



CAEPR Working Paper  
#010-2010

## Backtesting Portfolio Value-at-Risk with Estimated Portfolio Weights

Pei Pei  
Indiana University

November 21, 2010

This paper can be downloaded without charge from the Social Science Research Network electronic library at: <http://ssrn.com/abstract=1722177>.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: <http://www.indiana.edu/~caepr>. CAEPR can be reached via email at [caepr@indiana.edu](mailto:caepr@indiana.edu) or via phone at 812-855-4050.

©2010 by Pei Pei. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

# Backtesting Portfolio Value-at-Risk with Estimated Portfolio Weights

Pei, Pei \*  
Department of Economics  
Indiana University

## Abstract

This paper theoretically and empirically analyzes backtesting portfolio VaR with estimation risk in an intrinsically multivariate framework. For the first time in the literature, it takes into account the estimation of portfolio weights in forecasting portfolio VaR and its impact on backtesting. It shows that the estimation risk from estimating the portfolio weights as well as that from estimating the multivariate dynamic model of asset returns make the existing methods in a univariate framework inapplicable. And it proposes a general theory to quantify estimation risk applicable to the present problem and suggests practitioners a simple but effective way to carry out valid inference to overcome the effect of estimation risk in backtesting portfolio VaR. A simulation exercise illustrates our theoretical findings. In application, a portfolio of three stocks is considered.

**Keywords and Phrases:** Backtesting; Portfolio; Estimation Risk; Forecast evaluation; Risk management; Value at Risk.

**JEL:** C52, C32, G32

---

\*Contact information: Department of Economics, Indiana University, Wylie Hall 307, 100 S. Woodlawn, Bloomington, IN 47405. Email: ppei@indiana.edu. Phone number: 812-855-9531. I am especially grateful to Juan Carlos Escanciano, my advisor, for his valuable advice and encouragement during the formation of this paper. In addition, I thank Javier Mencia for sharing Matlab codes for his paper, Yongjin Lee, Kim Huynh and Todd Walker for their comments and Zaichao Du for helpful conversations. All errors are my own.

# 1 Introduction

The literature on financial risk management primarily focuses on the context of given or hypothetical portfolios, e.g. Giot and Laurent (2003), Ferreira and Lopez (2005)<sup>1</sup>, and little attention has been paid to the fact that portfolio weights are unknown and estimated in practice using some portfolio optimization theory, therefore, the extra uncertainty from estimating portfolio weights has been neglected in inference problems on portfolios such as backtesting portfolio Value-at-Risk (VaR). The objective of this paper is to study the impact of estimation risk on backtesting portfolio VaR with the consideration of portfolio choice and to suggest a practical way to carry out valid inference in backtesting portfolio VaR free of estimation risk.

VaR has become the standard risk measure used in financial institutions, since adopted by the Basel Accord (1996a). It is defined as the maximum expected loss on an investment over a specified horizon at a given confidence level, see Jorion (2001). Required as part of capital-adequacy framework, backtesting, which is a statistical framework to evaluate the out-of-sample forecast accuracy of the portfolio VaR model recommended by the Basel Accord (1996b), has become an important issue in practice.

This paper tackles backtesting portfolio VaR with estimation risk in a complete multivariate setting, since backtesting portfolio VaR is intrinsically a multivariate inference problem. As argued in Giot and Laurent (2003) and Bauwens et. al. (2006), whenever portfolio of assets are involved, a multivariate dynamic model of the component asset returns would be needed for determining portfolio weights as well as forecasting asset returns. Unlike univariate modeling, the multivariate models capture time-varying correlations between the component asset returns and are also more flexible for obtaining the implied distribution of any portfolio. In forecasting portfolio VaRs, portfolio returns are unobservable but can be directly computed from forecasted asset returns and estimated asset weights.

Consequently, there are two sources of estimation risk in backtesting portfolio VaR, one from estimating the multivariate dynamic model of asset returns and one from estimating portfolio weights. Without considering the impact of estimation risk in the standard backtesting procedure, wrong critical values may be used to assess market risk, see Escanciano and Olmo (2009). But the complication in our context makes the existing methods in the univariate framework inapplicable. Thus, we provide a general theory to quantify the two sources of estimation risk in the multivariate framework of backtesting portfolio VaR. As far as we

are concerned this is the first work to incorporate extra uncertainty about estimating portfolio weights into the backtesting procedure in a complete multivariate setting.

In fact, the estimation risk issue has not been paid much attention in risk management literature. It has been either neglected or overcome by means of complicated methods. For example, the effect of estimation risk on optimal portfolio choice first discussed in Klein and Bawa (1976) is mainly examined using the Bayesian predictive approach, e.g. Kandel and Stambaugh (1996) and Barberis (2000). Jorion (1996) and Dowd (2000) study the estimation risk issues on VaR, but just for the i.i.d. return case. Christoffersen and Gonçalves (2005) examines the issue in the generalized autoregressive conditional heteroscedasticity (GARCH) models using a bootstrap method, but their method is time-consuming and only valid for i.i.d. standardized innovations. Furthermore, Gao and Song (2008) provides an analytical method to deal with estimation risk in GARCH VaR and expected shortfall estimates, and Escanciano and Olmo (2009) quantifies the estimation risk in backtesting VaR. However, all the above works are restricted to the univariate framework. The theory to be presented in this paper is applicable in the multivariate framework, especially with the consideration of portfolio choice.

One of the theoretical findings of this paper is that the effect of estimation risk on backtesting portfolio VaR tends to vanish as the ratio of the out-of-sample size relative to the in-sample size goes to zero. We design a series of simulation experiments to illustrate our theoretical findings. The simulation results turn out to support our theoretical findings. We conclude that a simple but effective way to carry out valid inferences in backtesting procedures is to consider a small ratio of the out-of-sample size to the in-sample size.

Although the general theory to be presented in this paper does not require any particular distributional assumptions for asset returns and any particular method of portfolio choice, we consider these two problems in detail in our application. In modeling asset returns, we focus on the parametric multivariate generalized autoregressive conditional heteroscedasticity models (MGARCH), which is the most popular modelling to capture the salient empirical features of volatility clustering and time-varying correlations from financial time series. There is a large body of literature on MGARCH, see two recent surveys of Bauwens et. al. (2006) and Silvennoinen and Terasvirta (2008) for a review. Meanwhile, it is crucial to make the distributional assumption of the innovations. We will use the the general hyperbolic (GH) distribution. As analyzed in Mencia and

Sentana (2005), the GH distribution is adequate to model positive excess kurtosis and negative skewness of financial asset returns in conditionally heteroscedastic dynamic regression models, and it is a rather flexible distribution that contains many well-known subclasses, including the most important distributions already used in the literature such as the Normal and the skew-Student. Additionally, the GH distribution is closed under linear transformations, which will have important implication for the use of the model in applications such as portfolio allocation and portfolio VaR calculation.

In allocating assets, we use the so-called mean-variance-skewness (MVS) analysis other than the widely used mean-variance (MV) analysis. Firstly because the effect of higher order moments on asset allocation cannot be neglected considering the asymmetries of asset returns being modeled. Secondly, under our model setup asset returns will jointly follow a GH distribution, which can be expressed as a location-scale mixture of normals. Mencia and Sentana (2009) shows that the distribution of any portfolio whose components jointly follow a location-scale mixture of normals will be uniquely characterized by its mean, variance and skewness. Most attractively, the closed form solution for the optimal portfolio weights could be explicitly obtained under our model setup by this MVS method.

The remaining of this paper is structured as follows. Section 2 provides the general theory that quantifies estimation risk in backtesting optimal portfolio VaR without any particular distributional assumption for asset returns and any particular method of optimal portfolio choice, and proposes a way to overcome the effect of estimation risk on backtesting. Section 3 applies the general procedures to a multivariate parametric setting in which asset returns are model by a MGARCH model with standardized GH innovations and the optimal portfolio weights are chosen by MVS method. A series of simulation exercises is designed to illustrate our theoretical findings. Section 4 considers an application to a portfolio of three stocks and compares differences in the inferences by different ways of backtesting optimal portfolio VaR. Finally, section 5 concludes.

## 2 Backtesting portfolio VaR robust to estimation risk: A general theory

The essence of backtesting is the out-of-sample comparison of actual trading results with model-generated risk measures. In backtesting portfolio VaR, both asset returns and asset allocation needs to be considered. In order to examine the effects of estimation risk on backtesting portfolio VaR, we need to elaborate on the forecast evaluation problem first.

### 2.1 Forecast evaluation problem

Let us consider a portfolio of  $d$  assets. Let  $r_t = (r_{1t}, r_{2t}, \dots, r_{dt})'$  denote the  $d$ -dimensional vector of stationary asset returns combined in the portfolio. Assume that at time  $t - 1$  the investor's information set is given by  $I_{t-1}$ , which may contain past values of  $r_t$  and other relevant economic and financial variables  $z_t$ , i.e.  $I_{t-1} = (r'_{t-1}, z'_{t-1}, r'_{t-2}, z'_{t-2}, \dots)'$ , while the portfolio weights,  $w_t = (w_{1t}, w_{2t}, \dots, w_{dt})'$ , where  $w_t \in \mathbb{R}^d$  and  $\sum_{i=1}^d w_{it} = 1$ , are unknown and need to be estimated at time  $t$  conditioning on the information available up to time  $t - 1$  by using any portfolio choice method. To make this explicitly, we write  $w_t \equiv w(I_{t-1})$ , where  $w_t \in \mathcal{F}_{t-1}$ . Obviously,  $w_t$  can be treated as a constant at time  $t$ , once we condition on the information set at time  $t - 1$ . Notice that no particular portfolio choice method is specified here so that the theory to be presented covers all existing methods in the literature. Thus the unobserved portfolio return at time  $t$  can be calculated by the linear projection,  $Y_t(w_t) \equiv w'_t r_t$ . Assuming that the conditional distribution of the unobserved portfolio return  $Y_t$  given  $I_{t-1}$  is continuous, the conditional portfolio VaR at a given confidence level  $1 - \alpha$  given  $I_{t-1}$ ,  $m_\alpha(w_t, \theta_0, I_{t-1})$ , is defined as the  $\alpha^{th}$  quantile of the distribution of  $Y_t|I_{t-1}$  satisfying the equation

$$P(Y_t \leq m_\alpha(w_t, \theta_0, I_{t-1})|I_{t-1}) = \alpha, \text{ almost surely (a.s.), } \alpha \in (0, 1), \forall t \in \mathbb{Z}. \quad (1)$$

for some parameter  $\theta_0$  belonging to  $\Theta$ , with  $\Theta$  a compact set in an Euclidean space  $\mathbb{R}^p$ .

It is important to examine the accuracy of the portfolio VaR model,  $m_\alpha(w_t, \theta_0, I_{t-1})$ , since market risk capital requirements are directly linked to both the estimated level of portfolio VaR as well as the portfolio

VaR model's performance on backtests as laid out by the Basel Committee for Banking Supervision. One of the implications of definition (1) given by Christoffersen (1998) has been taken as the criterion for the out-of-sample evaluation of portfolio VaR forecasts,

$$\{h_{t,\alpha}(\theta_0)\} \text{ are } iid \text{ } Ber(\alpha) \text{ random variables for some } \theta_0, \quad (2)$$

where  $h_{t,\alpha}(\theta_0) := 1(Y_t \leq m_\alpha(w_t, \theta_0, I_{t-1}))$  and  $1(A)$  is the indicator function, i.e.  $1(A) = 1$  if the event  $A$  occurs and 0 otherwise, the variables  $\{h_{t,\alpha}(\theta_0)\}$  are the so-called “hits” or “exceedances”, and  $Ber(\alpha)$  stands for a Bernoulli random variable with parameter  $\alpha$ . The problem of evaluating the accuracy of portfolio VaR forecasts can be reduced to the problem of examining the unconditional coverage and independence properties of the hit sequence  $\{h_{t,\alpha}(\theta_0)\}$ . Based on such statistical properties of the hit sequence, the literature has proposed several tests, such as those in Kupiec (1995), Christoffersen (1998) and Engle and Manganelli (2004).

These testing problems are carried out in an out-of-sample forecast exercise. The forecast environment can be described as follows. Suppose we have a sample  $\{r'_t, z'_t\}_{t=1}^n$  of size  $n \geq 1$  that is used to evaluate portfolio VaR forecasts. For simplicity we only consider one-step-ahead forecasts. As it is known, portfolio choice methods use the estimating and forecasting results from the multivariate dynamic model of asset returns, so we could assume that  $\theta_0$  are the unknown parameters only from the multivariate dynamic model of asset returns without loss of generality, and the portfolio weights  $w_t$  will depend on both  $\theta_0$  and  $I_{t-1}$ , i.e.  $w_t \equiv w_t(\theta_0) = w(\theta_0, I_{t-1})$ . Assume that the first  $R$  observations are used to estimate  $\hat{\theta}_R$  and  $\hat{w}_{R+1}$  in the first forecast, and then we will have  $P = n - R$  predictions to be evaluated. The first VaR forecasts is  $VaR_{R+1,1}(\hat{\theta}_R, \hat{w}_{R+1}) = m_\alpha(w_{R+1}(\hat{\theta}_R), \hat{\theta}_R, I_R)$  and the further forecasts are  $VaR_{t+1,1}(\hat{\theta}_t, \hat{w}_{t+1}) = m_\alpha(w_{t+1}(\hat{\theta}_t), \hat{\theta}_t, I_t)$ ,  $R \leq t \leq n-1$ , where  $\hat{\theta}_t$  and  $\hat{w}_{t+1}$  are estimated using observations  $s = 1, \dots, t$ . For simplicity, we will only focus on studying the unconditional backtesting procedure, but the similar methodology could be applied to the independence tests.

## 2.2 Unconditional backtesting robust to estimation risks

The most popular unconditional backtest proposed by Kupiec (1995) is based on the absolute value of the standardized sample mean

$$K_P \equiv K(P, R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^n (h_{t,\alpha}(\theta_0) - \alpha) = \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [1(Y_t \leq m_\alpha(w_t, \theta_0, I_{t-1})) - \alpha]. \quad (3)$$

Under proper regularity conditions,  $(\alpha(1-\alpha))^{-\frac{1}{2}} K_P$  converges to a standard normal random variable. The standard backtests are implemented under the unrealistic assumptions of  $\theta_0$  and  $w_t$  being known and the portfolio return  $Y_t$  being observable, and using the critical values from the standard normal distribution. In practice, however, both the true parameters  $\theta_0$  and the portfolio weights  $w_t$  are not known and have to be estimated, and hence the portfolio return  $Y_t$  is unobservable. Thus the test statistic becomes

$$S_P \equiv S(P, R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left[ 1 \left( \widehat{Y}_t \leq m_\alpha(\widehat{w}_t, \widehat{\theta}_{t-1}, I_{t-1}) \right) - \alpha \right] \quad (4)$$

where  $\widehat{w}_t = w_t(\widehat{\theta}_{t-1})$  and  $\widehat{Y}_t = Y_t(w_t(\widehat{\theta}_{t-1})) = (w_t(\widehat{\theta}_{t-1}))' r_t$ .

Without considering the impact of estimation risk, the standard backtesting procedure may use wrong critical values, so we must implement backtesting procedures with estimation risk. In order to quantify estimation risk in the present framework, we must consider two estimators,  $\widehat{\theta}_{t-1}$  and  $\widehat{w}_t$ . In different words, the estimation risk in the multivariate VaR model comes from two sources: one is the estimation of unknown parameters in the multivariate dynamic model of asset returns, and the other is the estimation of the unknown portfolio weights.

From the expression of the test statistic in (4), it seems that portfolio VaR can be treated as a univariate parametric VaR model, however, there is an important difference that the portfolio weights  $w_t$  are not observable and must be estimated. As a result, the portfolio return  $Y_t$  is unobservable as well and turns out to be an explicit function of  $\theta_0$ , *i.e.*  $Y_t \equiv Y_t(w_t(\theta_0))$ . This subtle difference has important implications for our testing problem and marks departures from the existing literature. First, it shows that a purely univariate approach to portfolio VaR is in general not possible. Second, this difference makes the results for the univariate case in the literature not applicable to our present framework. As this paper will show



that not only the estimated parameter  $\theta_0$  but also the estimated  $w_t$  add extra terms in the estimation effect on portfolio backtesting. More concretely, we show that both components,  $\widehat{\theta}_{t-1}$  and  $w_t(\widehat{\theta}_{t-1})$ , respectively, introduce asymptotically an extra term in the, still normal, limiting distribution, changing the resulting asymptotic variance of  $S_P$ .

Denote the univariate conditional distributions of  $Y_t(w_t(\theta_0))$  given  $I_{t-1}$  as  $F_{Y_t(w_t(\theta_0))}(\cdot, w_t(\theta_0), \theta_0, I_{t-1})$ , which can be derived from the multivariate conditional distribution of  $r_t$  given  $I_{t-1}$ , and the derivative of  $m_\alpha(w_t, \theta, I_{t-1})$  as  $g_\alpha(w_t, \theta, I_{t-1})$ . We also need some assumptions which are similar to those in Escanciano and Olmo (2009).

**Assumption 1:**  $\{r'_t, z'_t\}_{t \in \mathbb{Z}}$  is strictly stationary and ergodic.

**Assumption 2:** The family of distribution functions  $\{F_x(\cdot), x \in \mathbb{R}^\infty\}$  has Lebesgue densities  $\{f_x(y), x \in \mathbb{R}^\infty\}$  that are uniformly bounded  $\sup_{x \in \mathbb{R}^\infty, y \in \mathbb{R}} |f_x(\cdot)| \leq C$  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 3:** The model  $m_\alpha(w_t, \theta, I_{t-1})$  is continuously differentiable in  $\theta$  and  $w_t$  (*a.s.*), and  $w_t(\theta)$  is also continuously differentiable in  $\theta$  (*a.s.*), such that for its derivative  $g_\alpha(w_t, \theta, I_{t-1})$ ,  $E[\sup_{\theta \in \Theta_0} |g_\alpha(w_t, \theta, I_{t-1})|^2] < C$ , for a neighborhood  $\Theta_0$  of  $\theta_0$ .

**Assumption 4:** The parameter space  $\Theta$  is compact in  $\mathbb{R}^p$ . The true parameter  $\theta_0$  belongs to the interior of  $\Theta$ . The estimator  $\widehat{\theta}_t$  satisfies the asymptotic expansion  $\widehat{\theta}_t - \theta_0 = H(t) + o_P(1)$ , where  $H(t)$  is a  $p \times 1$  vector such that  $H(t) = t^{-1} \sum_{s=1}^t l(r_s, I_{s-1}, \theta_0)$ ,  $R^{-1} \sum_{s=t-R+1}^t l(r_s, I_{s-1}, \theta_0)$  and  $R^{-1} \sum_{s=1}^R l(r_s, I_{s-1}, \theta_0)$  for recursive, rolling and fixed schemes, respectively. We assume that  $E[l(r_t, I_{t-1}, \theta_0) | I_{t-1}] = 0$  *a.s.* and positive definite  $V := E[l(r_t, I_{t-1}, \theta_0) l'(r_t, I_{t-1}, \theta_0)]$  exists. Moreover,  $l(r_t, I_{t-1}, \theta_0)$  is continuous (*a.s.*) in  $\theta$  in  $\Theta_0$  and  $E[\sup_{\theta \in \Theta_0} |l(r_t, I_{t-1}, \theta_0)|^2] \leq C$ , where  $\Theta_0$  is a small neighborhood around  $\theta_0$ .

**Assumption 5:**  $R, P \rightarrow \infty$ , and  $\lim_{n \rightarrow \infty} \frac{P}{R} = \pi$ ,  $0 \leq \pi < \infty$ .

With these assumptions we are ready to establish the first important result of this paper.

THEOREM 1: *Under Assumption A1-A5,*

$$\begin{aligned}
S_P &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [h_{t,\alpha}(\theta_0) - F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t(\theta_0), \theta_0, I_{t-1}))] \\
&\quad + \underbrace{E\left[\frac{\partial F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t(\theta_0), \theta_0, I_{t-1}))}{\partial \theta'_0}\right] \frac{1}{\sqrt{P}} \sum_{t=R+1}^n H(t-1)}_{\text{Estimation Risk}} \\
&\quad + \underbrace{\frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t(\theta_0), \theta_0, I_{t-1})) - \alpha]}_{\text{Model Risk}} + o_P(1)
\end{aligned}$$

where the score component in the estimation risk, say  $A = E\left[\frac{\partial F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t(\theta_0), \theta_0, I_{t-1}))}{\partial \theta'_0}\right]$ , can be partitioned into two components from the chain rule,

$$A = E \left[ \underbrace{\frac{\partial F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t(\theta_0), \theta, I_{t-1}))}{\partial \theta'}}_{\text{Due to estimation of dynamics}} + \underbrace{\frac{\partial F_{Y_t(w_t(\theta))}(m_\alpha(w_t(\theta), \theta_0, I_{t-1}))}{\partial \theta'}}_{\text{Due to estimation of portfolio weights}} \right]_{\theta=\theta_0}.$$

Notice that in the first component of  $A$  we only consider variation in  $\theta$  with  $w_t$  held fixed. As the score component  $A$  will be evaluated at  $\theta = \theta_0$ , in  $w_t$  we are entitled to replace  $\theta$  with  $\theta_0$  before we differentiate, therefore the first component in  $A$  is just due to estimation of dynamics. In an analogous manner, the second component of  $A$  is obtained by letting  $w_t$  vary and holding all other  $\theta$  outside of  $w_t$  as fixed, which is only due to estimation of portfolio weights.

Theorem 1 quantifies both estimation risk and model risk in the unconditional backtests. In this paper we assume the multivariate VaR model is correctly specified, i.e.  $F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t, \theta_0, I_{t-1})) = \alpha$ , then model risk vanishes, but we still have the estimation risk to deal with. Notice that there are two sources of estimation risk under the multivariate VaR model, one from estimating parameters in the multivariate model for asset returns, one from estimating the portfolio weights. Without accounting for any of those components, we may make wrong inference in the unconditional backtesting procedures. In addition, the theory can be applied to either optimal or suboptimal portfolios, as long as portfolios are estimated. However, the magnitude of the estimation risk due to the estimation of portfolio weights will depend on the property of the objective function in the portfolio optimization problem. It has been known in statistical literature

that, when the objective function is linear or symmetric, loss from estimation error tends to be small if the estimates are unbiased, so the magnitude of the estimation risk from estimating portfolio weights is not expected to be big enough to have influential effect on backtesting result. But when the objective function is highly non-linear and asymmetric, the estimation risk due to the estimation of portfolio weights tends to be moderate, see Im, Lim and Choi (2007). This will be examined in our simulation exercise.

COROLLARY 1: *Under Assumptions A1-A5 and (1),*

$$S_P \xrightarrow{d} N(0, \sigma_u^2)$$

where

$$\sigma_u^2 = \alpha(1 - \alpha) + 2\lambda_{hl}A\rho + \lambda_{ll}AVA'$$

with  $\rho = E[(h_{t,\alpha}(\theta_0) - \alpha)l(r_t, I_{t-1}, \theta_0)]$ , and where

<i>Forecast Scheme</i>	$\lambda_{hl}$	$\lambda_{ll}$
<i>recursive scheme</i>	$1 - \pi^{-1} \ln(1 + \pi)$	$2[1 - \pi^{-1} \ln(1 + \pi)]$
<i>rolling scheme with <math>\pi \leq 1</math></i>	$\pi/2$	$\pi - \pi^2/3$
<i>rolling scheme with <math>1 &lt; \pi &lt; \infty</math></i>	$1 - (2\pi)^{-1}$	$1 - (3\pi)^{-1}$
<i>fixed scheme</i>	$0$	$\pi$

Corollary 1 presents the asymptotic distribution of  $S_P$  with estimation risk, which suggests a way to carry out valid inference for unconditional backtests free of estimation risk. However, it is a difficult task to implement such backtesting procedure. First, the analytical formula for  $\sigma_u^2$  is too complex to estimate straightforwardly by conventional methods; Secondly, there exist computational complexities in deriving the score component  $A$  in the estimation risk in the multivariate VaR model.

From Corollary 1, it is obvious that the standard unconditional backtesting procedure is not reliable unless the asymptotic variance  $\sigma_u^2$  goes to  $\alpha(1 - \alpha)$ , the asymptotic variance of  $K_P$ . Fortunately, Corollary 1 also suggests a solution to this problem. According to the formulas of the coefficients  $\lambda_{hl}$  and  $\lambda_{ll}$ , as the parameter  $\pi = P/R$  goes to zero, both coefficients go to zero under all the three forecast schemes, which implies  $\sigma_u^2$  goes to  $\alpha(1 - \alpha)$  as the ratio of the out-of-sample size to the in-sample size goes to zero. In other

words, the effect of estimation risk on the standard unconditional backtesting tends to vanish as the ratio of the out-of-sample size relative to the in-sample size goes to zero. Therefore, to overcome the effect of estimation risk on the unconditional backtesting, we recommend financial institutions to use small ratios of the out-of-sample size to the in-sample size ratio, such that valid inference could be carried out. This will be confirmed in the simulation exercise.

### **3 An application under particular distributional assumptions**

The theoretical findings presented above are very general, since they do not require any particular distributional assumptions for asset returns and any particular method of portfolio choice, and they do not require the constructed portfolio to be optimal as well. In this section, we will consider several particular settings. In order to illustrate our theoretical findings, we also carry out a series of Monte Carlo simulation experiments by using the models from the simplest to the most realistic, since all the above theoretical findings are asymptotic, and we need to find out how they behave in finite samples.

#### **3.1 Multivariate dynamic model for asset returns**

We want to set up a multivariate parametric dynamic model with specified innovations satisfying the following properties: it not only takes into account the empirical features of volatility clustering and time-varying correlations but also the stylized facts of positive excess kurtosis and negative skewness from financial time series; Additionally, the specified multivariate distribution of asset returns is closed under linear transformations such that the distribution of returns of any portfolio whose components are modeled in such a framework is still in the same class, which will have important implication for the use of the model in applications such as portfolio allocation and portfolio VaR calculation. The most popular proposal for multivariate volatility modelling belongs to the family of multivariate generalized autoregressive conditional heteroscedasticity models (MGARCH). One type is to model the conditional covariance matrix directly, which includes the VEC model of Bollerslev, Engle and Wooldridge (1988) and BEKK model defined in Engle and Kroner (1995). Another type is to model the conditional variances and correlations instead of directly modelling the conditional covariance matrix, and the simplest one is the Constant Conditional Correlation (CCC)-GARCH

model of Bollerslev (1990), which is attractive to the practitioner due to its interpretable parameters and easy estimation. A major problem with most MGARCH model is that the number of parameters tends to explode with the dimension of the model. Therefore, factor models are motivated for parsimony, which either assume that asset returns are generated by underlying conditionally heteroscedastic factors, see Diebold and Nerlove (1989) and King, et al. (1994), or assume there is a time varying factor structure in the covariance matrix of returns, see Engle, Ng and Rothschild (1990). Both specifications are very appealing in finance because of their important implications for both the Arbitrage Pricing Theory and the Capital Asset Pricing Model. As for the distributional assumption of the innovations, there are several multivariate distributions in the literature that could be used in a multivariate dynamic model, for example, Bauwens and Laurent (2004) applies the multivariate skew-Student density to a dynamic conditional correlation (DCC) model, Mencia and Sentana (2005) analyses the general hyperbolic (GH) distribution in the multivariate dynamic regression model and Cajigas and Urga (2006) uses asymmetric multivariate Laplace (AML) innovations in the DCC model.

For illustrative purpose, this paper considers a multivariate conditionally heteroscedastic single factor model, since the single factor models just have exactly the same pricing ability as the multiple factor models, see Cochrane (2001). The model takes the form

$$r_t = \mu + cf_t + v_t$$

where  $f_t$  is the common latent factor with conditional mean  $E[f_t|I_{t-1}] = 0$  and time-varying conditional variance  $V(f_t|I_{t-1}) = \lambda_t$ ,  $v_t' = (v_{1t}, v_{2t}, \dots, v_{dt})$  is the  $d$ -dimensional vector of idiosyncratic risks satisfying  $E[v_t|I_{t-1}] = 0$  and  $V(v_t|I_{t-1}) = \Gamma = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_d)$  with nonnegative diagonal elements,  $v_t$  is assume to be conditionally orthogonal to  $f_t$ ,  $\mu$  is the  $d$ -dimensional vector of mean returns, and  $c' = (c_1, c_2, \dots, c_d)$  is the  $d$ -dimensional vector of factor loadings. These assumptions imply that the distribution of  $r_t$  conditional on  $I_{t-1}$  has mean  $\mu$ , and time varying covariance matrix  $\Sigma_t = cc' \lambda_t + \Gamma$ . Therefore, the data generating process of  $r_t$  can be expressed as  $r_t = \mu + \Sigma_t^{1/2} \varepsilon_t$ . We will assume that the conditional distribution of the innovations  $\varepsilon_t$  on  $I_{t-1}$  is the standardized  $d$ -dimensional GH distribution with parameters  $(\eta, \psi, b)$ , where  $\eta$  and  $\psi$  allow for flexible tail modeling and the vector  $b' = (b_1, b_2, \dots, b_d)$  introduces skewness, see Appendix

C for the detailed parameterization of this distribution. In particular, we suppose the conditional variances of the common factor follows the generalized quadratic autoregressive conditionally heteroscedastic model, GQARCH(1,1), given by

$$\lambda_t = \beta_0 + \omega f_{t-1} + \beta_1 f_{t-1}^2 + \beta_2 \lambda_{t-1},$$

where  $\beta_0, \beta_1, \beta_2 > 0$  and  $\omega^2 \leq 4\beta_1\beta_0$  to ensure the conditional variances to be positive. The GQARCH model originally proposed by Sentana (1995) has the advantage of capturing both an asymmetric effect on the conditional variance and higher excess kurtosis, compared with the standard GARCH model. Notice  $\lambda_t$  depends on past values of  $f_{t-1}$  and  $f_{t-1}^2$ , whose true values do not necessarily belong to the available information set  $I_{t-1}$ , but can be evaluated on the observables via the Kalman filter, see Harvey et al. (1992). Taking the common factor as the state, it is easy to derive the updating equations:  $f_{t|t} = E[f_t|I_t] = \omega_{t-1|t-1}c'\Gamma^{-1}r_{t-1}$  and  $\omega_{t|t} = V[f_t|I_t] = (\lambda_{t-1} + c'\Gamma^{-1}c)$ , and then the QGARCH model will be modified as

$$\lambda_t = \beta_0 + \omega f_{t-1|t-1} + \beta_1(f_{t-1|t-1}^2 + \omega_{t-1|t-1}) + \beta_2 \lambda_{t-1},$$

where  $f_{t-1|t-1}^2 + \omega_{t-1|t-1} = E[f_{t-1}^2|I_{t-1}]$  and  $\omega_{t-1|t-1}$  plays the role of correction error in the factor estimates. Such specifications are appealing in the following aspects: first, the factor model provides a relatively parsimonious representation, which has much less number of parameters involved than the other model specifications such as DCC and BEKK, which makes it feasible in large systems; Second, it is able to capture all the stylized facts of financial time series; Third, a positive (semi-)definite conditional covariance matrix for  $r_t$  is automatically guaranteed once we ensure that the conditional variances of the factors are non-negative. But due to the complexity of the GH distribution, the estimation of this model is still computationally demanding.

### 3.2 Portfolio selection

The estimating and forecasting results from the above model will be as the inputs to the portfolio selection problem. Under our model setup, asset returns jointly follow a GH distribution, which can be expressed as a location-scale mixture of normals and the skewness of asset returns is also considered. Therefore, we follow the MVS analysis of Mencia and Sentata (2009) who show that the distribution of any portfolio whose

components jointly follow a location-scale mixture of normals will be uniquely characterized by its mean, variance and skewness, and also derives the closed-form solution for the optimal portfolio weights, which can be expressed as a linear combination of the *skewness-variance* efficient portfolio  $\bar{b}$  and the *mean-variance* efficient portfolio  $\Sigma_t^{-1}\iota$ , where  $\iota$  is a  $(d \times 1)$  vector of ones. To save space, we provide the formula in Appendix B, see Mencia and Sentana (2009) for the details.

It is worthwhile mentioning that there are few papers in the literature considering the selection of portfolio weights in the forecasts of portfolio VaRs and no papers even considering the impact of estimation risk from estimating portfolio weights on backtesting portfolio VaR. One work that evaluates portfolio VaR with the estimated optimal portfolio weights is Rombouts and Verbeek (2004) which determines portfolio weights taking into account a VaR constraint. To the best of our knowledge, our paper is the first work to account for the estimation of portfolio weights and its influences on backtesting portfolio VaR.

For the simulation, we assume that  $\mu = 0$ . Hence, the portfolio VaR would take the form

$$m_\alpha(w_t, \theta_0, I_{t-1}) = \sqrt{w_t' \Sigma_t w_t} G^{-1}(\alpha),$$

where  $G^{-1}(\alpha)$  is the  $\alpha$ -th quantile of the univariate standardized GH distribution of  $\varepsilon_t^* = w_t' \Sigma_t^{\frac{1}{2}} \varepsilon_t / \sqrt{(w_t \Sigma_t^{\frac{1}{2}})' \Sigma_t^{\frac{1}{2}} w_t}$ .

### 3.3 Backtesting portfolio VaR

Theorem 1 allows us to quantify estimation risk for the unconditional backtests such that we could carry out valid inferences. For simplicity, we only consider the fixed forecasting scheme. The estimation risk term for the unconditional test with fixed forecasting scheme in the current setting takes the form

$$ER_u = E \left[ \underbrace{\frac{\partial F_{Y_t(w_t(\theta_0))}(m_\alpha(w_t(\theta_0), \theta, I_{t-1}))}{\partial \theta'}}_{\text{Due to estimation of dynamics}} + \underbrace{\frac{\partial F_{Y_t(w_t(\theta))}(m_\alpha(w_t(\theta), \theta_0, I_{t-1}))}{\partial \theta'}}_{\text{Due to estimation of portfolio weights}} \right] \Big|_{\theta=\theta_0} \sqrt{\pi} \sqrt{R} (\hat{\theta}_R - \theta_0)$$

### 3.4 Simulation exercise

The purpose of this section is to illustrate our theoretical findings. We show that the standard backtesting procedure without considering the effect of the estimation risk could be misleading, and the estimate of portfolio weights and the choice of different in-sample size to out-of-sample size ratio have important implications in backtesting. For illustrative purpose, we implement the same set of experiments by using three different models. The simplest one is a bivariate constant location-scale model with standardized Student-t innovations, which can be expressed as  $r_t = \Sigma_0^{\frac{1}{2}} \varepsilon_t$ , where  $\varepsilon_t$  follow a bivariate Student-t distribution with degrees of freedom  $\nu$ . The second one is a bivariate first-order BEKK model, which takes the form

$$\begin{aligned} r_t &= \Sigma_t^{\frac{1}{2}} \varepsilon_t \\ \Sigma_t &= C_0' C_0 + C_1' r_{t-1} r_{t-1}' C_1 + D' \Sigma_{t-1} D, \end{aligned}$$

where  $C_0$ ,  $C_1$  and  $D$  are in  $\mathbb{R}^{2 \times 2}$ ,  $C_0$  is an upper triangular matrix and  $\varepsilon_t$  follow a bivariate Student-t distribution with degrees of freedom  $\nu$  conditional on information set  $I_{t-1}$ . The third one, as the most realistic and complicated one, is a trivariate model of the conditionally heteroscedastic single factor model as described in the previous section. In allocating assets, we apply the MV method to the two simpler models and the MVS method to the third one.

As for the parameter values, we have chosen for the BEKK model that

$$C_0 = 10^{-3} \begin{pmatrix} 1.15 & .31 \\ 0 & .076 \end{pmatrix}, C_1 = \begin{pmatrix} .282 & -.050 \\ -.057 & .293 \end{pmatrix}, D = \begin{pmatrix} .939 & .028 \\ .025 & .939 \end{pmatrix} \text{ and } \nu = 5,$$

which are taken from the estimation results in Hardle, Kleinow and Stahl (2002), and for the simplest model that  $\Sigma_0$  is set to be the unconditional covariance matrix of the specified BEKK model. For the third model, we have set the model parameters  $c' = (1, 1, 1)$ ,  $\Gamma = \text{diag}(1, 1, 1)$ ,  $\beta_1 = 0.1$ ,  $\beta_2 = 0.6$ ,  $\omega = -0.2771$  and  $\beta_0 = 1 - \beta_1 - \beta_2$  and the distribution parameters  $\eta = 0.5$ ,  $\psi = 0.1$  and  $b' = (-0.1, -0.1, -0.1)$ .

We implement a series of simulation experiments based on the uncorrected standard unconditional backtesting test statistics  $S_P$ . For the purpose of comparison, we design four cases. Case I is the experiment with



both the true parameter values and the true portfolio weights, case II is the one with both the estimated parameters and the estimated portfolio weights, case III is the one with the estimated parameters but the true portfolio weights, and case IV is the one with the true parameter values but the estimated portfolio weights. In each case, we also consider four different pairs of the in-sample and out-of-sample size and three different levels of the nominal sizes. We calculate  $S_P$  for 1000 Monte Carlo simulations in each case, and then compare the size performance.

We carry out the designed simulation experiments using the three models with three different forecasting schemes, which are the fixed scheme, the recursive scheme and the rolling scheme, and for  $\alpha = 0.05$  and  $\alpha = 0.01$ , respectively. For the sake of space, we just report the results with the fixed forecasting scheme in Table 1, Table 2 and Table 3. Other results are available upon request.

There are five main conclusions from our simulation results. First, the standard unconditional backtesting only performs well when the true parameter values are known. As it is shown in case I, the empirical sizes are closer to the nominal sizes across all the experiments. Unsurprisingly, the estimation risk does have significant effect on the backtesting results when the estimated parameters are used. As the results of case II show, the empirical sizes are far away from the nominal sizes across all the experiments, especially for  $\pi = 1$ , the case usually being used in practice. For example, at the nominal size of 10%, the empirical size of backtesting the BEKK model forecasted portfolio VaR at  $\alpha = 0.05$  is as high as 42%. There exists a huge size distortion when the true parameter values are not known and have to be estimated, which will be the case in practice. Thus the standard backtesting procedure without considering the effect of the estimation risk could be misleading.

Second, as predicted by the theory, the empirical sizes of the unconditional test are closer to the nominal sizes as the ratio of the out-of-sample to in-sample size,  $\pi$ , goes to zero (See the results of Case II as  $\pi$  goes from 1 to 0.125. ). As the ratio of the out-of-sample to in-sample size gets smaller, the results improve at all levels of nominal sizes. In other words, the effect of estimation risk on backtesting portfolio VaR tends to vanish as the ratio of the out-of-sample size relative to the in-sample size goes to zero. Therefore, we recommend to use small values of the out-of-sample relative to the in-sample size to financial institutions, in order to make valid inference for unconditional backtests.

Third, estimation risk tends to be more important when the number of parameters gets larger. Among

the three models used, the BEKK model has the most parameters and the constant location-scale model has the least. Comparing the results of case II across the tables, we found the empirical sizes of the BEKK model are the highest and those of the constant location-scale model are the lowest. This result somewhat confirms the conjecture by Christoffersen and Goncalves (2005) that the estimation risk issue is probably even more important in multivariate modeling where the number of parameters is large.

Fourth, the only difference between case I and case IV and that between case II and case III is whether the portfolio weights are estimated. Comparing the results, the backtesting results in Table 1 and Table 2 are not very sensitive to whether the portfolio weights are estimated, but the results in Table 3 are, just as expected. This is due to the symmetric objective function used in the MV analysis and asymmetric one used in the MSV analysis.

Fifth, the asymptotic distribution of  $S_P$  does not provide an accurate approximation for small VaR values such as  $\alpha = 0.01$ . In such a case we need a different asymptotic theory based on  $\alpha \rightarrow 0$ , which is beyond the scope of this paper.

## 4 Application

As we have seen above, the findings in this paper suggest a simple but effective way to overcome the effect of estimation risk on unconditional backtests for financial institutions, such that they can make more reliable inference in backtesting portfolio VaR, which is to implement the standard unconditional backtesting procedure by using a small value of the out-of-sample size relative to the in-sample size and taking VaR level to be 5%.

As an example, we apply the proposed method to a portfolio of three US stocks of the Dow Jones Index: the Alcoa stock (AA), the MacDonald stock (MCD) and Merck stock (MRK). The data originally used in Giot and Laurent (2003) range from January 1990 to May 2002 (3000 daily observations). Daily returns are constructed as the first difference of logarithmic prices multiplied by 100. The main features of all returns series include fat tails, skewness, the excess of positive kurtosis and volatility clustering. Given these characteristics, we fit a trivariate conditionally heteroscedastic single factor model and choose the MSV method to obtain the portfolio weights as specified previously in the paper. We take the last 250 observations

as the out-of-sample period, i.e.  $P=250$ , and the in-sample period of  $R=2750$  observations. For the purpose of comparison, we also choose several different in-sample sizes:  $R=250, 500, 1000$  and  $2000$ , and implement the backtesting procedure at 1% VaR as well. The in-sample period is then used to estimate the model, calculate the portfolio weights, and make forecasts over the out-of-sample period. For simplicity, we only consider the fixed forecast scheme.

Table 4 reports the backtesting results, which include the number of violations, the unconditional backtesting test statistic ( $S_D$ ) and the multiplication factor<sup>2</sup> ( $mf_t$ ) for risk-based capital requirements under the VaR levels, 5% and 1%, respectively. The results show that the portfolio VaR model is rejected at 5% significance level for  $\alpha = 0.05$  and  $R = 2750$ , and for  $\alpha = 0.01$  and  $R = 2000$  and  $2750$ , but not for the other cases with smaller values of the in-sample size. Since the out-of-sample period is fixed, the larger the in-sample size is, the smaller the ratio,  $\pi$ , is. Based on our theory, the effect of estimation risk on the unconditional backtesting results declines with the value of  $\pi$ , therefore, the results from using a larger in-sample size are more reliable. We find that the backtesting results are substantially different between using the large in-sample size and the small one. With the small value of  $R$ , there are less number of violations of the model forecasted portfolio VaR, which implies the model provides a sufficient coverage of trading risk, however, with large value of  $R$ , the number of violations turns out to be larger, which implies the existence of excessive trading risk not covered by the portfolio VaR model. For example, for  $\pi = 1$ , the number of violations of the model forecasted 5% VaRs is just 9 out of 250, but for  $\pi = 0.0909$ , we have 20 violations of the model forecasted 5% VaRs out of 250. Most importantly, as laid out by the Basle Committee on Banking Supervision, the forecasted portfolio VaR and the backtesting result are directly related to the determination of risk-based capital requirement, in which the multiplication factor plays a role of the penalty for the backtest. The multiplication factor varies with backtesting results. As is shown in Table 4, the multiplication factor is 3.3 for  $\pi = 0.0909$ , which is slightly larger than those of the other cases with relatively larger  $\pi$ . This is because a forecasted VaR that is violated more frequently results in a larger multiplication factor and accordingly a larger risk based capital requirement. Therefore, if financial institutions implement backtests using a large value of  $\pi$ , estimation risk will have an influential effect on their backtesting results, so they may make the wrong inference, compute the inappropriate multiplication factor and accordingly determine the insufficient risk capital requirement. As this paper suggests, we should backtest portfolio VaR using a small value of  $\pi$

in order to overcome the effect of estimation risk on backtests. As for the current application, we shall use multiplication factor equal to at least 3.3 in determination of risk capital requirement.

## 5 Conclusion

This paper proposes the general unconditional backtesting procedure robust to estimation risk for portfolio VaR with consideration of portfolio weights estimation. It extends the theory of quantifying estimation risk in backtesting procedures from the univariate case to a multivariate case, which is intrinsically the framework for backtesting portfolio VaR. We also apply the general theory to a particular setting in which asset returns are modeled by a multivariate location-scale dynamic model with standardized GH innovations and use the MVS analysis to obtain the portfolio allocation. The simulation exercise in the paper supports the theoretical findings. In order to overcome the effect of estimation risk on backtesting portfolio VaR, we suggest a simple, practical but effective way, which is to implement the standard unconditional backtesting procedure by using a small ratio of the out-of-sample to the in-sample size and the 5% VaR level. Our proposed method is of great importance in practice, and helps practitioners to set aside more accurate risk capital requirement. The findings in this paper suggest that inferences on portfolios are subject to estimation risk. Although we only study the impact of estimation risk on backtesting portfolio VaR, our methodology can be applied to other out-of-sample problems involving estimated portfolios.

## References

- Basle Committee on Banking Supervision. 1996a. Amendment to the Capital Accord to Incorporate Market Risks. Bank for International Settlements.
- Basle Committee on Banking Supervision. 1996b. Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements. Bank for International Settlements.
- Bauwens, Luc and Sebastien Laurent. 2004. "A new class of multivariate skew density, with application to GARCH models." working paper.
- Bauwens, Luc, Sebastien Laurent and Jeroen V.K. Rombouts. 2006. Multivariate GARCH models: A survey. *Journal of Applied Economics*, 21 79-109.
- Barberis, Nicholas. 2000. Investing for the long run when returns are predictable, *Journal of Finance*, 55, 225-264.
- Bollerslev, T., R.F. Engle and J. M. Wooldridge. 1988. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96, 116-131.
- Cajigas, Juan-Pablo and Giovanni Urga. 2006. Dynamic Conditional Correlation Models with Asymmetric Multivariate Laplace Innovations, working paper.
- Christoffersen, P. and S. Gonçves. 2005. Estimation risk in financial risk management, *Journal of Risk*, 7, 1-28.
- Cochrane, John H. 2001. *Asset Pricing*, Princeton University Press.
- Delgado, M.A. and Escanciano, J.C. 2006. Nonparametric Tests for Conditional Symmetry in Dynamic Models, *Journal of Econometrics*, 141, 652-682.
- Diebold, Francis X. and Nerlove Marc. 1989. The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor Arch Model, *Journal of Applied Econometrics*, Vol. 4, No. 1, pp. 1-21.

- Down, K. 2000. Assessing VaR accuracy. *Derivatives Quarterly*, 6, 61-63.
- Engle, R.F. and K.F. Kroner. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122-150.
- Engle, R.F., V. K. Ng and M. Rothschild. 1990. Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics*, 45, 213-238.
- Escanciano, J. Carlos and Jose Olmo. 2009. Backtesting Value-at-Risk with Estimation Risk. Forthcoming in *Journal of Business and Economic Statistics*.
- Ferreira, M.A. and J.A. Lopez. 2005. Evaluating Interest Rate Covariance Models Within a Value-at-Risk Framework, *Journal of Financial Econometrics*, 3, 126-168.
- Hall, P. and Heyde, C. 1980. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics, Academic Press. New York.
- Hardle, W., Kleinow, T. and Stahl, G. 2002. *Applied Quantitative Finance: Theory and Computational Tool*.
- Harvey, Andrew, Ruiz, Esther and Sentana Enrique. 1992. Unobserved component time series models with ARCH disturbances. *Journal of Econometrics* 52(1992), 129-157, North-Holland.
- Im, Ji Jung, Hyun Soo Lim and Sung Sub Choi. 2007. Portfolio Selection under Parameter Uncertainty Using a Predictive Distribution. *Annals of Economics and Finance*, 8-2, 305-312.
- Gao, Feng and Fengming Song. 2008. Estimation risk in GARCH VaR and ES estimations. *Econometric Theory*, 24, 2008, 1404-1424.
- Giot, Pierre and Laurent, Sebastien. 2003. Value-at-Risk for long and short trading positions. *Journal of Applied Economics*, 18: 641-664.
- Gourieroux, Christian and Joann Jasiak. 2002. Value-at-Risk, forthcoming *Handbook of Financial Econometrics*. ed. Ait-Sahalia, Hansen.

- Jennrich, R.I. 1969. Asymptotic Properties of Nonlinear Least Square Estimators. *Annals of Mathematical Statistics*, 40, 633-643.
- Jorion, P. 1996. Risk: Measuring the risk in value-at-risk. *Financial Analysts Journal*, 52, 47-56.
- Jorion, P. 2001. *Value-at-Risk: the New Benchmark for Controlling Market Risk*. McGraw-Hill, New York.
- Kandel, shmuel, and Robert Stambaugh. 1996. On the predictability of stock returns: An asset allocation perspective. *Journal of Finance*, 51,385-424.
- Klein, R., and V. Bawa. 1976. The Effect of Estimation Risk on Optimal Portfolio Choice, *Journal of financial Economics*, 3:215-31.
- McCracken. 2000. Robust Out-of-sample Inference, *Journal of Econometrics*, 99,5,195-223.
- McNeil, Alexander J., Rudiger Frey and Paul Embrechts. 2005. *Quantitative Risk Management*, Princeton University Press, Princeton and Oxford.
- Mencia, F.Javier and Enrique Sentana. 2005. "Estimation and Testing of Dynamic models with generalized Hyperbolic innovations." Working paper.
- Mencia, F.Javier and Enrique Sentana. 2009. Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Journal of Econometrics*. 153(2009), 105-121.
- Rombouts, Jeroen V.K. and Marno Verbeek. 2004. "Evaluating Portfolio Value-at-Risk using Semi-Parametric GARCH Models." Working paper.
- Sentana, enrique. 1995. Quadratic ARCH models, *The Review of Economic Studies*, Vol. 62, No. 4, pp. 639-661.
- Silvennoinen, Annastiina and Terasvirta, Timo. 2008. Multivariate GARCH models. To appear in T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch, eds. *Handbook of Financial Time Series*, New York: Springer.

## Notes

<sup>1</sup>Giot and Laurant (2003) uses a hypothetical portfolio of three US stocks in order to fully test their method. Ferreira and Lopez (2005) considers an equally weighted portfolio of short-term fixed income positions in the U.S. dollar, German deutschemark, and Japanese yen.

<sup>2</sup>The multiplication factors determined by classifying the number of VaR violations in the previous 250 days,  $N$ , into three zones.

For a true coverage level of 99%,

$$mf_t = \begin{cases} 3.0, & \text{if } N \leq 4, \text{ the green zone.} \\ 3 + 0.2(N - 4), & \text{if } 5 \leq N \leq 9, \text{ the yellow zone.} \\ 4.0, & \text{if } N \geq 10, \text{ the red zone.} \end{cases}$$

For a true coverage of 95%,

$$mf_t = \begin{cases} 3.0, & \text{if } N \leq 17, \text{ the green zone.} \\ 3 + 0.1(N - 17), & \text{if } 18 \leq N \leq 27, \text{ the yellow zone.} \\ 4.0, & \text{if } N \geq 28, \text{ the red zone.} \end{cases}$$

According to Basle Committee on Banking Supervision (1996b), the yellow zone begins at the point such that the probability of obtaining that number or fewer violations equals or exceeds 95%, and the red zone begins at the point such that the probability of obtaining that number or fewer violations equal or exceeds 99.99%.



Table 1: Empirical test sizes in % at  $\alpha = 0.05$  (The constant location-scale model)

		$\alpha = 0.05$				$\alpha = 0.01$			
$\pi$	<i>Nominal size</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>	<i>Case IV</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>	<i>Case IV</i>
1	10%	9.0	19.0	17.0	8.9	9.1	42.0	39.2	9.6
	5%	4.7	12.8	11.2	5.1	5.0	35.8	33.4	5.2
	1%	1.4	5.3	4.3	1.3	1.2	25.8	22.2	1.5
0.5	10%	8.6	12.5	12.6	8.5	8.8	30.6	30.6	8.3
	5%	4.9	7.9	7.2	4.9	4.3	24.3	24.3	4.4
	1%	1.2	2.0	2.4	1.3	0.6	16.9	15.5	0.6
0.25	10%	9.0	10.9	10.1	8.9	9.3	19.6	19.3	9.2
	5%	4.9	6.9	6.4	4.7	4.7	12.6	12.4	4.8
	1%	1.6	2.2	2.1	1.3	0.8	6.8	5.8	0.7
0.125	10%	8.8	9.0	9.0	8.5	9.3	14.3	13.6	9.0
	5%	4.4	4.4	4.2	4.2	4.4	7.7	7.7	4.2
	1%	1.0	1.7	1.5	1.2	1.0	3.1	2.8	0.9

Table 2: Empirical test sizes in % at the VaR level  $\alpha = 0.05$  (BEKK Model)

		$\alpha = 0.05$				$\alpha = 0.01$			
$\pi$	<i>Nominal size</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>	<i>Case IV</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>	<i>Case IV</i>
1	10%	9.1	42.0	39.2	9.6	2.3	33.0	26.6	2.0
	5%	5.0	35.8	33.4	5.2	2.3	33.0	26.6	2.0
	1%	1.2	25.8	22.2	1.5	0.4	26.5	19.9	0.3
0.5	10%	8.8	30.6	30.6	8.3	2.3	22.9	20.9	2.5
	5%	4.3	24.3	24.3	4.4	2.3	22.9	20.9	2.5
	1%	0.6	16.9	15.5	0.6	0.6	16.8	14.3	0.6
0.25	10%	9.3	19.6	19.3	9.2	1.6	17.5	16.1	1.7
	5%	4.7	12.6	12.4	4.8	1.6	17.5	16.1	1.7
	1%	0.8	6.8	5.8	0.7	0.6	9.1	8.0	0.5
0.125	10%	9.3	14.3	13.6	9.0	1.7	12.6	11.9	1.7
	5%	4.4	7.7	7.7	4.2	1.7	12.6	11.9	1.7
	1%	1.0	3.1	2.8	0.9	0.5	6.2	6.2	0.5

Table 3: Empirical test sizes in % at the VaR level  $\alpha = 0.05$  (The factor model)

		$\alpha = 0.05$				$\alpha = 0.01$			
$\pi$	<i>Nominal size</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>	<i>Case IV</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>	<i>Case IV</i>
1	10%	9.9	20.6	15.3	15.3	5.1	6.1	13.0	5.1
	5%	5.9	16.2	10.3	13.1	5.1	6.1	13.0	5.1
	1%	1.2	11.4	4.2	9.7	2.2	2.6	7.1	2.3
0.5	10%	8.5	17.8	12.5	15.2	5.7	4.4	9.0	5.4
	5%	4.6	12.8	7.7	10.7	5.7	4.4	9.0	5.4
	1%	1.0	8.7	3.4	7.9	1.6	2.6	5.0	2.2
0.25	10%	7.9	13.7	9.1	11.8	4.2	4.5	6.1	4.5
	5%	4.3	8.5	5.4	7.4	4.2	4.5	6.1	4.5
	1%	0.9	5.3	1.7	5.0	1.2	1.6	2.8	1.5
0.125	10%	9.2	9.7	9.5	9.4	5.6	5.4	8.3	5.4
	5%	4.5	5.9	6.1	4.9	5.5	5.3	8.2	5.4
	1%	1.0	2.6	1.9	2.2	1.9	2.3	3.2	2.2

Table 4: Application Results

		$\alpha = 0.05$			$\alpha = 0.01$		
$R$	$\pi$	# of violation	$S_P$	$mf_t$	# of violation	$S_P$	$mf_t$
2750	0.0909	20	2.1764	3.3	6	2.2247	3.4
2000	0.1250	17	1.3059	3.0	6	2.2247	3.4
1000	0.2500	12	0.1451	3.0	3	0.3178	3.0
500	0.5000	10	0.7255	3.0	3	0.3178	3.0
250	1.0000	9	0.4353	3.0	4	0.9535	3.0

## Appendix A: Mathematical Proofs

First of all, we show how to get the conditional distribution of  $Y_t$  given  $I_{t-1}$  from the linear transformation  $Y_t = w_t' r_t$  and the multivariate conditional distribution of  $r_t$  given  $I_{t-1}$ . Note that  $w_t$  is treated as a constant at time  $t$ . Construct a one-to-one mapping between  $r_t$  and a constructed vector  $Z_t$  with the first element being  $Y_t = w_t' r_t$ ,

$$Z_t = \begin{bmatrix} Y_t \\ r_{2t} \\ \vdots \\ r_{dt} \end{bmatrix} = \begin{bmatrix} w_t' r_t \\ r_{2t} \\ \vdots \\ r_{dt} \end{bmatrix} = \underbrace{\begin{bmatrix} w_{1t} & w_{2t} & \cdots & w_{dt} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{J(\theta_0)} \begin{bmatrix} r_{1t} \\ r_{2t} \\ \vdots \\ r_{dt} \end{bmatrix} = J(\theta_0) r_t,$$

where  $J(\theta_0)$  is the constructed positive definite  $d \times d$  matrix known at time  $t$ , so  $r_t = J^{-1}(\theta_0) Z_t$ , where

$$J^{-1}(\theta_0) = \begin{bmatrix} \frac{1}{w_{1t}} & -\frac{w_{2t}}{w_{1t}} & \cdots & -\frac{w_{dt}}{w_{1t}} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

and  $w_{it} \neq 0$ ,  $i = 1, \dots, d$ . Since the conditional distribution of  $r_t$  given  $I_{t-1}$  is specified as  $r_t | I_{t-1} \sim F_{r_t}(\cdot, \theta_0, I_{t-1})$ , then the multivariate conditional distribution of  $Z_t | I_{t-1}$  can be obtained as follows

$$\Pr\{Z_t \leq z | I_{t-1}\} = \Pr\{J(\theta_0) r_t \leq z | I_{t-1}\} = \Pr\{r_t \leq J^{-1}(\theta_0) z | I_{t-1}\} = F_{r_t}(J^{-1}(\theta_0) z, \theta_0, I_{t-1})$$

i.e.  $F_{Z_t}(z_t, \theta_0, I_{t-1}) = F_{r_t}(J^{-1}(\theta_0) z_t, \theta_0, I_{t-1})$ . And its density is

$$f_{Z_t}(z_t, \theta_0, I_{t-1}) = \frac{\partial F_{Z_t}(z_t, \theta_0, I_{t-1})}{\partial z_t} = J^{-1}(\theta_0) f_{r_t}(J^{-1}(\theta_0) z_t, \theta_0, I_{t-1}).$$

We are only interested in the first element of  $Z_t$ , then the marginal density of  $Y_t$  can be derived by integrating out all the other elements:

$$f_{Y_t}(Y_t, \theta_0, I_{t-1}) = \int \cdots \int f_{Z_t}(z) dz_{2t} \cdots dz_{dt} = \int \cdots \int f_{r_t}(J^{-1}(\theta_0)z, \theta_0, I_{t-1}) J^{-1}(\theta_0) dz_{2t} \cdots dz_{dt}.$$

Notice that  $Y_t = w_t' r_t$  can be explicitly written as a function of  $\theta_0$  and  $I_{t-1}$ , i.e.  $Y_t \equiv Y_t(\theta_0) = Y(\theta_0, I_{t-1})$ , since  $w_t$  depends on  $\theta_0$  and  $I_{t-1}$ ,  $w_t \equiv w(\theta_0) = w(\theta_0, I_{t-1})$ .

Next, we prove Theorem 1 using empirical processes theory and a small variation of a weak convergence theorem in Delgado and Escanciano (2006). For simplicity, we write  $F_{Y_t(\theta_0)}(\theta_0) = F_{Y_t(\theta_0)}(m_\alpha(w_t, \theta_0, I_{t-1}))$  and  $f_{Y_t(\theta_0)}(\theta_0) = f_{Y_t(\theta_0)}(m_\alpha(w_t, \theta_0, I_{t-1}))$ .

Define the process

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

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

LEMMA A1: *Under Assumption A1-A5, the process  $K_n(c)$  is asymptotically tight with respect to  $c \in \mathcal{C}_K$ .*

The proof of Lemma A1 can be found in EO.

PROOF OF THEOREM 1: Simple but tedious algebra shows that for each  $c \in \mathcal{C}_K$ ,

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

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

$$|K_n(\hat{c}) - K_n(0)| = o_P(1). \tag{5}$$

Now, we will apply this argument with  $\hat{c} := \max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t - \theta_0)$ , with  $R$  denoting the in-sample sample size.

Thus, we should prove that under our three forecasting schemes

$$\max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t - \theta_0) = O_P(1) \tag{6}$$

holds.

- (i) Recursive: Our assumptions imply that  $\sqrt{t}S_t = \sum_{s=1}^t l(r_s, I_{s-1}, \theta_0)$  is a martingale with respect to  $\mathcal{F}_{t-1}$ , where  $S_t$  is implicitly defined. Hence, by Corollary 2.1 in Hall and Heyde (1980) and A5

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

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

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

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

Then, (5) holds for  $\hat{c} = \max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t - \theta_0)$ , and hence

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

which implies the decomposition

$$\begin{aligned} \frac{1}{\sqrt{P}} \sum_{t=R+1}^n (h_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha) &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [h_{t,\alpha}(\theta_0) - F_{Y_t(\theta_0)}(\theta_0)] \\ &\quad + \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{Y_t(\hat{\theta}_{t-1})}(\hat{\theta}_{t-1}) - F_{Y_t(\theta_0)}(\theta_0)] \\ &\quad + \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{Y_t(\theta_0)}(\theta_0) - \alpha] + o_P(1). \end{aligned} \tag{7}$$

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

$$\begin{aligned} A_{1n} &:= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left[ F_{Y_t(\hat{\theta}_{t-1})}(\hat{\theta}_{t-1}) - E[F_{Y_t(\hat{\theta}_{t-1})}(\hat{\theta}_{t-1})] - F_{Y_t(\theta_0)}(\theta_0) + E[F_{Y_t(\theta_0)}(\theta_0)] \right] \\ &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left( \frac{\partial F_{Y_t(\tilde{\theta}_{t-1})}(\tilde{\theta}_{t-1})}{\partial \tilde{\theta}'_{t-1}} - E \left[ \frac{\partial F_{Y_t(\tilde{\theta}_{t-1})}(\tilde{\theta}_{t-1})}{\partial \tilde{\theta}'_{t-1}} \right] \right) (\hat{\theta}_{t-1} - \theta_0), \end{aligned}$$

where  $\tilde{\theta}_{t-1}$  is between  $\hat{\theta}_{t-1}$  and  $\theta_0$ .

Note that A2 and A3 imply that  $E \left[ \sup_{\theta \in \Theta_0} \left| \frac{\partial F_{Y_t(\theta)}(\theta)}{\partial \theta'} \right| \right] < C$ . Hence, by the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2) and (6), then  $A_{1n} = o_P(1)$  holds. Similarly,

$$\begin{aligned} & \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left[ E[F_{Y_t(\hat{\theta}_{t-1})}(\hat{\theta}_{t-1})] - E[F_{Y_t(\theta_0)}(\theta_0)] \right] \\ &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n E \left[ \frac{\partial F_{Y_t(\theta_0)}(\theta_0)}{\partial \theta'_0} \right] (\hat{\theta}_{t-1} - \theta_0) + \\ & \quad + \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left[ E \left[ \frac{\partial F_{Y_t(\tilde{\theta}_{t-1})}(\tilde{\theta}_{t-1})}{\partial \tilde{\theta}'_{t-1}} \right] - E \left[ \frac{\partial F_{Y_t(\theta_0)}(\theta_0)}{\partial \theta'_0} \right] \right] (\hat{\theta}_{t-1} - \theta_0) \\ &: = B_{1n} + B_{2n}. \end{aligned}$$

Now, by the ULLN and (6), then  $B_{2n} = o_P(1)$  holds. Hence,

$$\left| \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left[ F_{Y_t(\hat{\theta}_{t-1})}(\hat{\theta}_{t-1}) - F_{Y_t(\theta_0)}(\theta_0) \right] - E \left[ \frac{\partial F_{Y_t(\theta_0)}(\theta_0)}{\partial \theta'_0} \right] \frac{1}{\sqrt{P}} \sum_{t=R+1}^n H(t-1) \right| = o_P(1).$$

The theorem follows from (7) and the last display.  $\square$

PROOF OF COROLLARY 1: Once Theorem 1 has been established, the proof follows the same arguments as in McCracken (2000, Theorem 2.3.1). Details are omitted to save space.  $\square$

## Appendix B: The Efficient Mean-Variance-Skewness Portfolios



The efficient mean-variance-skewness portfolios yield the maximum asymmetry for every feasible combination of mean and variance. The problem can be expressed as follows:

$$\max_{w_t \in \mathbb{R}^d} \varphi_t(\theta, \eta, \psi, b) \text{ s.t. } \begin{cases} w_t' \mu_t(\theta) = \mu_{0t} \\ w_t' \Sigma_t(\theta) w_t = \sigma_{0t}^2 \end{cases}$$

where  $\varphi_t(\theta, \eta, \psi, b)$  is the skewness and for a given expected return  $\mu_{0t}$ , the target variance  $\sigma_{0t}^2$  must be greater than or equal to that of the mean-variance frontier, that is  $\sigma_{0t}^2 \geq \mu_{0t}^2 / (\mu_t'(\theta) \Sigma_t^{-1}(\theta) \mu_t(\theta))$ . Following Mencia and Sentana (2009), we have the following proposition:

PROPOSITION B1: The efficient mean-variance-skewness portfolios that solve the above problem can be expressed as either

$$w_{1t} = \frac{\mu_{0t} + \Delta_t^{-1} \mu_t'(\theta) b}{\mu_t'(\theta) \Sigma_t^{-1}(\theta) \mu_t(\theta)} \Sigma_t^{-1}(\theta) \mu_t(\theta) - \frac{1}{\Delta_t} b$$

or

$$w_{2t} = \frac{\mu_{0t} - \Delta_t^{-1} \mu_t'(\theta) b}{\mu_t'(\theta) \Sigma_t^{-1}(\theta) \mu_t(\theta)} \Sigma_t^{-1}(\theta) \mu_t(\theta) + \frac{1}{\Delta_t} b$$

where

$$\Delta_t = \sqrt{\frac{(b' \Sigma_t(\theta) b)(\mu_t'(\theta) \Sigma_t^{-1}(\theta) \mu_t(\theta)) - (\mu_t'(\theta) b)^2}{\sigma_{0t}^2 (\mu_t'(\theta) \Sigma_t^{-1}(\theta) \mu_t(\theta)) - \mu_{0t}^2}}.$$

## Appendix C: The General Hyperbolic Distribution

Following McNeil, Frey and Embrechts (2005), the GH distribution can be introduced as a normal mean-variance mixture, in which the mixing variable is Generalized Inverse Gaussian (GIG) distributed.

DEFINITION C1: *The random vector  $X = (X_1, \dots, X_d)'$  is said to follow a  $d$ -dimensional GH distribution with parameters  $\lambda, \chi, \varphi, \alpha, \beta$  and  $\Upsilon$ , in short  $X \sim GH_d(\lambda, \chi, \varphi, \alpha, \beta, \Upsilon)$ , if*

$$X \stackrel{d}{=} \alpha + \xi \Upsilon \beta + \xi^{\frac{1}{2}} \Upsilon^{\frac{1}{2}} Z,$$

where  $\alpha, \beta \in R^d$ , and  $\Upsilon$  is a positive definite matrix of order  $d$ ,  $Z \sim N_d(0, I_d)$  follows a  $d$ -dimensional normal distribution,  $\xi \sim GIG(\lambda, \chi, \varphi)$  is a positive, scalar random variable independent of  $Z$ .

The conditional distribution of  $X$  given  $\xi$  is normal with conditional mean  $\alpha + \xi\Upsilon\beta$  and covariance matrix  $\xi\Upsilon$ , i.e.  $X|\xi \sim N(\alpha + \xi\Upsilon\beta, \xi\Upsilon)$ , which explains the so-called normal mean-variance mixture. Thus the mixing variable  $\xi$  could be interpreted as a stochastic volatility factor. The parameters of the mixing variable distribution,  $\lambda$ ,  $\chi$ , and  $\varphi$ , allow for flexible tail modeling;  $\alpha$  and  $\Upsilon$  play the roles of location vector and dispersion matrix; and the vector  $\beta$  introduces skewness into this distribution. We could reparametrize the GH distribution to get the standardized GH distribution with zero mean vector and identity covariance matrix.

DEFINITION C2: *The random vector  $X^* = (X_1^*, \dots, X_d^*)' \sim GH_d(\lambda, \chi, \varphi, \alpha, \beta, \Upsilon)$  is said to follow a  $d$ -dimensional standardized GH distribution, if*

$$\chi = 1, \alpha = -c(\beta, \lambda, \varphi)\beta, \text{ and } \Upsilon = \frac{\psi}{R_\lambda(\varphi)} \left[ I_d + \frac{c(\beta, \lambda, \varphi) - 1}{\beta'\beta} \beta\beta' \right]$$

where  $R_\lambda(\varphi) = \frac{K_{\lambda+1}(\varphi)}{K_\lambda(\varphi)}$ ,  $D_{\lambda+1}(\varphi) = \frac{K_{\lambda+2}(\varphi)K_\lambda(\varphi)}{K_{\lambda+1}^2(\varphi)}$  and  $c(\beta, \lambda, \varphi) = \frac{-1 + \sqrt{1 + 4[D_{\lambda+1}(\varphi) - 1]\beta'\beta}}{2[D_{\lambda+1}(\varphi) - 1]\beta'\beta}$ .

The parameters are reduced to be  $\lambda$ ,  $\varphi$  and  $\beta$  after the standardization. For analytical convenience,  $\lambda$  and  $\varphi$  are always replaced by  $\eta$  and  $\psi$ , where  $\eta = -0.5\lambda^{-1}$  and  $\psi = (1 + \varphi)^{-1}$ .