops a framework for answering the following questions about VaRs: 1) How can a risk manager test that the VaR measure at hand is properly specified, given the history of asset returns? 2) Given two different VaR measures, how can the risk manager compare the two and pick the best in a statistically meaningful way? Finally, 3) How can the risk manager combine two or more different VaR measures in order to obtain a single statistically superior measure? The usefulness of the methodology is illustrated in an application to daily returns on the S&P500. In the application, competing VaR measures are calculated from either historical or option-price based volatility measures, and the VaRs are then tested and compared.

JEL Codes: G10, C22, C53

Keywords: Risk management, volatility, nonnested testing, options

1 Motivation

Sparked by the stock market crash in 1987, the past decade has witnessed a surge in the effort financial market participants devote to risk management. In a recent survey of risk management software, Risk (1999) lists more than *eighty* commercial vendors offering enterprise or trading risk management information systems. This effort has been encouraged by regulatory authorities imposing risk-based capital adequacy requirements on financial institutions (see Dimson and Marsh, 1995, and Wagster, 1996). The recent turbulence in emerging markets, starting in Mexico in 1995, continuing in Asia in 1997, and spreading to Russia and Latin America in 1998 has further extended the interest in risk management to companies outside the traditional sphere of banking and insurance.

Two important developments, one in academia and one on Wall Street have facilitated the advancement in knowledge about risk management. First, the development of volatility models for measuring and forecasting volatility dynamics began in academics with Engle (1982). The hundreds of papers following Engle's original work – many of them finding applications to financial data – have had important implications for modern risk management techniques. Second, the introduction of RiskMetrics by JP Morgan (1996) has enabled companies with just a minimum of computational power and technical ability to compute simple measures of market risk for a given portfolio of assets. RiskMetrics has also aroused the interest of academics as it offers a benchmark methodology upon which improvements can be made, and against which alternatives can be tested. Research in this tradition is reported in Jorion (1996), Duffie (1997), and Dowd (1998).

An important contribution of the RiskMetrics methodology is the introduction of the Value-at-Risk (VaR) concept which collapses the entire distribution of the portfolio returns into a single number which investors have found useful and easily interpreted as a measure of market risk. The VaR is essentially a p-percent quantile of the conditional distribution of portfolio returns.

In RiskMetrics, the VaR measure has only a few unknown parameters which are simply calibrated to values found to work quite well in common situations. However, several studies such as Danielsson and de Vries (1997), Christoffersen (1998), and Engle and Manganelli (1999) have found significant improvements possible when deviations from the relatively rigid RiskMetrics framework are explored. But, when one attempts to apply the results which have emerged from the GARCH and related literatures to risk management, several questions remain open. We ask the following questions: 1) Given a VaR measure, how can the risk manager test that the particular measure at hand is appropriately specified? 2) Given two different VaR measures, say one using GARCH and one using implied volatility, how can the risk manager compare the two and pick the best in a statistically meaningful way? Finally, 3) How can the risk manager explore the possibility of combining two or more different VaR measures in order to obtain a single statistically optimal measure? Choosing an optimal VaR measure is important. Beder (1995) finds for example that applying different VaR models to the same portfolio can yield significantly different assessments of risk.

We illustrate the usefulness of our approach in an application to daily returns on the S&P500

index. We test and compare VaR measures based on GARCH-type volatilities estimated from historical returns with measures based on implied and estimated volatilities from options contracts written on the S&P500 index. We use the volatility measures constructed by Chernov and Ghysels (1998).

The development of a specification testing methodology is complicated by the fact that the VaR concept introduces an important nondifferentiability which invalidates existing statistical testing procedures. In addition, when comparing two competing measures, it is essential to allow for them to be nonnested. We tackle these challenges by extending the recent results by Kitamura (1997) to allow for nondifferentiability.

The remainder of our paper is structured as follows: In Section 2, we establish some notation and develop a moment-based framework for VaR specification testing, nonnested VaR comparison testing, and VaR model combination. In Section 3, we introduce the econometric methodology and show how it can be applied to testing VaR models. In Section 4, we apply our methodology to returns on the S&P500 index, comparing traditional time series based VaR measures to VaRs based on implied volatilities from options prices. Section 5 concludes and gives directions for future research.

2 Value-at-Risk with Conditional Moment Restrictions

We set out by defining the notation necessary for establishing our testing framework.

2.1 Defining Value-at-Risk

Let the asset return process under consideration be denoted by

$$y_t = \mu_t + \varepsilon_t,$$

where $\varepsilon_t | \Psi_{t-1} \sim \left(0, \sigma_t^2\right)$, and where Ψ_{t-1} is the time-$t-1$ information set. Then the Value-at-Risk measure with coverage probability, p, is defined as the conditional quantile, $F_{t|t-1}(p)$, where

$$\Pr\left(y_t \leq F_{t|t-1}(p) \middle| \Psi_{t-1}\right) = p.$$

The conditionality of the VaR measure is key. Throughout this paper, we will assume that y_t is appropriately demeaned so that $\mu_t = 0$ and $y_t = \varepsilon_t$. But volatility will be allowed to be time-varying.

2.2 Specifying Volatility

Risk managers have a plethora of volatility measures to choose from when calculating Value-at-Risk (VaR) measures. Time series models of volatility range from exponentially smoothed and simple autoregressive models, over single-shock GARCH models, to two-shock stochastic volatility models. Furthermore, the risk manager can use option based measures of volatility to measure risk. Let us therefore first give a brief overview of available volatility models.

The benchmark measure advocated in JP Morgan's (1996) RiskMetrics sets the conditional mean constant, and specifies the variance as an exponential filter

$$\sigma_t^2 = (1-\lambda)\,\varepsilon_{t-1}^2 + \lambda \sigma_{t-1}^2, \tag{1}$$

where λ is simply set to .94 for daily data. The innovations are assumed to be Gaussian, thus the VaR measure is

$$F_{t|t-1}^{RM}(p) = \mu + \Phi^{-1}(p)\,\sigma_t.$$

Obviously, for $p = .05$, we would have $\Phi^{-1}(p) = -1.64$. In the standard Gaussian GARCH(1,1) case (Bollerslev 1986) the conditional variance evolves as

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{2}$$

and the one-step ahead conditional quantile with coverage p is

$$F_{t|t-1}^{G}(p) = \mu_{t|t-1} + \Phi^{-1}(p)\,\sigma_t.$$

Stochastic Volatility models instead assume volatility is driven by an unobserved factor. In the simplest case,

$$y_t - \mu_{t|t-1} = \varepsilon_t \exp\left(\frac{\sigma_t}{2}\right),$$

where

$$\sigma_t = \gamma_0 + \gamma_1\sigma_{t-1} + \eta_{t-1}.$$

Within each type of volatility model, many variants exist, based on considerations regarding long versus short memory, nonlinear versus linear specifications, and exogenous variables such as seasonal and trading volume effects.

GARCH, RiskMetrics and stochastic volatility models are all based solely on the history of the return y_t itself. But information on volatility may also be obtained from current market data such as option prices. In an effort to incorporate the market's belief about future returns, the risk manager can apply implied volatilities from options prices. Given data options contracts traded, the Black and Scholes (1972) implied volatility of a European call option can be found as the σ which solves

$$C = S\cdot\Phi(d_1) - \exp(-r(T-t))\,K\cdot\Phi(d_2), \tag{3}$$

where C is the quoted options price, $\Phi(\cdot)$ is the standard normal c.d.f., and

$$d_1 = \frac{\log\frac{S}{K} + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2 = d_1 - \sigma\sqrt{T-t}.$$

where $K, S, r, T-t$, and σ denote the strike price, the underlying asset price, the risk-free interest rate, the time-to-maturity, and the implied volatility respectively. Jorion (1995), for example, has found implied volatilities to work well as predictors of future volatility when using a standard mean squared error criterion.

One can also use option prices and asset returns to estimate a more realistic model of returns allowing for time-varying volatility. A benchmark model in this tradition is found in Heston (1993), who assumes that the price of the underlying asset, $S(t)$, evolves according to

$$dS(t) = \mu S dt + \sqrt{v(t)} S dz_1(t),$$

and volatility, $v(t)$, evolves according to

$$dv(t) = \kappa[\theta - v(t)]dt + \sigma\sqrt{v(t)}dz_2(t), \tag{4}$$

where the two Wiener process innovations, $dz_1(t)$ and $dz_2(t)$ are allowed to be correlated. Heston (1993) derives a closed-form solution for a European call option price which is similar in structure to equation (3). Chernov and Ghysels (1998) show how the parameters can be estimated using data on options and returns.

Other measures of volatility, which differ in the return data applied, include Garman and Klass (1980), and Gallant and Tauchen (1998) who incorporate daily high and low quotes, and Andersen and Bollerslev (1998) and Andersen, Bollerslev, Diebold and Labys (1999), who average intraday squared returns to estimate daily volatility.

In the empirical application at the end of the paper we will study VaR measures based on volatility measures from equations (1), (2), (3), and (4) respectively.

2.3 Conditional Moment Restrictions

Implicit in the context of risk management and the related pursuit of a good measure of volatility is an assumption that the return standardized by its conditional mean and some transformation of volatility, say $\xi(\sigma_t)$, is *i.i.d.*: If $(y_t - \mu_t)/\xi(\sigma_t)$ is not *i.i.d.* for any transformation $\xi(\cdot)$ of volatility, then volatility alone would not be sufficient for characterization of conditional quantile. Typically, we make an implicit assumption that y_t belongs to a *location-scale family*: We assume that $(y_t - \mu_t)/\sigma_t$ is *i.i.d.*, which would imply that the conditional quantile is some linear function of volatility, where the relevant coefficients of such a linear function is determined by the common distribution of the standardized return. Therefore, one can think of the VaR measure as the outcome of a quantile regression. Treating volatility as a regressor, and ignoring conditional mean dynamics, we have for example, that

$$F_{t|t-1}(\beta_p) = \beta_{p,1} + \beta_{p,2}\sigma_t$$

for *some* $\beta_{p,1}$ and $\beta_{p,2}$. Notice that the parameters will vary with the chosen coverage, p. A different VaR measure, based on a different volatility model σ_t^*, or a different distributional assumption, or both, could be written as

$$F^*_{t|t-1}(\theta_p) = \theta_{p,1} + \theta_{p,2}\sigma^*_t.$$

At least three questions now arise: First, "How do we evaluate the appropriateness of the specification of these measures?" Second, "How do we compare them?" And third, "Can we combine them to make an even better measure?"

In order to answer these questions, we apply the following conditional moment framework: Consider first the specification testing question. Given the risk manager's information set, Ψ_{t-1}, and under the null that the VaR measure is correctly specified, the following must hold:

Definition 1 *The VaR is efficient with respect to the information set,* Ψ_{t-1}, *when*

$$E\left[I\left(y_t < F_{t|t-1}\left(\beta_p\right)\right) - p \middle| \Psi_{t-1}\right] = 0,$$

where $I(\cdot)$ *is the indicator function.*

This moment condition states that no information available to the risk manager at time $t-1$ should help predict whether time t's return falls above or below the VaR measure reported at time $t-1$. The VaR measure should in other words be efficient with respect to the information set Ψ_{t-1}. We will refer to this as the *efficient VaR condition*. The first question can now be restated as, "Does a particular VaR measure satisfy the efficient VaR condition?"

It seems plausible that most VaRs are potentially misspecified. After all, it is hard to imagine that any econometric model underlying a VaR is an exact description of the data generating process. This would for instance be the case if the true distribution did not belong to a location-scale family. Under these circumstances, the conditional quantile of interest may not be a function of the conditional variance only, and conditional kurtosis, for example, may play an additional role in characterizing the conditional quantile. It is then likely that every VaR measure would be rejected given a sufficiently large amount of observations. We therefore want our testing methodology to allow for the possibility of misspecification.[1]

Our second research question may be restated as, "How do we compare misspecified VaR measures?" In order to answer this question, consider now again the competing VaR measure, $F^*_{t|t-1}(\theta_p)$. We can write

$$E\left[\left(I\left(y_t < F^*_{t|t-1}(\theta_p)\right) - p\right) |\Psi_{t-1}\right] = 0.$$

We then want to test whether $F_{t|t-1}\left(\beta_p\right)$ is significantly better than $F^*_{t|t-1}(\theta_p)$ in a statistically meaningful way, using these moment conditions.

Even if one VaR measure dominates the other, we can still ask whether we can improve on it. This could be motivated again by questioning the implicit location-scale assumption. If this

[1] Of course, in finite samples, even statistical acceptance of the efficient VaR condition for some particular VaR measure does not neccesarily imply that the efficient VaR condition is satisfied in population, as a lack of power against the relevant alternative could be the culprit.

assumption is violated, it is entirely possible that one VaR measure is superior to the other one, but neither of them satisfies the efficient VaR condition in population. Under these circumstances, it is possible to improve even the "better" VaR measure by combining it with a "worse" volatility measure. This concept of forecast combination was introduced to the literature on conditional mean forecasting by Bates and Granger (1969), and Stock and Watson (1998) have recently reported new evidence of its success in forecasting macroeconomic time series. So far, however, it has remained unexplored in the context of risk management.

We have now established a moment condition framework for VaR measures but we still need to find the distribution of the moment conditions. This task is complicated by the presence of the indicator function. As it always takes on a value of either zero or one, it introduces a nondifferentiability into the moment conditions. We will resolve this complication partly by using the results of Pakes and Pollard (1989) and partly by extending the framework of Kitamura (1997).

3 Methodology

Recall that, if the VaR measure is correctly specified, we must have

$$E\left[I\left(y_t < F_{t|t-1}\left(\beta_p\right)\right) - p\middle|\, \Psi_{t-1}\right] = 0. \tag{5}$$

Suppose that the instruments $\{z_t, z_{t-1}, \ldots\}$ are contained in the information set Ψ_t. Note that, by the law of iterated expectations, we should have

$$E\left[\left(I\left(y_t < F_{t|t-1}\left(\beta_p\right)\right) - p\right) \times k\left(z_{t-1}, z_{t-2}, \ldots\right)\right] = 0 \tag{6}$$

for every measurable vector-valued function $k(\cdot)$ of $\{z_t, z_{t-1}, \ldots\}$. For simplicity, omitting the time and p-subscripts, we may write equation (6) generically as $E\left[f(x, \beta)\right] = 0$, where the vector x contains the elements of z_t and y_t as well as σ_t.

3.1 VaR Specification Testing

Hansen's (1982) GMM overidentification test, sometimes known as the J-test, can be used to test the implication in (6). The test statistic is defined as

$$T\bar{f}_T\left(\widehat{\beta}\right)' W\bar{f}_T\left(\widehat{\beta}\right), \tag{7}$$

where

$$\widehat{\beta} = \arg\min_{\beta} \overline{f}_T\left(\beta\right)' W\overline{f}_T\left(\beta\right), \quad \overline{f}_T\left(\beta\right) = \frac{1}{T}\sum_{t=1}^{T} f\left(x_t,\beta\right),$$

and W is the optimal weighting matrix making GMM a consistent and asymptotically efficient estimator. It is clear that, due to the presence of the indicator function, $I\left(\cdot\right)$, the moment function $f\left(x, \beta\right)$ is not differentiable in β, which presents an econometric challenge. For specification testing, this challenge has been resolved by Pakes and Pollard (1989) who apply simulation-based techniques.

Although the standard GMM framework is thus suitable for specification testing of VaR measures, it is ill suited for nonnested comparisons of possibly misspecified models. This is the topic to which we now turn.

3.2 Nonnested VaR Comparison

For the specification test described at the end of the preceding subsection, we could in principle have relied on the information theoretic alternative to GMM due to Kitamura and Stutzer (1997), who consider solving the sample analog of the unconstrained problem

$$\beta^* = \underset{\beta}{\operatorname{argmax}} \min_{\gamma} E_\mu \left[\exp \left(\gamma' f \left(x, \beta \right) \right) \right]$$

i.e.,

$$\widehat{\beta}_T = \underset{\beta}{\operatorname{argmax}} \min_{\gamma} M_T \left(\beta, \gamma \right) = \underset{\beta}{\operatorname{argmax}} \min_{\gamma} \frac{1}{T} \sum_{t=1}^{T} \exp \left(\gamma' f \left(x_t, \beta \right) \right). \tag{8}$$

Their estimator is based on the intuition that, under correct specification, β^* minimizes the Kullback-Leibler Information criterion (KLIC). Interestingly, their interpretation has a nice generalization to the nonnested hypothesis testing as discussed by Kitamura (1997).

Suppose now that we are given two VaR measures, $F_{t|t-1}\left(\beta_p\right)$, and $F^*_{t|t-1}\left(\theta_p\right)$, the moment conditions of which can be written as:

$$E \left[f \left(x, \beta_p \right) \right] \equiv E \left[\left(I \left(y_t < F_{t|t-1} \left(\beta_p \right) \right) - p \right) \times k \left(z_{t-1} \right) \right] = 0$$

and

$$E \left[g \left(x, \theta_p \right) \right] \equiv E \left[\left(I \left(y_t < F^*_{t|t-1} \left(\theta_p \right) \right) - p \right) \times k \left(z_{t-1} \right) \right] = 0,$$

where $k(\cdot)$ is a given finite-dimensional vector-valued function. Note that neither VaR measure nests the other, and traditional nested hypothesis testing cannot be used for comparing these two VaR measures. This alone presents a theoretical challenge for VaR comparisons. We take an even more ambitious position by assuming that both specifications are potentially incorrect.

Kitamura (1997) proposed to deal with such nonnested hypothesis testing by comparing the KLIC distance of the two moment restrictions in population. Under his proposal, the moment restriction with smaller KLIC distance will be accepted: Our test will be based on the difference between the KLIC distances

$$M_T \left(\widehat{\beta}_T, \widehat{\gamma}_T \right) = \max_{\beta} \min_{\gamma} M_T \left(\beta, \gamma \right) \left(= \frac{1}{T} \sum_{t=1}^{T} \exp \left[\gamma' f \left(x_t, \beta \right) \right] \right)$$

and

$$N_T \left(\widehat{\theta}_T, \widehat{\lambda}_T \right) = \max_{\theta} \min_{\lambda} N_T \left(\theta, \lambda \right) \left(= \frac{1}{T} \sum_{t=1}^{T} \exp \left[\lambda' g \left(x_t, \theta \right) \right] \right).$$

Kitamura (1997) established the properties of such nonnested hypothesis testing for the case where both f and g are differentiable. Due to the indicator function, differentiability is violated in our application. We therefore generalize his result to our nondifferentiable case, and obtain the following result.

Theorem 1 *Under the null that* $M(\beta^*, \gamma^*) = N(\theta^*, \lambda^*)$, *we have*

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - N_T\left(\widehat{\theta}_T, \widehat{\lambda}_T\right)\right) \to N\left(0, \sigma_\infty^2\right),$$

where $\sigma_\infty^2 = \lim_{T\to\infty} \text{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \exp[\gamma^{*\prime} f(x_t, \beta^*)] - \exp[\lambda^{*\prime} g(x_t, \theta^*)]\right)$.

Proof. See Appendix.

Thus, a significantly large value of the test statistic will cause a rejection of the hypothesis that the two measures match the efficient VaR condition equally well in favor of the VaR model denoted by $E[g(x, \theta^*)] = 0$.

3.3 VaR Combination

Consider now that last of our three research questions: How can the risk manager optimally combine competing VaR measures? For simplicity and practicality, we focus on the linear combination of VaR measures. We consider combining the following n VaR measures

$$\Pr\left(y_t \leq F_{t|t-1}^{(i)}\left(\beta_p^{(i)}\right)\Big|\, \Psi_{t-1}^{(i)}\right) = p, \tag{9}$$

for $i = 1, 2, ..., n$. Let $\beta = \left(\beta_p^{(1)}, \beta_p^{(2)}, \cdots, \beta_p^{(n)}\right)$, $\pi(\pi_0, \pi_1, \cdots, \pi_n)$, and $\Psi_{t-1} = \cap_{i=1}^n \Psi_{t-1}^{(i)}$. Denote the combined VaR measure by

$$F_{t|t-1}(\beta, \pi) = \pi_0 + \sum_{i=1}^n \pi_i F_{t|t-1}^{(i)}\left(\beta_p^{(i)}\right).$$

Note that, if any of the individual VaR measures is correctly specified, then

$$\Pr\left(y_t \leq F_{t|t-1}(\beta, \pi)\big|\, \Psi_{t-1}\right) = p \tag{10}$$

for some π trivially. Even if none of the VaR measures are correctly specified, it is plausible that some linear combination would improve on all individual VaR measures. Notice that above we wrote the moment condition as

$$E\left[\left(I\left(y_t \leq F_{t|t-1}(\beta, \pi)\right) - p\right) \times k(z_{t-1})\right] = 0$$

which in turn can be written simply as

$$E[f(z, \beta, \pi)] = 0.$$

Note that we can consider π a part of the parameter vector in the moment condition, and the asymptotic results underlying Theorem 1 can be applied again. The resulting combined VaR measure is optimal in the sense that minimizes the Kullback-Leibler distance in population.

4 Application to Daily Returns on the S&P500

The focus of this application is to assess and compare the usefulness of different volatility measures in risk management. We apply our testing methodology to a portfolio consisting of a long position in the S&P500 index with an investment horizon of one day. The data applied was graciously provided to us by Chernov and Ghysels (1998). They provide us with S&P500 index returns which are recorded daily from November 1985 to October 1994, corresponding to 2209 observations. They also supply a daily European options price on the at-the-money, nearest to maturity, call option contract on the S&P500 index. Using the efficient GMM methodology of Gallant and Tauchen (1996), Chernov and Ghysels (1998) estimate the Heston's (1993) model in equation (4), and obtain a series of daily "fitted" volatilities, using the reprojection algorithm in Gallant and Tauchen (1998). We shall refer to these as reprojected volatilities below. In addition to the reprojected volatilities from Heston's model, Chernov and Ghysels produce a set of daily implied Black-Scholes volatilities defined from equation (3).

In addition to the two volatility series calculated from option prices, we apply two volatility measures based on the historical daily returns data. One is an estimated GARCH(1,1) volatility as in equation (2), the other is the so called RiskMetrics volatility which is constructed simply as an exponential filter of the squared returns, as in equation (1). As in RiskMetrics, we set the smoothing parameter, λ, to .94. The four standard deviation series are plotted in Figure 1.

For each of the volatility series, and at each desired VaR coverage, p, we run a simple quantile regression of returns on a constant and the time-varying standard deviation to get initial parameter estimates. We then optimize this first estimate using equation (8) to get a final parameter estimate, and thus a final VaR(p) measure for each model. We then turn to the testing of the four volatility measures for VaR purposes across a range of coverage values, p.

4.1 VaR Specification Testing

When testing each of the four VaRs for misspecification, we could of course use the well-known GMM J-test suggested in equation (7). However, in order to maintain continuity with the ensuing comparison tests, we will instead apply Kitamura and Stutzer's (1997) $\kappa-$test. The $\kappa-$test is the information theoretic version of the J-test, and it takes the form

$$\hat{\kappa}_T = -2T \log M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) = -2T \log\left(\frac{1}{T}\sum_{t=1}^{T} \exp\left[\hat{\gamma}' f\left(x_t, \hat{\beta}\right)\right]\right) \rightarrow \chi^2_{r-m},$$

where r is the number of moments, and m is the number of estimated parameters. We will test the VaR measures constructed from GARCH volatilities, RiskMetrics volatilities, implied volatilities, and reprojected volatilities from the daily S&P500 returns. We use a constant as well as the first lag of the four volatility measures as our linear conditioning information. As we are estimating two parameters: the constant and the slope on volatility. We have $r - m = (1 + 4) - 2 = 3$ degrees of freedom in the asymptotic χ^2 distribution. The specification testing results are summarized in the following table.

Table 1: Specification Testing Across VaR Coverage Rates

VaR Measure	$p = .01$	$p = .05$	$p = .10$	$p = .25$
GARCH Volatility	3.84	4.85	3.41	11.35
RiskMetrics	0.33	8.09	7.58	10.40
Implied Volatility	4.76	9.72	10.07	14.44
Reprojected Volatility	7.20	8.02	7.44	9.04

The $\chi^2(3)$ distribution has a 5 percent critical value of 7.82 and a 10 percent critical value of 6.25. Choosing the 5 percent level of significance, we see that no VaRs are rejected when the coverage rate $p = .01$, all but the GARCH VaR are rejected when $p = .05$, the implied volatility VaR is rejected when $p = .10$, and all VaRs are rejected when $p = .25$.

As volatility is inherently a symmetric phenomenon one might try, instead of using a VaR moment condition, to use an interval type moment condition, such as

$$E\left[I\left(-F_{t|t-1}(\beta_p) < y_t < F_{t|t-1}(\beta_p)\right) - 2p|\Psi_{t-1}\right] = 0. \tag{11}$$

We get the following results when testing the four VaR specifications using the double-sided interval:

Table 2: Double-sided Specification Testing

VaR Measure	$p = .01$	$p = .05$	$p = .10$	$p = .25$
GARCH Volatility	8.07	18.93	19.22	18.50
RiskMetrics	1.70	25.61	29.73	11.70
Implied Volatility	9.73	8.22	6.94	11.79
Reprojected Volatility	12.67	17.90	14.64	14.40

We now reject all VaR across p values with the exception of RiskMetrics for $p = .01$, and Implied volatility for $p = .10$.

An important implication of these results is that different VaRs might be optimal for different levels of coverage. This is not surprising as all the VaR models are no doubt misspecified. The important thing to note is that our testing framework allows the user to assess the quality of a VaR measure given the desired coverage probability, p. Should a risk manager want to test a model across a set of coverage rates, he or she could simply stack the moment conditions corresponding to each p in the set and run the test on all the conditions simultaneously.

4.2 Nonnested VaR Comparison Testing

In this section we perform the nonnested VaR comparison tests using the asymptotic result in Theorem 1. The results from performing pairwise comparison testing of the four competing VaRs are as follows:

Table 3: VaR Comparisons Across Coverage Rates

VaR Model 1 vs VaR Model 2:	$p = .01$	$p = .05$	$p = .10$	$p = .25$
GARCH vs RiskMetrics	-0.88	0.54	1.32	-0.24
GARCH vs Implied	0.21	0.65	0.93	0.48
GARCH vs Reprojected	0.59	0.40	0.55	-0.34
RiskMetrics vs Implied	1.30	0.17	0.32	1.06
RiskMetrics vs Reprojected	1.67	-0.01	-0.02	-0.28
Implied vs Reprojected	0.63	-0.44	-0.62	-1.90

Each entry in the table represents the test value from the null hypothesis of VaR Model 1 and VaR Model 2 being equally suitable. A value larger than 1.96 in absolute terms denotes a rejection of the null hypothesis at the 5 percent significance level, and a value larger than 1.65 denotes a rejection at the 10 percent level. A positive value indicates that VaR Model 1 is preferred, and a negative value that VaR Model 2 is preferred.

From the table, only a few rejections are possible, and only at the 10 percent significance level. At a VaR coverage of 1 percent, the RiskMetrics VaR is preferred to the reprojected volatility VaR. For $p = .25$, the reprojected volatility VaR is preferred to the implied volatility VaR.

When comparing the VaRs using the interval rather than the VaR criterion, we get the following results:

Table 4: Double-sided VaR Comparisons

VaR Model 1 vs VaR Model 2:	$p = .01$	$p = .05$	$p = .10$	$p = .25$
GARCH vs RiskMetrics	-1.24	0.93	1.33	-1.27
GARCH vs Implied	0.24	-0.92	-0.96	-0.53
GARCH vs Reprojected	0.57	-0.08	-0.33	-0.30
RiskMetrics vs Implied	1.26	-1.32	-1.65	0.01
RiskMetrics vs Reprojected	1.66	-0.52	-1.01	0.22
Implied vs Reprojected	0.43	1.17	1.16	0.36

Again, we see that RiskMetrics is preferred to the reprojected volatility VaR when $p = .01$, and now the implied volatility VaR is preferred to RiskMetrics when $p = .10$.

Notice that the comparison testing results in general correspond well to the inference drawn from the specification testing exercise above. For example, two VaRs which were both rejected in the specification tests typically receive a comparison test value close to zero. Notice also that even though we do not find a lot of evidence to significantly discern between VaR measures in the comparison tests, the test values will allow for an informal pairwise ranking of nonnested VaRs, even if their differences are not statistically significant.

Finally, we note that we might be able to significantly rank more models if we change the investment horizon from one to five or ten trading days. The GARCH and RiskMetrics models typically provide very similar short-term variance forecasts, but they have very different implications for the longer term.

5 Summary and Directions for Further Work

Risk managers have an abundance of Value-at-Risk methodologies to choose from. Consequently, we have considered specification tests of various VaR measures. From the perspective that relevant VaR measures should satisfy an efficient VaR condition, which we define, we have provided various methodologies with which such relevance can be tested. The methodology can (i) test whether a VaR measure satisfies the efficient VaR condition; (ii) compare two misspecified VaR measures; and (iii) combine several misspecified VaR measures. The usefulness of the new methodology was illustrated in an application to daily returns on the S&P500 index.

Several issues are left for future research. We have implicitly assumed away estimation errors in the volatility measures which is not completely justified in the GARCH(1,1) case. We have also assumed that the volatility measures are stationary. This is not without loss of generality, but we do not yet found an adequate yet theoretically rigorous way of incorporate such problems. In future work, we intend to address these issues.

A Appendix: Proof of Theorem 1

A.1 Assumptions

1. The process x_t is stationary.

2. $\beta \in \Theta$, a compact, m-dimensional set.

3. There exists a unique solution, (β^*, γ^*), to

$$\max_{\beta} \min_{\gamma} E\left[\exp\left(\gamma' f\left(x, \beta\right)\right)\right]$$

4. For sufficiently small $\delta > 0$,

$$E\left[\sup_{\beta' \in \Gamma(\beta,\delta)} \exp\left(g' f\left(x, \beta'\right)\right)\right] < \infty$$

   for all vectors g in the neighborhood of γ^*. Here, $\Gamma\left(\beta, \delta\right)$ denotes an open ball of radius δ around β.

5. $E[f(x, \beta) f(x, \beta)']$ is nonsingular for all β in Θ.

6. $f(x, \beta)$ belongs to a measurable V-C subgraph class of functions the p-th moment of which envelope function is finite.

7. The process x_t is β mixing with β mixing coefficients β_k satisfying

$$k^{p/(p-2)} \left(\log k\right)^{2(p-1)/(p-2)} \beta_k \to 0$$

   for some $2 < p < \infty$.

8. $D \equiv \frac{\partial^2}{\partial\gamma\partial\beta'} E[\exp(\gamma^{*\prime} f(x, \beta^*))]$ is of full column rank.
   $V \equiv \lim_{T\to\infty} \text{Var}\left[\frac{1}{T}\sum_{t=1}^{T} \exp(\gamma^{*\prime} f(x_t, \beta^*)) f(x_t, \beta^*)\right]$ is positive definite.

9. x_t is a continuous random variable, and there is an integrable $\Re^r-$valued function $F(x)$ such that

$$\begin{aligned} \left|\exp(\gamma' f(x, \beta)) f(x, \beta)\right| &\leq F(x), \\ \left|\exp(\gamma' f(x, \beta)) f(x, \beta)^2\right| &\leq F(x), \\ \left|\exp(\gamma' f(x, \beta)) f(x, \beta)^3\right| &\leq F(x), \end{aligned}$$

   for all x in a neighborhood of (β^*, γ^*), where $|\cdot|$, power, and $\leq$ are element-by-element.

We also impose conditions on θ, λ, and $g\left(x, \theta\right)$ which correspond to Assumptions 2-8.

A.2 Theorem

Let

$$\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) = \underset{\beta}{\operatorname{argmax}} \min_{\gamma} M_T(\beta, \gamma) \left(= \frac{1}{T} \sum_{t=1}^{T} \exp(\gamma' f(x_t, \beta)) \right),$$

$$\left(\widehat{\theta}_T, \widehat{\lambda}_T\right) = \underset{\theta}{\operatorname{argmax}} \min_{\lambda} N_T(\theta, \lambda) \left(= \frac{1}{T} \sum_{t=1}^{T} \exp(\lambda' g(x_t, \theta)) \right).$$

Under Assumptions 1-8,

(a) $\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) \xrightarrow{p} (\beta^*, \gamma^*)$.

(b)

$$\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right) \xrightarrow{d} N\left(0, \left(D'SD\right)^{-1} D'S^{-1}VS^{-1}D\left(D'SD\right)^{-1}\right)\right),$$

$$\sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right) \xrightarrow{d} N\left(0, \left(I_r - \left(D'SD\right)^{-1} D'\right) S^{-1}VS^{-1}\left(I_r - D\left(D'SD\right)^{-1}\right)\right),$$

where $S \equiv \frac{\partial^2}{\partial\gamma\partial\gamma'} E[\exp(\gamma^{*\prime} f(x, \beta^*))]$.

(c) Under the null hypothesis that $E\left[(\gamma^*)' f(x, \beta^*)\right] = E\left[(\lambda^*)' g(x, \theta^*)\right]$,

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - N_T\left(\widehat{\theta}_T, \widehat{\lambda}_T\right)\right) \xrightarrow{d} N\left(0, \sigma_\infty^2\right),$$

where $\sigma_\infty^2 = \lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\exp\left[(\gamma^*)' f(x_t, \beta^*)\right] - \exp\left[(\lambda^*)' g(x_t, \theta^*)\right]\right)\right)$.

A.3 Proofs

A.3.1 Proof of (a): consistency

Kitamura and Stutzer's (1997) consistency proof, which does not require differentiability, establishes that

1. $\gamma(\beta)$ is continuous in β under Assumption 5.

2. For some $L \equiv E\left\{\exp\left[\gamma(\beta^*)' f(x, \beta^*)\right]\right\}$

$$E\left\{\exp\left[\gamma(\beta)' f(x, \beta)\right]\right\} < L$$

and

$$\lim_{\delta\to 0} E\left[\sup_{\beta'\in\Gamma(\beta,\delta)} \exp\left(\gamma(\beta')' f(x, \beta')\right)\right] = E\left\{\exp\left[\gamma(\beta)' f(x, \beta)\right]\right\}.$$

3. For all $\delta > 0$, there exists some $h > 0$ such that

$$\lim_{T\to\infty} \Pr\left[\sup_{\beta'\in\Theta-\Gamma(\beta^*,\delta)} \frac{1}{T} \sum_{t=1}^{T} \exp\left(\gamma(\beta')' f(x_t, \beta')\right) > L - h\right] = 0. \quad (12)$$

4. By definition,

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T(\beta)' f(x_t,\beta)\right) \leq \frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma(\beta)' f(x_t,\beta)\right),$$

which, when combined with (12), yields

$$\lim_{T\to\infty}\Pr\left[\sup_{\beta'\in\Theta-\Gamma(\beta^*,\delta)}\frac{1}{T}\sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T(\beta')' f(x_t,\beta')\right) > L-h\right]=0. \tag{13}$$

Consider the minimization of

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f(x_t,\beta^*)\right)$$

with the corresponding minimizer $\widehat{\gamma}_T(\beta^*)$. Note that, for fixed g, we have by Taylor expansion

$$\begin{aligned}
&\sum_{t=1}^{T}\exp\left(\left(\gamma^*+\frac{1}{\sqrt{T}}g\right)' f(x_t,\beta^*)\right)-\sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f(x_t,\beta^*)\right)\\
&\quad=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f(x_t,\beta^*)\right) f(x_t,\beta^*)'\cdot g\\
&\qquad+\frac{1}{2}g'\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2}{\partial\gamma\partial\gamma'}\exp\left(\gamma^{*\prime} f(x_t,\beta^*)\right)\right\}g+o_p(1),
\end{aligned}$$

from which we obtain

$$\begin{aligned}
&\sum_{t=1}^{T}\exp\left(\left(\gamma^*+\frac{1}{\sqrt{T}}g\right)' f(x_t,\beta^*)\right)-\sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f(x_t,\beta^*)\right)\\
&= \frac{1}{2}g'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime} f(x,\beta^*)\right)\right]\right\}g\\
&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f(x_t,\beta^*)\right) f(x_t,\beta^*)'\cdot g+o_p(1).
\end{aligned}\tag{14}$$

Pollard's (1991) convexity lemma strengthens the convergence to uniform convergence: The last $o_p(1)$ is now over any compact set of gs. By replicating Pollard's (1991) argument, we can prove that

$$\begin{aligned}
\sqrt{T}\left(\widehat{\gamma}_T(\beta^*)-\gamma(\beta^*)\right) &= -\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma' f(x,\beta^*)\right)\right]\right\}^{-1}\\
&\quad\times\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f(x_t,\beta^*)\right] f(x_t,\beta^*)\right\}+o_p(1).
\end{aligned}\tag{15}$$

Now, observe that

$$E\left[\exp\left[\gamma^{*\prime} f(x,\beta^*)\right] f(x,\beta^*)\right]=0$$

because it is the first order condition for minimization of $E\left[\exp\left[\gamma' f\left(x,\beta^*\right)\right]\right]$ with respect to γ. Central limit theorem then implies that

$$\widehat{\gamma}_T\left(\beta^*\right)=\gamma\left(\beta^*\right)+O_p\left(\frac{1}{\sqrt{T}}\right)=\gamma\left(\beta^*\right)+o_p\left(1\right). \tag{16}$$

Furthermore, plugging (15) into (14), we obtain

$$\begin{aligned}
&\sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T\left(\beta^*\right)' f\left(x_t,\beta^*\right)\right)-\sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f\left(x_t,\beta^*\right)\right)\\
&= \quad -\frac{1}{2}\frac{1}{T}\left\{\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t,\beta^*\right)\right] f\left(x_t,\beta^*\right)\right\}'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime} f\left(x,\beta^*\right)\right)\right]\right\}^{-1}\\
&\qquad\times\left\{\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t,\beta^*\right)\right] f\left(x_t,\beta^*\right)\right\}+o_p\left(1\right).
\end{aligned} \tag{17}$$

Note that (16) in particular implies that

$$\lim_{n\to\infty}\Pr\left[\frac{1}{T}\sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T\left(\beta^*\right)' f\left(x_t,\beta^*\right)\right)<L-h\right]=0. \tag{18}$$

Combination of (13) and (18) delivers the consistency of $\widehat{\beta}_T$.

We now show that $\widehat{\gamma}_T=\widehat{\gamma}_T\left(\widehat{\beta}\right)\xrightarrow{p}\gamma^*$. This can be accomplished by the convexity lemma again. For this purpose, we note that

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\widehat{\beta}\right)\right)\xrightarrow{p}E\left[\exp\left(\gamma' f\left(x,\beta^*\right)\right)\right]$$

uniformly in γ within some compact neighborhood of γ^*. This can be proved as follows. Theorem 2.1 of Arcones and Yu (1994) and Theorem 10.2 of Pollard (1990) imply the stochastic equicontinuity of $\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\cdot\right)\right)$. For fixed γ, the consistency of $\widehat{\beta}_T$ and this stochastic equicontinuity deliver

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\widehat{\beta}_T\right)\right)-\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\beta^*\right)\right)=h\left(\gamma,\widehat{\beta}_T\right)-h\left(\gamma,\beta^*\right)+o_p\left(1\right),$$

where

$$h\left(\gamma,\beta\right)\equiv E\left[\exp\left(\gamma' f\left(x,\beta\right)\right)\right].$$

For fixed γ, the almost sure continuity of $f\left(x,\beta\right)$ in β combined with Assumption 9 delivers

$$h\left(\gamma,\widehat{\beta}_T\right)-h\left(\gamma,\beta^*\right)=o_p\left(1\right).$$

We therefore have

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\widehat{\beta}_T\right)\right)-\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\beta^*\right)\right)=o_p\left(1\right)$$

pointwise in γ. Note that

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\beta^*\right)\right) = E\left[\exp\left(\gamma' f\left(x,\beta^*\right)\right)\right] + o_p\left(1\right)$$

by the ergodic theorem. We therefore have

$$\frac{1}{T}\sum_{t=1}^{T}\exp\left(\gamma' f\left(x_t,\widehat{\beta}_T\right)\right) = E\left[\exp\left(\gamma' f\left(x,\beta^*\right)\right)\right] + o_p\left(1\right)$$

for each fixed γ. Convexity strengthens it to uniform convergence, and the consistency of $\widehat{\gamma}_T\left(\widehat{\beta}_T\right)$ follows.

A.3.2 Proof of (b): Root-T-Consistency

We first show that

$$\frac{1}{T}\sum_{t=1}^{T} f\left(x_t,\widehat{\beta}_T\right)\exp\left[\gamma^{*\prime} f\left(x_t,\widehat{\beta}_T\right)\right] = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Given an arbitrary vector g, let

$$g_T = \gamma^* + \frac{1}{\sqrt{T}}g.$$

By definition, we should have

$$\sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T\left(\beta^*\right)' f\left(x_t,\beta^*\right)\right) \leq \sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T\left(\widehat{\beta}_T\right)' f\left(x_t,\widehat{\beta}_T\right)\right) \leq \sum_{t=1}^{T}\exp\left(g_T, f\left(x_t,\widehat{\beta}_T\right)\right).$$

By (17), we have

$$\begin{aligned}
&\sum_{t=1}^{T}\exp\left(\widehat{\gamma}_T\left(\beta^*\right)' f\left(x_t,\beta^*\right)\right) - \sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f\left(x_t,\beta^*\right)\right) \\
&= -\frac{1}{2}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t,\beta^*\right)\right] f\left(x_t,\beta^*\right)\right\}'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime} f\left(x,\beta^*\right)\right)\right]\right\}^{-1} \\
&\quad\times\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t,\beta^*\right)\right] f\left(x_t,\beta^*\right)\right\} + o_p\left(1\right). \qquad (19)
\end{aligned}$$

As for the last expression, we can make use of basically the same argument that leads to the quadratic approximation (14), and conclude that

$$\begin{aligned}
&\sum_{t=1}^{T}\exp\left(\left(\gamma^* + \frac{1}{\sqrt{T}}g\right)' f\left(x_t,\widehat{\beta}_T\right)\right) - \sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f\left(x_t,\beta^*\right)\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime} f\left(x_t,\widehat{\beta}_T\right)\right) f\left(x_t,\widehat{\beta}_T\right)'\cdot g \\
&\quad+\frac{1}{2}g'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime} f\left(x,\beta^*\right)\right)\right]\right\} g + o_p\left(1\right). \qquad (20)
\end{aligned}$$

We therefore have

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime}f\left(x_t,\widehat{\beta}_T\right)\right)f\left(x_t,\widehat{\beta}_T\right)'\cdot g+\frac{1}{2}g'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime}f\left(x,\beta^*\right)\right)\right]\right\}g+o_p\left(1\right)\\
&\qquad\geq-\frac{1}{2}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime}f\left(x_t,\beta^*\right)\right]f\left(x_t,\beta^*\right)\right\}'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime}f\left(x,\beta^*\right)\right)\right]\right\}^{-1}\\
&\qquad\qquad\times\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime}f\left(x_t,\beta^*\right)\right]f\left(x_t,\beta^*\right)\right\}.
\end{aligned}
$$

With a little bit of algebra, we can show that this actually implies that

$$
\begin{aligned}
&\frac{1}{2}\left(g+\xi_T\right)'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime}f\left(x,\beta^*\right)\right)\right]\right\}\left(g+\xi_T\right)\\
&\quad\geq\quad\left(\xi_T-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime}f\left(x_t,\widehat{\beta}_T\right)\right)f\left(x_t,\widehat{\beta}_T\right)'\right)\cdot g+o_p\left(1\right).
\end{aligned}
\tag{21}
$$

where

$$
\xi_T=\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime}f\left(x,\beta^*\right)\right)\right]\right\}^{-1}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime}f\left(x_t,\beta^*\right)\right]f\left(x_t,\beta^*\right)\right\}=O_p\left(1\right).
$$

Now recall from the convexity lemma that the $o_p(1)$ terms in (19) and (20) hold uniformly in compact sets of g. Therefore, we can replace g in (21) with any arbitrary $O_p(1)$ random vectors, say g_T, and rewrite (21) as

$$
\begin{aligned}
&\frac{1}{2}\left(g_T+\xi_T\right)'\left\{\frac{\partial^2}{\partial\gamma\partial\gamma'}E\left[\exp\left(\gamma^{*\prime}f\left(x,\beta^*\right)\right)\right]\right\}\left(g_T+\xi_T\right)\\
&\qquad\qquad\geq\left(\xi_T-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime}f\left(x_t,\widehat{\beta}_T\right)\right)f\left(x_t,\widehat{\beta}_T\right)'\right)\cdot g_T+o_p\left(1\right).
\end{aligned}
$$

If this is to hold for arbitrary $g_T=O_p(1)$, we should have

$$
\xi_T-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left(\gamma^{*\prime}f\left(x_t,\widehat{\beta}_T\right)\right)f\left(x_t,\widehat{\beta}_T\right)'=O_p\left(1\right).
$$

We therefore conclude that

$$
\frac{1}{T}\sum_{t=1}^{T}f\left(x_t,\widehat{\beta}_T\right)\exp\left[\gamma^{*\prime}f\left(x_t,\widehat{\beta}_T\right)\right]=O_p\left(\frac{1}{\sqrt{T}}\right).
$$

Now, note that

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime}f\left(x_t,\widehat{\beta}_T\right)\right]f\left(x_t,\widehat{\beta}_T\right)\\
&\quad=\quad\frac{1}{T}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime}f\left(x_t,\beta^*\right)\right]f\left(x_t,\beta^*\right)+D\left(\widehat{\beta}_T-\beta^*\right)+o_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}
\tag{22}
$$

by the usual stochastic equicontinuity argument, where the stochastic equicontinuity follows from Theorem 2.1 of Arcones and Yu and Theorem 10.2 of Pollard (1990).

Because

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f\left(x_t, \widehat{\beta}_T\right) = O_p(1), \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \exp\left[\gamma^{*\prime} f(x_t, \beta^*)\right] f(x, \beta^*) = O_p(1),$$

we have

$$D\left(\widehat{\beta}_T - \beta^*\right) = O_p\left(\frac{1}{\sqrt{T}}\right)$$

or

$$\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right) = O_p(1).$$

A.3.3 Asymptotic Distribution

Now, combine (20) and (22), and obtain

$$\begin{aligned}\sum_{t=1}^{T} \exp\left(\left(\gamma^* + \frac{1}{\sqrt{T}}g\right)' f\left(x_t, \widehat{\beta}_T\right)\right) &= \sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f(x_t, \beta^*)\right) \\ &+ \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \exp\left[\gamma^{*\prime} f(x_t, \beta^*)\right] f(x_t, \beta^*) + D\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right)\right] \cdot g + \frac{1}{2}g' S g + o_p(1),\end{aligned}$$

where the $o_p(1)$ is uniform in compact sets of gs. By repeating Pollard's (1991) argument, we can obtain

$$\sqrt{T}\left(\widehat{\gamma}_T\left(\widehat{\beta}_T\right) - \gamma^*\right) \tag{23}$$

$$= S^{-1} \times \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \exp\left[\gamma^{*\prime} f(x_t, \beta^*)\right] f(x_t, \beta^*) + D\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right)\right] + o_p(1), \tag{24}$$

from which we can obtain

$$\begin{aligned}\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right) &= -\left(D' S^{-1} D\right)^{-1} D' S^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(x_t, \beta^*) \exp\left[\gamma^{*\prime} f(x_t, \beta^*)\right] \\ &\quad -\sqrt{T} D' \left(\widehat{\gamma}_T - \gamma^*\right) + o_p(1).\end{aligned} \tag{25}$$

Now, consider the maximization of

$$M_T\left(\widehat{\gamma}_T, \beta\right) \equiv \frac{1}{T}\sum_{i=1}^{n} \exp\left[\widehat{\gamma}_T' f(x_t, \beta)\right]$$

with respect to β. First note that, by definition, we have

$$M_T\left(\widehat{\gamma}_T, \beta^*\right) \leq M_T\left(\widehat{\gamma}_T, \widehat{\beta}_T\right).$$

An empirical process argument again delivers

$$M_T\left(\widehat{\gamma}_T, \widehat{\beta}_T\right) = M\left(\widehat{\gamma}_T, \widehat{\beta}_T\right) + M_T\left(\widehat{\gamma}_T, \beta^*\right) - M\left(\widehat{\gamma}_T, \beta^*\right) + o_p\left(\frac{1}{\sqrt{T}}\right).$$

Combined with the above inequality, we obtain

$$M\left(\widehat{\gamma}_T, \widehat{\beta}_T\right) - M\left(\widehat{\gamma}, \beta^*\right) + o_p\left(\frac{1}{\sqrt{T}}\right) \geq 0.$$

Now, let $\widetilde{\beta}_T$ be an arbitrary sequence of random vectors in the $\sqrt{T}$ neighborhood of β^*: $\sqrt{T}\left(\widetilde{\beta}_T - \beta^*\right) = O_p(1)$. Recall the quadratic approximation:

$$\sum_{t=1}^{T} \exp\left(\left(\gamma^* + \frac{1}{\sqrt{T}} g\right)' f\left(x_t, \widehat{\beta}_T\right)\right) - \sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \beta^*\right)\right)$$
$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widehat{\beta}_T\right)\right) f\left(x_t, \widehat{\beta}_T\right)' \cdot g + \frac{1}{2} g' \left\{\frac{\partial^2}{\partial\gamma\partial\gamma'} E\left[\exp\left(\gamma^{*\prime} f\left(x, \beta^*\right)\right)\right]\right\} g + o_p(1).$$

Therefore, we have

$$M_T\left(\widehat{\gamma}_T, \widehat{\beta}_T\right) = \frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \beta^*\right)\right) + \frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widehat{\beta}_T\right)\right) f\left(x_t, \widehat{\beta}_T\right)' \cdot \sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right)$$
$$+ \frac{1}{2}\left(\widehat{\gamma}_T - \gamma^*\right)' \left\{\frac{\partial^2}{\partial\gamma\partial\gamma'} E\left[\exp\left(\gamma^{*\prime} f\left(x, \beta^*\right)\right)\right]\right\}\left(\widehat{\gamma}_T - \gamma^*\right) + o_p\left(\frac{1}{\sqrt{T}}\right).$$

We can similarly establish

$$M_T\left(\widehat{\gamma}_T, \widetilde{\beta}_T\right) = \frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \beta^*\right)\right) + \frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widetilde{\beta}_T\right)\right) f\left(x_t, \widetilde{\beta}_T\right)' \cdot \sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right)$$
$$+ \frac{1}{2}\left(\widehat{\gamma}_T - \gamma^*\right)' \left\{\frac{\partial^2}{\partial\gamma\partial\gamma'} E\left[\exp\left(\gamma^{*\prime} f\left(x, \beta^*\right)\right)\right]\right\}\left(\widehat{\gamma}_T - \gamma^*\right) + o_p\left(\frac{1}{\sqrt{T}}\right)$$

By definition, we should have

$$M_T\left(\widehat{\gamma}_T, \widehat{\beta}_T\right) \geq M_T\left(\widehat{\gamma}_T, \widetilde{\beta}_T\right)$$

Therefore, we should have

$$\frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widehat{\beta}_T\right)\right) f\left(x_t, \widehat{\beta}_T\right)' \cdot \left(\widehat{\gamma}_T - \gamma^*\right)$$
$$\geq \frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widetilde{\beta}_T\right)\right) f\left(x_t, \widetilde{\beta}_T\right)' \cdot \left(\widehat{\gamma}_T - \gamma^*\right) + o_p\left(\frac{1}{\sqrt{T}}\right).$$

By the usual stochastic equicontinuity argument, we can establish

$$\frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widehat{\beta}_T\right)\right) f\left(x_t, \widehat{\beta}_T\right) - \frac{1}{T}\sum_{t=1}^{T} \exp\left(\gamma^{*\prime} f\left(x_t, \widetilde{\beta}_T\right)\right) f\left(x_t, \widetilde{\beta}_T\right)$$
$$= D\left(\widehat{\beta}_T - \widetilde{\beta}_T\right) + o_p\left(\frac{1}{\sqrt{T}}\right)$$

Therefore, we obtain

$$\left(\widehat{\beta}_T - \widetilde{\beta}_T\right)' D\left(\widehat{\gamma}_T - \gamma^*\right) + o_p\left(\frac{1}{T}\right) \geq 0$$

or

$$\sqrt{T}\left(\widehat{\beta}_T - \widetilde{\beta}_T\right)' D\sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right) \geq o_p\left(1\right).$$

Because this holds for arbitrary sequence $\widetilde{\beta}_T$ in the $\sqrt{T}$ neighborhood of β^*, we should have

$$\sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right)' D = o_p\left(1\right).$$

Getting back to (25), we obtain

$$\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right) = -\left(D'S^{-1}D\right)^{-1} D'S^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t, \beta^*\right)\right] f\left(x_t, \beta^*\right) + o_p\left(1\right).$$

Returning to (24), we also obtain

$$\begin{aligned}\sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right) = S^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t, \beta^*\right)\right] f\left(x_t, \beta^*\right) \\ - S^{-1}D.\left(D'S^{-1}D\right)^{-1} D'S^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\exp\left[\gamma^{*\prime} f\left(x_t, \beta^*\right)\right] f\left(x_t, \beta^*\right) + o_p\left(1\right).\end{aligned}$$

A.3.4 Proof of (c): nonnested hypothesis testing

Suppose we want to compare

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - N_T\left(\widehat{\theta}_T, \widehat{\lambda}_T\right)\right)$$

Note that

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - M\left(\beta^*, \gamma^*\right)\right) = \sqrt{T}\left(M\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - M_T\left(\beta^*, \gamma^*\right)\right) + o_p\left(1\right)$$

by the usual stochastic equicontinuity argument. Also note that

$$\sqrt{T}\left(M\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - M\left(\beta^*, \gamma^*\right)\right) = \frac{\partial M}{\partial \beta'}\cdot\sqrt{T}\left(\widehat{\beta}_T - \beta^*\right) + \frac{\partial M}{\partial \gamma'}\cdot\sqrt{T}\left(\widehat{\gamma}_T - \gamma^*\right) + o_p\left(1\right) = o_p\left(1\right),$$

which follows from

$$\frac{\partial M}{\partial \beta'} = 0, \quad \frac{\partial M}{\partial \gamma'} = 0$$

which in turn follows from the first order condition in the population. Therefore, we have

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - M\left(\beta^*, \gamma^*\right)\right) = \sqrt{T}\left(M_T\left(\beta^*, \gamma^*\right) - M\left(\beta^*, \gamma^*\right)\right) + o_p\left(1\right).$$

Similarly, we obtain

$$\sqrt{T}\left(N_T\left(\widehat{\theta}_T, \widehat{\lambda}_T\right) - N\left(\theta^*, \lambda^*\right)\right) = \sqrt{T}\left(N_T\left(\theta^*, \lambda^*\right) - N\left(\theta^*, \lambda^*\right)\right) + o_p\left(1\right).$$

Under the null that

$$M\left(\beta^*, \gamma^*\right) = N\left(\theta^*, \lambda^*\right),$$

we have

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - N_T\left(\widehat{\theta}_T, \widehat{\lambda}_T\right)\right) = \sqrt{T}\left(M_T\left(\beta^*, \gamma^*\right) - N_T\left(\theta^*, \lambda^*\right)\right) + o_p\left(1\right).$$

Therefore, under the null

$$\sqrt{T}\left(M_T\left(\widehat{\beta}_T, \widehat{\gamma}_T\right) - N_T\left(\widehat{\theta}_T, \widehat{\lambda}_T\right)\right) \rightarrow N\left(0, \sigma_\infty^2\right).$$

References

[1] Andersen, T. and T. Bollerslev (1998), "Answering the Critics: Yes, ARCH Models do Provide Good Volatility Forecasts," *International Economic Review*, 39, 885-905.

[2] Andersen, T., T. Bollerslev, F. Diebold and P. Labys (1999), "The Distribution of Exchange Rate Volatility," Manuscript, University of Pennsylvania.

[3] Arcones, M.A. and B. Yu (1994) "Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences," *Journal of Theoretical Probability,* 7, 47–71.

[4] Bates, J. and C. Granger (1969), "The Combination of Forecasts," *Operations Research Quarterly*, 20, 451-468.

[5] Beder, T. (1995), "VaR: Seductive but Dangerous," *Financial Analysts Journal*, September-October, 12-24.

[6] Black, F. and M. Scholes (1972), "The Valuation of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.

[7] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

[8] Chernov, M. and E. Ghysels (1998), "What Data Should be Used to Price Options?" Manuscript, Department of Finance, Penn State University.

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

[10] Danielsson, J. and C. de Vries (1997), "Value at Risk and Extreme Returns," Manuscript, London School of Economics.

[11] Dimson, E. and P. Marsh (1995), "Capital Requirements for Securities Firms," *Journal of Finance*, 50, 821-851.

[12] Dowd, K. (1998), *Beyond Value at Risk: The New Science of Risk Management*. Wiley.

[13] Duffie, D. (1997), "An Overview of Value at Risk," *Journal of Derivatives*, 4, 7-49.

[14] Engle. R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

[15] Engle, R. and S. Manganelli (1999), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," Manuscript, UCSD.

[16] Gallant, R. and G. Tauchen (1996), "Which Moments to Match?," *Econometric Theory*, 12, 657-681.

[17] Gallant, R. and G. Tauchen (1998), "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions, *Journal of the American Statistical Association*, 93, 10-24.

[18] Garman, M. and M. Klass (1980), "On the Estimation of Security Price Volatilities from Historical Data," *Journal of Business*, 53, 67-78.

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

[20] Heston, S. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6, 327-343.

[21] Jorion, P. (1995), "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance*, 50, 507-528.

[22] Jorion, P. (1996), *Value at Risk: The New Benchmark for Controlling Market Risk*. Irwin Professional.

[23] JP Morgan (1996), *RiskMetrics*, Technical Document. 4th Edition. New York.

[24] Kitamura, Y. (1997), "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," Manuscript, Department of Economics, University of Wisconsin, Madison.

[25] Kitamura, Y. and M. Stutzer (1997), "Information Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica,* 65, 861–874.

[26] Pakes, A., and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027 - 57.

[27] Pollard, D. (1990), *Empirical Processes: Theory and Applications,* Institute of Mathematical Statistics, Hayward, CA.

[28] Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory,* 7, 186–199.

[29] Risk (1999), "Risk Software Survey 1998" *Risk*, January, pp. 67-80.

[30] Stock, J. and M. Watson (1998), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," Manuscript, Kennedy School of Government, Harvard University.

[31] Wagster, J. (1996), "Impact of the 1988 Basle Accord on International Banks," *Journal of Finance*, 51, 1321-1346.

Figure 1: Four Volatility Measures of Daily S&P500 Returns