

# Backtesting Portfolio Value-at-Risk with Estimation Risk

Pei Pei \*  
Department of Economics  
Indiana University

## Abstract

Nowadays the most extensively used risk measure by financial institution is known as Value-at-Risk (VaR), which is defined as the maximum expected loss on an investment over a specified horizon at a given confidence level. To evaluate the accuracy and quality of the out-of-sample VaR forecast (backtesting procedures) is an important issue in practice. The purpose of this paper is to quantify the estimation risk in backtesting portfolio VaR and then to propose the corrected standard backtesting procedures robust to the estimation risk so that valid inferences could be carried out in out-of-sample portfolio VaR forecasts evaluations. Portfolio VaR forecast is intrinsically a problem in a multivariate setting where portfolio return is directly computed from asset returns and asset allocation. In this paper, a multivariate parametric dynamic model with standardized Generalized Hyperbolic (GH) innovations is used to model asset returns and the mean-variance-skewness approach is used to estimate the optimal portfolio weights such that the unobserved portfolio return could be estimated. As a result there are three sources of the estimation risk in the standard backtesting procedures of portfolio VaR, one from estimating the multivariate dynamic model, one from estimating the optimal portfolio weights and the other from estimating the unobserved portfolio return, which distinguishes this paper from the others. Escanciano and Olmo (2007) has considered the estimation risk in backtesting VaR but in the case of univariate financial time series such that the estimation risk only comes from estimating the univariate dynamic model. Finally, a simulation exercise illustrates the theoretical findings and a parametric bootstrap is used to improve the approximation by the asymptotic theory.

---

\*Contact information: Department of Economics, Indiana University, Wylie Hall 344, 100 S. Woodlawn, Bloomington, IN 47405. Email: ppei@indiana.edu. I am especially grateful to Juan Carlos Escanciano, my advisor, for his valuable advice, guidance and encouragement during the formation of this paper. This draft is very preliminary and incomplete. All errors are my own.

# 1 Introduction

Value-at-Risk (VaR) is one of the most popular risk measures used in financial institutions. It has also been extended to the portfolio VaR measure used for managing risks and returns under a multiple-asset portfolio. In financial terms, VaR is defined as the maximum expected loss on an investment over a specified horizon at a given confidence level, see Jorion (2001). In statistical terms, VaR is a quantile of the conditional distribution of portfolio returns given investor's information set. The Basel Accord allows financial institutions to have the freedom to develop their own model to compute VaR and also recommends a statistical framework to assess the accuracy and quality of the VaR forecast techniques, which was denominated backtesting. This paper investigates the backtesting of portfolio VaR with estimation risk, which is intrinsically a testing problem in a multivariate setup in which two estimators appear in the portfolio VaR predictions. The two estimators are the estimated parameters in a multivariate dynamic model for asset returns and the estimated optimal portfolio weights. This paper shows that the use of the standard backtesting procedures can be misleading, since the tests do not consider the impact of estimation risk from those two estimators and may use wrong critical values to assess market risk. The purpose of this paper is to quantify the estimation risk and to correct standard backtesting procedures to provide valid inference in out-of-sample analyses.

The effect of estimation risk on backtesting techniques has been studied in the case of univariate time series in Escanciano and Olmo (2007, EO hereafter). However, backtesting portfolio VaR is a multivariate inference problem, so the theory in EO cannot be directly applied for our problem. For an actively traded portfolio, its composition will change during each holding period, so in forecasting portfolio VaR, besides the parameters in the multivariate model for asset returns the optimal portfolio weights also need to be estimated at each time in the out-of-sample period. Most of the literature has assumed, however, that the optimal portfolio weights are known. The standard backtesting techniques assume that the true parameters and the optimal weights are known, but in practice, only the estimated ones are used in the portfolio VaR predictions. Therefore there exists uncertainty about the estimators coming from the in-sample data and the out-of-sample data depending on the forecasting scheme used. The use of the estimators adds additional terms in the

unconditional and independence backtesting procedures. Therefore, the estimation risk must be taken into account to construct valid inferences in out-of-sample portfolio VaR forecasts evaluations.

In order to quantify the estimation risk in portfolio backtesting procedures, there are two challenges. First of all, we need set up a general multivariate dynamic model for assets returns, estimate the model and derive the asymptotic distribution for the estimated parameters. Secondly, we need to build a proper optimal portfolio choice problem, which is analytically tractable to estimate the optimal portfolio weights and thus derive the asymptotic distribution for the estimated weights.

In modeling asset returns, we focus on a multivariate parametric dynamic model with standardized generalized hyperbolic (GH) innovations, see Mencia and Sentana (2005). There are several reasons to use this framework. First of all, it is more realistic than the normal distribution, since it is able to capture the stylized facts that the distribution of financial assets returns is non-normal but with negative skewness and positive excess kurtosis. Second, a nice property of the GH distribution is closed under linear transformations, i.e. a linear combination of the elements of GH random vector is a univariate GH random variable. Without derivatives included in a portfolio, the portfolio return will be a linear combination of the component asset returns. Thus, the VaR of any portfolio whose components are modeled in such a framework can be partly determined from the multivariate model estimates and the corresponding multivariate distribution of the innovations. Last but not the least, GH distribution is a very flexible parametric class. It contains many subclasses, which are of the most important multivariate distributions already used in the literature such as Normal, symmetric Student  $t$ , skewed Student  $t$ , asymmetric Normal-Gamma, Normal Inverse Gaussian and Hyperbolic distributions.

In allocating assets, we use the so-called mean-variance-skewness (MVS) analysis other than the widely used mean-variance analysis. Firstly because higher order moments cannot be neglected considering the asymmetric distribution of financial asset returns. Secondly, under our multivariate model setup asset returns will jointly follow a GH distribution, which can be expressed as a location-scale mixture of normals. Mencia and Sentana (2008) 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 close 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 portfolio VaR backtesting procedures robust to estimation risk without making particular distributional assumptions for asset returns. Section 3 applies the general procedures to a multivariate parametric setting in which asset returns are model by a multivariate parametric dynamic model with standardized GH innovations and the optimal portfolio weights are chosen by MVS method. A simulation exercise using a simple case illustrates the theoretical findings in section 4. Finally, section 5 concludes.

## 2 Portfolio VaR backtesting techniques robust to estimation risk: A general theory

This section proposes general portfolio VaR backtesting procedures robust to estimation risk. Portfolio VaR backtesting is the procedure to evaluate the accuracy of a portfolio VaR model forecasts in which both asset returns and asset allocation needs to be considered. Therefore, in order to examine the effects of estimation risk on the procedures, we need to elaborate 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 and  $w_t = (w_{1t}, w_{2t}, \dots, w_{dt})'$  the vector of portfolio weights, where  $w_t \in \mathbb{R}^d$  and  $\sum_{i=1}^d w_{it} = 1$ . 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.  $I_{t-1} = (r'_{t-1}, z'_{t-1}, r'_{t-2}, z'_{t-2} \dots)'$ , and the conditional distribution of  $r_t$  given  $I_{t-1}$  is a multivariate conditional distribution denoted as  $F_{r_t}(\cdot, \theta_0, I_{t-1})$ . Since optimal portfolio allocation is related to the joint distribution of asset returns and is carried out at time  $t$  conditioning on the information at time  $t - 1$ , then the optimal portfolio weights  $w_t^*$  will depend on the distribution parameter  $\theta_0$  and the information set  $I_{t-1}$ . We make this explicitly by writing  $w_t^* \equiv w^*(\theta_0) = w^*(\theta_0, I_{t-1})$ , where  $w_t^* \in \mathcal{F}_{t-1}$ . Therefore the portfolio return at time  $t$  is  $Y_t = w_t^{*'} r_t$ . By this linear

projection  $Y_t = w_t^* r_t$ , we can also derive the univariate conditional distribution of  $Y_t$  given  $I_{t-1}$  denoted as  $Y_t | I_{t-1} \sim F_{Y_t}(\cdot, w_t^*, \theta_0, I_{t-1})$  and its associated density  $f_{Y_t}(\cdot)$ , in which  $w_t^*$  can be treated as constants since it is known at time  $t-1$ . Assuming that  $F_{Y_t}(\cdot, w_t^*, \theta_0, I_{t-1})$  is continuous, then the conditional portfolio VaR at a given confidence level  $1 - \alpha$  given  $I_{t-1}$  is defined as the  $\alpha^{th}$  quantile of the distribution of  $Y_t | I_{t-1}$  satisfying the equation

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

Actually  $m_\alpha(\theta, I_{t-1})$  is just the inverse of  $F_{Y_t}(\cdot, w_t^*, \theta_0, I_{t-1})$  at the level  $\alpha$  with respect to the first argument, i.e.  $m_\alpha(\theta, I_{t-1}) = F_{Y_t}^{-1}(\alpha, w_t^*, \theta_0, I_{t-1})$ . Therefore, it seems that portfolio VaR can be treated as a univariate parametric VaR model, similar to the ones studied in EO. However, there is an important difference that the optimal portfolio weights  $w_t^*$  are not observable and must be estimated, which result in the portfolio return  $Y_t$  being unobservable. This subtle difference has important implications for our testing problem. First, it shows that a purely univariate approach to portfolio VaR is in general not possible. Second, this difference makes the results in EO not applicable to our present framework. This paper will show that the estimated  $w_t^*$  in  $Y_t$  also add an extra term in the estimation effect on portfolio backtesting.

Denote the above multivariate portfolio VaR model as  $m_\alpha(w^*, \theta, I_{t-1})$  and its derivative with respect to  $\theta$  as  $g_\alpha(w^*, \theta, I_{t-1})$ . Notice that the optimal portfolio weights add an additional term in the derivative  $g_\alpha(w^*(\theta), \theta, I_{t-1})$  comparing to that in EO. Namely,

$$g_\alpha(w^*, \theta, I_{t-1}) = m_{\alpha,1}(w^*, \theta, I_{t-1}) \frac{\partial w^*(\theta)}{\partial \theta} + m_{\alpha,2}(w^*, \theta, I_{t-1}).$$

where  $m_{\alpha,j}(w^*, \theta, I_{t-1})$  is the derivative of  $m_\alpha(w^*, \theta, I_{t-1})$  with respect to the  $j^{th}$  argument.

By definition (1), this parametric portfolio VaR model is correctly specified if and only if

$$E[h_{t,\alpha}(\theta) | I_{t-1}] = \alpha \text{ a.s. for some } \theta_0, \quad (2)$$

where  $h_{t,\alpha}(\theta) := 1(Y_t \leq m_\alpha(w^*, \theta, I_{t-1}))$  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”. Our inference procedure will focus on testing one of the most popular implications of condition (2) given by Christoffersen (1998),

$$E \left[ h_{t,\alpha}(\theta_0) \mid \tilde{h}_{t-1,\alpha}(\theta_0) \right] = \alpha \text{ a.s. for some } \theta_0, \quad (3)$$

where  $\tilde{h}_{t-1,\alpha}(\theta_0) := (h_{t-1,\alpha}(\theta_0), h_{t-2,\alpha}(\theta_0) \dots)'$ . This condition in turn is equivalent to

$$\{h_{t,\alpha}(\theta_0)\} \text{ are iid } Ber(\alpha) \text{ random variables for some } \theta_0, \quad (4)$$

where  $Ber(\alpha)$  represents for a Bernoulli random variable with parameter  $\alpha$ . Therefore, the problem of evaluating the accuracy of VaR forecasts can be reduced to the problem of examining the unconditional coverage and independence properties of the hit sequence  $\{h_{t,\alpha}(\theta_0)\}$ . To test for  $E[h_{t,\alpha}(\theta_0)] = \alpha$  is called unconditional backtesting and to test for  $\{h_{t,\alpha}(\theta_0)\}$  being *i.i.d* is called independence test.

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 the VaR forecasts. For simplicity we only consider one-step-ahead forecasts. Assume that the first  $R$  observations are used to estimate  $\hat{\theta}_R$  and  $w^*(\hat{\theta}_R)$  in the first forecast, and then we will have  $P = n - R$  predictions to be evaluated. The first VaR forecasts is  $VaR_{R+1,1}(\hat{\theta}_R) = m_\alpha(w^*(\hat{\theta}_R), \hat{\theta}_R, I_R)$  and the further forecasts are  $VaR_{t+1,1}(\hat{\theta}_t) = m_\alpha(w^*(\hat{\theta}_t), \hat{\theta}_t, I_t)$ ,  $R \leq t \leq n - 1$ , where  $\hat{\theta}_t$  and  $w^*(\hat{\theta}_t)$  are estimated using observations  $s = 1, \dots, t$ .

## 2.2 Unconditional backtesting robust to estimation risks

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

$$K_p \equiv K(P, R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^n (h_{t,\alpha}(\theta_0) - \alpha) = \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [1(Y_t \leq m_\alpha(w^*, \theta, I_{t-1})) - \alpha] \quad (5)$$

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 assumptions of the parameter  $\theta_0$  being known and the portfolio return  $Y_t$  being observable, and using the critical values from the standard normal distribution. In practice, however, the true parameter  $\theta_0$  is not known and must be estimated, and the portfolio return  $Y_t$  is also unobservable

and can be estimated from asset returns  $r_t$  and optimal portfolio weights  $w_t^*$ , using some portfolio choice procedure, e.g. mean-variance analysis. Thus the test statistic becomes

$$S_p \equiv S(P, R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left[ 1 \left( \widehat{Y}_t \leq m_\alpha(w^*(\widehat{\theta}_{t-1}), \widehat{\theta}_{t-1}, I_{t-1}) \right) - \alpha \right] \quad (6)$$

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

Therefore, to quantify estimation risk in the multivariate VaR model, we must consider the two estimators,  $\widehat{\theta}_{t-1}$  and  $w^*(\widehat{\theta}_{t-1})$ , as well as the calculated portfolio return  $\widehat{Y}_t$ . In different words, there should be three sources of estimation risk in the multivariate VaR model, while in the case of univariate time series studied in EO (2007), the estimation risk only comes from the estimated parameters in the univariate VaR model. In this sense, the theory developed in EO (2007) cannot be directly applied but still can be adapted to our case as this paper will show. More concretely, we show that those three components,  $\widehat{\theta}_{t-1}$ ,  $w^*(\widehat{\theta}_{t-1})$  and  $Y_t(\widehat{\theta}_{t-1})$ , respectively, introduce asymptotically an extra term in the, still normal, limiting distribution, changing the resulting asymptotic variance of  $S_p$ .

We need some assumptions which are similar to those in EO (2007).

**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  $m_\alpha(w^*, \theta, I_{t-1})$  is continuously differentiable in  $\theta$  and  $w^*$  (*a.s.*), and  $w^*(\theta)$  is also continuously differentiable in  $\theta$  (*a.s.*), such that for its derivative  $g_\alpha(w^*, \theta, I_{t-1})$ ,  $E[\sup_{\theta \in \Theta_0} |g_\alpha(w^*, \theta, I_{t-1})|^2] < C$ , 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  $V := E \left[ l(r_t, I_{t-1}, \theta_0) l'(r_t, I_{t-1}, \theta_0) \right]$  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,*

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

where the score component in estimation risk, say  $A = E\left[\frac{\partial F_{Y_t}(m_\alpha(w^*, \theta_0, I_{t-1}))}{\partial \theta'_0}\right]$ , can be partitioned into three components  $A = A_1 + A_2 + A_3$ , where

$$\begin{aligned}
A_1 &= E \left[ \underbrace{m'_{\alpha,1}(w^*, \theta_0, I_{t-1}) \frac{\partial w^*(\theta_0)}{\partial \theta_0} f_{Y_t}(m_\alpha(w^*, \theta_0, I_{t-1}))}_{\text{Due to portfolio weights estimation}} \right] \\
A_2 &= E \left[ \underbrace{m'_{\alpha,2}(w^*, \theta_0, I_{t-1}) f_{Y_t}(m_\alpha(w^*, \theta_0, I_{t-1}))}_{\text{Due to estimation of dynamics}} \right] \\
A_3 &= E \left[ \underbrace{\int^{m_\alpha(w^*, \theta_0, I_{t-1})} [f_{Y_{t,1}}(y_t, \theta_0, I_{t-1}) \frac{\partial y_t(\theta_0)}{\partial \theta'_0} + f'_{Y_{t,2}}(y_t, \theta_0, I_{t-1})] dy_t}_{\text{Due to unobserved optimal portfolio}} \right]
\end{aligned}$$

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.  $F(m_\alpha(w^*, \theta_0, I_{t-1})) = \alpha$ , then model risk vanishes, but we still have the estimation risk to deal with. Notice that there are three sources of estimation risk under the multivariate VaR model, one from estimating parameters in the multivariate model for asset returns, one from estimating the optimal portfolio weights and the other from estimating the unobserved portfolio return. Without any of those components, we may make wrong inference in the unconditional backtesting procedures.

Next corollary presents the asymptotic distribution of  $S_p$  with estimation risk, which provides the necessary corrections to carry out valid asymptotic inference for unconditional backtests free of estimation risk,

COROLLARY 1: *Under Assumptions A1-A5 and (2),*

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

where

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

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

		Forecast Scheme	$\lambda_{hl}$	$\lambda_{ll}$
<b>Corollary 1</b>		<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>	$0$	$\pi$

Therefore, as long as we get the consistent estimates,  $\hat{\rho}$ ,  $\hat{\pi}$ ,  $\hat{V}$ , and  $\hat{A}$ , then the asymptotic variance  $\sigma_u^2$  can be consistently estimated by  $\hat{\sigma}_u^2 = \alpha(1 - \alpha) + 2\hat{\lambda}_{hl}\hat{A}\hat{\rho} + \hat{\lambda}_{ll}\hat{A}\hat{V}\hat{A}'$ . The ratio  $\pi$  can be consistently estimated by  $\hat{\pi} = \frac{P}{R}$ ; the consistent estimators of  $V$  and  $\rho$  are  $\hat{V} = \frac{1}{P} \sum_{t=R+1}^n l(r_t, I_{t-1}, \hat{\theta}_{t-1})l'(r_t, I_{t-1}, \hat{\theta}_{t-1})$  and  $\hat{\rho} = \frac{1}{P} \sum_{t=R+1}^n (h_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha)l(r_t, I_{t-1}, \hat{\theta}_{t-1})$ , respectively. The vector  $A_1 + A_2$  can be consistently estimated, for instance, by the estimator introduced in Giacomini and Komunjer (2005)

$$\hat{A}_\tau = -\frac{1}{P} \sum_{t=R+1}^n \frac{1}{\tau} \exp[(Y_t - m_\alpha(w^*, \hat{\theta}_{t-1}, I_{t-1}))/\tau] h_{t,\alpha}(\hat{\theta}_{t-1}) [g'_\alpha(w^*, \hat{\theta}_{t-1}, I_{t-1})]$$

with  $\tau \rightarrow 0$  as  $n \rightarrow \infty$ .

Then, the corrected standard unconditional backtesting test statistic can be proposed as follows:

$$\tilde{S}_p \equiv \tilde{S}(P, R, \hat{\theta}_{t-1}) := \frac{1}{\hat{\sigma}_u \sqrt{P}} \sum_{t=R+1}^n (h_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha)$$

which converges to a standard normal random variable by Corollary 2 in EO, i.e.

$$\tilde{S}_p \xrightarrow{d} N(0, 1)$$

In order to make valid inference, we should use the corrected test statistic  $\tilde{S}_p$  in the unconditional backtesting procedures. If we carry out inference for  $S_p$  as if it were  $K_p$ , we may use the wrong critical values.

### 2.3 Independence tests robust to estimation risks

In independence tests the hypothesis of interest is that the sequence of hits  $\{h_{t,\alpha}(\theta_0)\}_{t=R+1}^n$  are *iid*. A likelihood ratio test introduced by Christoffersen (1998) is an early and influential one. Nowadays more general tests have been proposed based on the autocovariances

$$\zeta_j = Cov(h_{t,\alpha}(\theta_0), h_{t-j,\alpha}(\theta_0)), \text{ for } j \geq 1$$

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

$$\zeta_{P,j} = \frac{1}{P-1} \sum (h_{t,\alpha}(\theta_0)h_{t-j,\alpha}(\theta_0) - \alpha^2), \text{ for } j \geq 1$$

In fact, tests based on  $\{\zeta_{P,j}\}$  are joint tests of the iid and the unconditional hypothesis, since the fact that  $E[h_{t,\alpha}(\theta_0)] = \alpha$  is explicitly used in those tests.

Similarly with the analysis of unconditional backtests, in practice it is the estimated parameters, the estimated optimal portfolio weights and the estimated unobserved portfolio return that are used to test for  $\{h_{t,\alpha}(\theta_0)\}_{t=R+1}^n$  being *iid*, such as

$$\hat{\zeta}_{P,j} = \frac{1}{P-1} \sum (h_{t,\alpha}(\hat{\theta}_{t-1})h_{t-j,\alpha}(\hat{\theta}_{t-1-j}) - \alpha^2), \text{ for } j \geq 1$$

where  $h_{t,\alpha}(\hat{\theta}_{t-1}) = 1 \left( \hat{Y}_t \leq m_\alpha(w^*, \hat{\theta}_{t-1}, I_{t-1}) \right) = 1 \left( (w^*(\hat{\theta}_{t-1}))' r_t \leq m_\alpha(w^*, \hat{\theta}_{t-1}, I_{t-1}) \right)$ . Therefore, in order to examine the effect of estimation risk on independence backtests, we should also take into account the three sources considered in the above analysis of unconditional backtests.

Next theorem is the equivalent to Theorem 1 for the joint and independence backtesting tests. Denote  $B \equiv B_j := E \left[ \frac{\partial F(m_\alpha(w^*, \theta_0, I_{t-1}))}{\partial \theta_0'} \{h_{t-j,\alpha}(\theta_0) + \alpha\} \right]$ , and  $\eta \equiv \eta_j := E \left[ (h_{t,\alpha}(\theta_0)h_{t-j,\alpha}(\theta_0) - \alpha^2) l(r_t, I_{t-1}, \theta_0) \right]$ .

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

$$\sqrt{P-j}(\hat{\zeta}_{P,j} - \zeta_{P,j}) = B \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^n H(t-j-1) + o_P(1)$$

where the vector  $B$  could also be partitioned into three components as with vector  $A$  in Theorem 1. Of the three components, one is from the estimation of dynamics, one is from the estimation of the optimal weights and the other is from estimating the observed portfolio return. Details are omitted to save space. By Corollary 3 in EO (2007), for each  $j \geq 1$ ,

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

where  $\sigma_c^2 = \alpha^2(1-\alpha)^2 + 2\lambda_{hl}B\eta + \lambda_{ll}BV B'$  with  $\lambda_{hl}$  and  $\lambda_{ll}$  defined as in Corollary 2. The vector  $B$  could be consistently estimated by  $\widehat{B}_\tau$ , where  $\widehat{B}_\tau = -\frac{1}{P-j} \sum_{t=R+j+1}^n \frac{1}{\tau} \exp[(Y_t - m_\alpha(w^*, \widehat{\theta}_{t-1}, I_{t-1}))/\tau] h_{t,\alpha}(\widehat{\theta}_{t-1}) \{h_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) + \alpha\} [g'_\alpha(w^*, \widehat{\theta}_{t-1}, I_{t-1})]$ , with  $\tau \rightarrow 0$  as  $n \rightarrow \infty$ . And  $\eta$  could be consistently estimated by

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

$\pi$  and  $V$  could be consistently estimated by  $\widehat{\pi}$  and  $\widehat{V}$  showed in the previous section. As a result,  $\sigma_c^2$  could be consistently estimated by  $\widehat{\sigma}_c^2 = \alpha^2(1-\alpha)^2 + 2\widehat{\lambda}_{hl}\widehat{B}\widehat{\eta} + \widehat{\lambda}_{ll}\widehat{B}_\tau\widehat{V}\widehat{B}'_\tau$ .

Therefore, we could propose the corrected standard independence backtesting test statistic as follows

$$\widetilde{\zeta}_{P,j} \equiv \widetilde{\zeta}(P, R, \widehat{\theta}_{t-1}) = \frac{\sqrt{P-j}\widehat{\zeta}_{P,j}}{\widehat{\sigma}_c} \xrightarrow{d} N(0, 1)$$

### 3 An application under a particular distributional assumption

The previous section proposes the corrected standard portfolio VaR backtesting procedures robust to estimation risk for a general case. This theory is applicable to many portfolio VaR models, because there are no particular distributional assumptions for asset returns. In this section, we will apply the general theory to a particular setting in which asset returns are model by a multivariate parametric dynamic model with standardized GH innovations and the optimal portfolio weights are chosen by MVS method.

### 3.1 Multivariate dynamic model for asset returns

We turn to a realistic distribution for asset returns, the General Hyperbolic distribution introduced by Barndorff-Nielsen (1977), since it is able to capture the stylized facts of fat tails and negative asymmetry from financial time series. Another attractive property of the GH distribution is that it is closed under linear transformations such that the distribution of any portfolio combining the assets is still a GH distribution. In addition, as we mentioned in the introduction, it is a rather flexible distribution that contains many well-known subclasses that have been the most important distributions already used in the literature.

#### 3.1.1 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 1: *The random vector  $X = (X_1, \dots, X_d)'$  is said to follow a  $d$ -dimensional GH distribution with parameters  $\lambda, \chi, \psi, \alpha, \beta$  and  $\Upsilon$ , in short  $X \sim GH_d(\lambda, \chi, \psi, \alpha, \beta, \Upsilon)$ , 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,  $\xi \sim GIG(\lambda, \chi, \psi)$  is a positive, scalar random variable independent of  $Z$ .

We observe that 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  $\psi$ , 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, see Proposition 1 in Mencia and Sentana (2005), to get the standardized GH distribution with zero mean vector and identity covariance matrix.

DEFINITION 2: The random vector  $X^* = (X_1^*, \dots, X_d^*)' \sim GH_d(\lambda, \chi, \psi, \alpha, \beta, \Upsilon)$  is said to follow a  $d$ -dimensional standardized GH distribution, if

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

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

The parameters are reduced to be  $\lambda$ ,  $\psi$  and  $\beta$  after the standardization.

The GH distribution is closed under linear transformations. This means that any linear combination of the GH random vector must have a univariate GH distribution. Take the standardized GH distribution as example, see Proposition 2 in Mencía and Sentana (2005):

PROPOSITION 1: Let  $X^*$  be a  $d$ -dimensional standardized GH random vector with parameters  $\lambda$ ,  $\psi$  and  $\beta$ . Then, for any vector  $\omega \in R^d$ , with  $\omega \neq 0$ ,  $s^* = \frac{\omega' X^*}{\sqrt{\omega' \omega}}$  is distributed as a univariate standardized GH random variable with parameter  $\lambda$ ,  $\psi$  and  $\beta(\omega) = \frac{c(\beta, \lambda, \psi)(\omega' \beta) \sqrt{\omega' \omega}}{\omega' \omega + [c(\beta, \lambda, \psi) - 1](\omega' \beta)^2 / (\beta' \beta)}$ .

Note that only the skewness parameter is affected by the weights  $\omega$ . This result will be used later such that we could relatively easily compute VaR based on the so-called variance-covariance method.

### 3.1.2 Multivariate dynamic model

We model asset returns by a multivariate location-scale dynamic model with standardized GH innovations. The model can be written as follows

$$r_t = \mu_t(\theta) + \Sigma_t^{\frac{1}{2}}(\theta) \varepsilon_t, \quad (7)$$

where  $\mu_t(\theta) = \mu(I_{t-1}; \theta)$  is a  $d$ -dimensional vector of conditional mean returns and  $\Sigma_t(\theta) = \Sigma(I_{t-1}; \theta)$  is a symmetric  $d \times d$  conditional covariance matrix of returns. Both  $\mu_t(\theta)$  and  $\Sigma_t(\theta)$  are conditional upon the information set  $I_{t-1}$ . The  $k \times 1$  vector  $\theta$  collects all the unknown parameters, and  $\theta_0$  denotes the true parameter values.  $\Sigma_t^{\frac{1}{2}}(\theta)$  is some  $d \times d$  "square root" matrix such that  $\Sigma_t^{1/2}(\theta) \Sigma_t^{\frac{1}{2}}(\theta) = \Sigma_t(\theta)$ , and  $\varepsilon_t$  is a vector martingale difference sequence satisfying  $E(\varepsilon_t | I_{t-1}; \theta_0) = 0$  and  $V(\varepsilon_t | I_{t-1}; \theta_0) = I_d$ . As a consequence,  $E(r_t | I_{t-1}; \theta_0) = \mu_t(\theta_0)$  and  $V(r_t | I_{t-1}; \theta_0) = \Sigma_t(\theta_0)$ ; for the portfolio return  $Y_t = w_t' r_t$ , since  $w_t \in \mathcal{F}_{t-1} = \sigma(I_{t-1})$ , the corresponding moments are  $E(Y_t | I_{t-1}; \theta_0) = w_t' \mu_t(\theta_0)$  and  $V(Y_t | I_{t-1}; \theta_0) = w_t' \Sigma_t(\theta_0) w_t$ .

We also assume that  $\varepsilon_t|I_{t-1}$  follow a standardized GH distribution, i.e.  $\varepsilon_t|I_{t-1} \sim GH_d(\lambda, 1, \psi, \alpha, \beta, \Upsilon)$ , where  $\alpha$  and  $\Upsilon$  are functions of  $\lambda$ ,  $\psi$  and  $\beta$  defined as in Definition 5. Hence, this parametric multivariate dynamic model is able to capture asymmetries and kurtosis that have been found in many empirical studies with financial time series.

Given that  $\varepsilon_t$  is not generally observable, the parametrization will be variant to the choice of “square root” matrix,  $\Sigma_t^{\frac{1}{2}}(\theta)$ , except in univariate models or in multivariate models in which either  $\Sigma_t(\theta)$  is time-invariant or  $\beta = 0$ . As a result, the conditional distribution of  $r_t|I_{t-1}$  will depend on the choice of  $\Sigma_t^{\frac{1}{2}}(\theta)$ , which is not desirable, so we still follow Mencia and Sentana (2005) to parametrize  $\beta$  as a function of past information and a new vector of parameters  $b$  in the following way:

$$\beta_t(\theta, b) = \Sigma_t^{\frac{1}{2}}(\theta)b \quad (8)$$

then it follows that the resulting GH log-likelihood function is irrelevant to the choice of  $\Sigma_t^{\frac{1}{2}}(\theta)$ . In addition, for the analytical convenience we replace  $\lambda$  and  $\psi$  by  $\eta = 0.5\lambda^{-1}$  and  $\delta = (1 + \psi)^{-1}$ .

Finally, we collect all the unknown parameters into the vector  $\phi = (\theta', \eta, \delta, b')'$  with  $k + d + 2$  elements. Under such a framework, these parameters can be estimated simultaneously by conditional maximum likelihood (ML) method. The conditional log-likelihood function of a sample of size  $n$  takes the form  $L = \sum_{t=1}^n l(r_t|I_{t-1}; \phi)$ , where  $l(r_t|I_{t-1}; \phi)$  is the conditional log-density of  $r_t$  given  $I_{t-1}$  and  $\phi$ . To overcome the latency of the mixing variable  $\xi$  the EM (expectation-maximization) algorithm is used. The score function is

$$s_t(\phi) = \frac{\partial l(r_t|I_{t-1}; \phi)}{\partial \phi} = E\left[\frac{\partial l(r_t|\xi_t, I_{t-1}; \phi)}{\partial \phi}\right] + E\left[\frac{\partial l(\xi_t|I_{t-1}; \phi)}{\partial \phi}\right]$$

The information matrix is

$$\mathcal{I}(\phi) = E[s_t(\phi)s_t'(\phi)]$$

We will borrow the results from Mencia and Sentana (2005), who derives the analytical formulae for the score function  $s_t(\phi)$  and the information matrix  $\mathcal{I}(\phi)$ . Let  $\hat{\phi}_t$  denote the ML estimator of the true parameter  $\phi_0$ . Given correct specification, under certain regularity conditions,  $\hat{\phi}_t$  will be consistent and asymptotically normally distributed with

a covariance matrix which is the inverse of the information matrix, i.e.  $\sqrt{t}(\widehat{\phi}_t - \phi_0) \xrightarrow{d} N_{k+d+2}(0, \mathcal{I}(\phi_0)^{-1})$ .

By definition, we have that

$$\begin{aligned} \alpha &= P(Y_t \leq m_\alpha(w^*, \theta_0, I_{t-1}) | I_{t-1}) = P(w'_t r_t \leq m_\alpha(w^*, \theta_0, I_{t-1}) | I_{t-1}) \\ &= P\left(\frac{w'_t r_t - w'_t \mu_t}{\sqrt{w'_t \Sigma_t w_t}} \leq \frac{m_\alpha(w^*, \theta_0, I_{t-1}) - w'_t \mu_t}{\sqrt{w'_t \Sigma_t w_t}} \middle| I_{t-1}\right) \\ &= P\left(\frac{w'_t \Sigma_t^{\frac{1}{2}} \varepsilon_t}{\sqrt{(w_t \Sigma_t^{\frac{1}{2}})' \Sigma_t^{\frac{1}{2}} w_t}} \leq \frac{m_\alpha(w^*, \theta_0, I_{t-1}) - w'_t \mu_t}{\sqrt{(w_t \Sigma_t^{\frac{1}{2}})' \Sigma_t^{\frac{1}{2}} w_t}} \middle| I_{t-1}\right) \end{aligned}$$

Denote  $v_t = \frac{w'_t \Sigma_t^{\frac{1}{2}} \varepsilon_t}{\sqrt{(w_t \Sigma_t^{\frac{1}{2}})' \Sigma_t^{\frac{1}{2}} w_t}}$ , where  $v_t \in \mathbb{R}^d$ . By Proposition 1, we find that  $v_t | I_{t-1}$  is a univariate standardized GH random variable with parameters  $\lambda, \psi, \beta_{v_t}, \alpha_{v_t}$  and  $\Upsilon_{v_t}$ , where  $\beta_{v_t} = \frac{c(\beta, \lambda, \psi)((w_t \Sigma_t^{\frac{1}{2}})' \beta) \sqrt{w'_t \Sigma_t w_t}}{w'_t \Sigma_t w_t + [c(\beta, \lambda, \psi) - 1]((w_t \Sigma_t^{\frac{1}{2}})' \beta)^2 / (\beta' \beta)}$ . By Definition 2, we get  $\alpha_{v_t} = -c(\beta_{v_t}, \lambda, \psi) \beta_{v_t}$  and  $\Upsilon_{v_t} = \frac{\psi}{R_\lambda(\psi)} c(\beta_{v_t}, \lambda, \psi)$ . In short,  $v_t | I_{t-1} \sim GH(\lambda, 1, \psi, \alpha_{v_t}, \beta_{v_t}, \Upsilon_{v_t})$ . Assume the cumulative distribution function of  $v_t | I_{t-1}$  is  $Q_{v_t}(\cdot)$ , then the VaR for the portfolio with portfolio weights  $w_t$  at a given level  $\alpha$  can be reduced to the well-known variance-covariance formula

$$m_\alpha(w^*, \theta_0, I_{t-1}) = w_t^* \mu_t + \sqrt{w_t^{*'} \Sigma_t w_t^*} Q_{v_t}^{-1}(\alpha) \quad (9)$$

where  $Q_{v_t}^{-1}(\alpha)$  is the  $\alpha$ -th quantile of the univariate standardized GH distribution of  $v_t | I_{t-1}$ . Generally, the portfolio VaR is a function of the pre-specified level  $\alpha$ , the  $\alpha$ -th quantile of the univariate standardized GH distribution  $Q_{v_t}^{-1}(\alpha)$ , the portfolio weights  $w_t$ , the conditional mean return vector  $\mu_t$  and of the conditional covariance matrix  $\Sigma_t$ , among which except  $\alpha$  all depend on the information set  $I_{t-1}$  and the true parameters  $\phi_0$ . In practice, the true parameters  $\phi_0$  are unknown and must be estimated from the data to get the estimated mean return vector  $\widehat{\mu}_t = \mu_t(I_{t-1}; \widehat{\theta}_{t-1})$ , the estimated covariance matrix  $\widehat{\Sigma}_t = \Sigma_t(I_{t-1}; \widehat{\theta}_{t-1})$  by forecasting and the estimated quantile  $Q_{v_t}^{-1}(\alpha)$ . Consequently, in order to apply variance-covariance formula (9) to forecast the portfolio VaR based on information set up to time  $t-1$ , we should replace  $\mu_t, \Sigma_t$  and  $Q_{v_t}^{-1}(\alpha)$  by  $\widehat{\mu}_t, \widehat{\Sigma}_t$  and

$\widehat{Q}_{v_t}^{-1}(\alpha)$ , respectively, and also replace  $w_t$  by its estimate as well. The estimate of the optimal portfolio weights  $w_t$  will be discussed in the next section.

Additionally, there is a linear relationship between  $r_t$  and  $\varepsilon_t$  based on the model setup, and  $\varepsilon_t|I_{t-1}$  is assumed to follow a standardized GH distribution,  $\varepsilon_t|I_{t-1} \sim GH_d(\lambda, 1, \psi, \alpha, \beta, \Upsilon)$ , so the vector of asset returns  $r_t$  can also be expressed as a location-scale mixture of normals,

$$r_t = \alpha_r + \xi \Upsilon_r \beta_r + \xi^{\frac{1}{2}} \Upsilon_r^{\frac{1}{2}} Z,$$

where  $\alpha_r = \mu_t + \alpha \Sigma_t^{\frac{1}{2}}$ ,  $\beta_r = \beta \Sigma_t^{-\frac{1}{2}}$  and  $\Upsilon_r = \Sigma_t \Upsilon$ . By Definition 4,  $r_t|I_{t-1} \sim GH_d(\lambda, 1, \psi, \alpha_r, \beta_r, \Upsilon_r)$ .

There is another linear relationship between  $Y_t$  and  $v_t$ , which is  $Y_t = w_t' \mu_t + \sqrt{w_t' \Sigma_t w_t} v_t$ , and we have  $v_t|I_{t-1} \sim GH(\lambda, 1, \psi, \alpha_{v_t}, \beta_{v_t}, \Upsilon_{v_t})$ . Similarly, by Definition 1 we could obtain that the conditional distribution of  $Y_t$  given  $I_{t-1}$  is a univariate GH distribution as follows.

$$Y_t|I_{t-1} \sim GH(\lambda, 1, \psi, \alpha_{Y_t}, \beta_{Y_t}, \Upsilon_{Y_t}),$$

where  $\alpha_{Y_t} = w_t' \mu_t + \alpha_{v_t} \sqrt{w_t' \Sigma_t w_t}$ ,  $\beta_{Y_t} = \frac{\beta_{v_t}}{\sqrt{w_t' \Sigma_t w_t}}$  and  $\Upsilon_{Y_t} = w_t' \Sigma_t w_t \Upsilon_{v_t}$ .

### 3.2 Optimal portfolio selection

In terms of optimal portfolio selection, we will use the so-called mean-variance-skewness analysis other than the widely used mean-variance analysis. Firstly of all, the above parametric multivariate model for asset returns is under the assumption of the GH distribution and is able to capture the asymmetric distribution of financial asset returns, thus higher order moments cannot be neglected in asset allocation problem, while the mean-variance analysis is the natural approach under Gaussian distributional assumption.

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 Sentata (2008) 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. Naturally, the MVS analysis becomes the most appropriate method of optimal portfolio allocation.

Most attractively, the close form solution for the optimal portfolio weights could be explicitly obtained under our model setup by this MVS method. To save space, we only

provide the optimal portfolio weights in closed-form obtained by MVS analysis. There are two potential solutions, both of which can be expressed as linear combination of the mean-variance efficient portfolio  $\Sigma_t^{-1}(\theta)\mu_t(\theta)$  and the skewness-variance efficient portfolio  $b$ . They are

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

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

where  $\Delta_t = \sqrt{\frac{(b'\Sigma_t(\theta)b)(\mu_t'(\theta)\Sigma_t^{-1}(\theta)\mu_t(\theta)) - (\mu_t'(\theta)b)^2}{\sigma_{0t}^2(\mu_t'(\theta)\Sigma_t^{-1}(\theta)\mu_t(\theta)) - u_{0t}^2}}$ ,  $u_{0t}$  and  $\sigma_{0t}^2$  are the target expected return and variance, respectively.

### 3.3 Backtesting portfolio VaR

In this section, we will apply the corrected backtesting procedures robust to estimation risk presented in section 2 to the above setting. Theorem 1 and 3 allow us to quantify estimation risk for the unconditional and independence backtests such that we could carry out valid inferences. We only consider the fixed forecasting schemes. The estimation risk term for the unconditional tests with fixed forecasting scheme takes the form

$$\begin{aligned} ER_u &= E \left[ \sqrt{w_t'\Sigma_t w_t} \right] f_{Y_t}(m_\alpha(t-1))\sqrt{\pi}\sqrt{R} \left( Q_{v,R}^{-1}(\alpha) - Q_v^{-1}(\alpha) \right) \\ &+ E \left[ w_t' \frac{\partial \mu_t}{\partial \theta_0}' + \frac{(\text{vec}(w_t w_t'))' Q_v^{-1}(\alpha)}{2\sqrt{w_t'\Sigma_t w_t}} \frac{\partial \text{vec}\Sigma_t}{\partial \theta_0}' \right] f_{Y_t}(m_\alpha(t-1))\sqrt{\pi}\sqrt{R} \left( \hat{\theta}_R - \theta_0 \right) \\ &+ E \left[ \left( \mu_t' + \frac{w_t'\Sigma_t Q_v^{-1}(\alpha)}{\sqrt{w_t'\Sigma_t w_t}} \right) \frac{\partial w_t}{\partial \theta_0}' \right] f_{Y_t}(m_\alpha(t-1))\sqrt{\pi}\sqrt{R} \left( \hat{\theta}_R - \theta_0 \right) \\ &+ E \left[ \int^{m_\alpha(w^*, \theta_0, I_{t-1})} [f_{Y_{t,1}}(y_t, \theta_0, I_{t-1}) \frac{\partial Y_t(\theta_0)}{\partial \theta_0'} + f'_{Y_{t,2}}(y_t, \theta_0, I_{t-1})] dy_t \right] \sqrt{\pi}\sqrt{R} \left( \hat{\theta}_R - \theta_0 \right) \end{aligned}$$

where  $Q_{v,R}^{-1}(\alpha)$  is an  $\alpha$ -quantile estimator of the innovation distribution. Similarly, the estimation risk for the joint test is

$$\begin{aligned}
ER_c &= E \left[ \sqrt{w_t' \Sigma_t w_t} \{h_{t-j, \alpha}(\theta_0) + \alpha\} \right] f(m_\alpha(t-1)) \sqrt{\pi} \sqrt{R} \left( Q_{v,R}^{-1}(\alpha) - Q_v^{-1}(\alpha) \right) \\
&+ E \left[ \left( w_t' \frac{\partial \mu_t}{\partial \theta_0'} + \frac{(\text{vec}(w_t w_t'))' Q_v^{-1}(\alpha) \partial \text{vec} \Sigma_t}{2 \sqrt{w_t' \Sigma_t w_t}} \frac{\partial \text{vec} \Sigma_t}{\partial \theta_0'} \right) \{h_{t-j, \alpha}(\theta_0) + \alpha\} \right] f(m_\alpha(t-1)) \sqrt{\pi} \sqrt{R} \left( \hat{\theta}_R - \theta_0 \right) \\
&+ E \left[ \left( \mu_t' + \frac{w_t' \Sigma_t Q_v^{-1}(\alpha)}{\sqrt{w_t' \Sigma_t w_t}} \right) \frac{\partial w_t}{\partial \theta_0'} \{h_{t-j, \alpha}(\theta_0) + \alpha\} \right] f(m_\alpha(t-1)) \sqrt{\pi} \sqrt{R} \left( \hat{\theta}_R - \theta_0 \right) \\
&+ E \left[ \int^{m_\alpha(w^*, \theta_0, I_{t-1})} [f_{Y_{t,1}}(y_t, \theta_0, I_{t-1}) \frac{\partial Y_t(\theta_0)}{\partial \theta_0'} + f'_{Y_{t,2}}(y_t, \theta_0, I_{t-1})] dy_t \{h_{t-j, \alpha}(\theta_0) + \alpha\} \right] \sqrt{\pi} \sqrt{R} \left( \hat{\theta}_R - \theta_0 \right)
\end{aligned}$$

## 4 Simulation exercise

In this section, we illustrate our theoretical findings through a BEKK model.

## 5 Conclusion

This paper proposes the general unconditional and independence backtesting procedures robust to estimation risk for portfolio VaR. It extends the theory of quantifying estimation risk in backtesting procedures in EO (2007) from the univariate case to a multivariate case. We also apply the general theory to a very general location-scale parametric dynamic model with innovations following a very flexible class of distribution, the GH distribution. The simulation exercise on a simple example also supports our theoretical findings. In practice, our correction to the standard backtesting procedures for portfolio VaR is of great importance. Neglecting the effect of estimation risk on those procedures may result in the backtesting procedures being invalid. The future work will be an application to real financial data to see how estimation risk affects the decisions in inferences.

## Appendix A: Mathematical Proofs

First of all, we show how to get the conditional distribution of  $Y_t|I_{t-1}$  from the linear transformation  $Y_t = w_t^* r_t$ , where  $w_t^*$  is treated as constant, and the multivariate conditional distribution of  $r_t|I_{t-1}$ . Construct a one-to-one mapping between  $r_t$  and a constructed vector  $Z_t$  with the first element being  $Y_t = w_t^* r_t$ . The constructed vector is

$$Z_t = \begin{bmatrix} Y_t \\ r_{2t} \\ \vdots \\ r_{dt} \end{bmatrix} = \begin{bmatrix} w_t^* r_t \\ r_{2t} \\ \vdots \\ r_{dt} \end{bmatrix} = \underbrace{\begin{bmatrix} w_{1t}^* & w_{2t}^* & \cdots & w_{dt}^* \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{J(\theta_0)} \begin{bmatrix} r_{1t} \\ r_{2t} \\ \vdots \\ r_{dt} \end{bmatrix} = J(\theta_0)r_t,$$

where  $J(\theta_0)$  is a positive definite  $d \times d$  matrix. So  $r_t = J^{-1}(\theta_0)Z_t$ , where

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

Since we specify  $r_t|I_{t-1} \sim F_{r_t}(\cdot, \theta_0, I_{t-1})$ , then the multivariate conditional distribution of  $Z_t|I_{t-1}$  as follows

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

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 over the other elements

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

where  $Y_t = w_t^* r_t$  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  $w_t^*$  will depend on the distribution parameter  $\theta_0$  and the information set

$I_{t-1}$ ,  $w_t^* \equiv w^*(\theta_0) = w^*(\theta_0, I_{t-1})$ . Notice that  $Y_t \equiv Y_t(\theta_0)$  will add an extra term in the derivative of  $f_{Y_t}(y_t, \theta_0, I_{t-1})$  with respect to  $\theta_0$ . Namely,

$$\frac{\partial f_{Y_t}(y_t, \theta_0, I_{t-1})}{\partial \theta_0} = f_{Y_{t,1}}(y_t, \theta_0, I_{t-1}) \frac{\partial Y_t(\theta_0)}{\partial \theta_0} + f_{Y_{t,2}}(y_t, \theta_0, I_{t-1})$$

where  $f_{Y_{t,j}}(Y_t, \theta_0, I_{t-1})$  is the derivative of  $f_{Y_t}(Y_t, \theta_0, I_{t-1})$  with respect to the  $j^{\text{th}}$  argument.

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

Define the process

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

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 (2007b).

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

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

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

$$|K_n(\widehat{c}) - K_n(0)| = o_P(1). \tag{10}$$

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

$$\max_{R \leq t \leq n} \sqrt{t}(\widehat{\theta}_t - \theta_0) = O_P(1) \tag{11}$$

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

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

which implies the decomposition

$$\begin{aligned} \frac{1}{\sqrt{P}} \sum_{t=R+1}^n (h_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha) &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [h_{t,\alpha}(\theta_0) - F_{Y_t}(\theta_0)] \\ &\quad + \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{Y_t}(\hat{\theta}_{t-1}) - F_{Y_t}(\theta_0)] \\ &\quad + \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{Y_t}(\theta_0) - \alpha] + o_P(1). \end{aligned} \tag{12}$$

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

$$\begin{aligned} A_{1n} &:= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{Y_t}(\hat{\theta}_{t-1}) - E[F_{Y_t}(\hat{\theta}_{t-1})] - F_{Y_t}(\theta_0) + E[F_{Y_t}(\theta_0)]] \\ &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \left( \frac{\partial F_{Y_t}(\tilde{\theta}_{t-1})}{\partial \tilde{\theta}_{t-1}} - E\left[\frac{\partial F_{Y_t}(\tilde{\theta}_{t-1})}{\partial \tilde{\theta}_{t-1}}\right] \right) (\hat{\theta}_{t-1} - \theta_0), \end{aligned}$$

where  $\tilde{\theta}_{t-1}$  is between  $\hat{\theta}_{t-1}$  and  $\theta_0$ , and the derivative of  $F_{Y_t}(\theta)$  with respect to  $\theta'$  is

$$\begin{aligned}
\frac{\partial F_{Y_t}(\theta)}{\partial \theta'} &= \frac{\partial}{\partial \theta'} \int^{m_\alpha(w^*, \theta, I_{t-1})} f_{Y_t}(y_t, \theta, I_{t-1}) dy_t \\
&= g'_\alpha(w^*, \theta, I_{t-1}) f_{Y_t}(\theta) + \int^{m_\alpha(w^*, \theta, I_{t-1})} \frac{\partial f_{Y_t}(y_t, \theta, I_{t-1})}{\partial \theta'} dy_t \\
&= [m'_{\alpha,1}(w^*, \theta, I_{t-1}) \frac{\partial w^*(\theta)}{\partial \theta} + m'_{\alpha,2}(w^*, \theta, I_{t-1})] f_{Y_t}(\theta) \\
&\quad + \int^{m_\alpha(w^*, \theta, I_{t-1})} [f_{Y_{t,1}}(y_t, \theta, I_{t-1}) \frac{\partial Y_t(\theta)}{\partial \theta'} + f'_{Y_{t,1}}(y_t, \theta, I_{t-1})] dy_t
\end{aligned}$$

Note that A2 and A3 imply that  $E \left[ \sup_{\theta \in \Theta_0} \left| \frac{\partial F_{Y_t}(\theta)}{\partial \theta'} \right| \right] < C$ . Hence, by the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2) and (11), then  $A_{1n} = o_P(1)$  holds. Similarly,

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

Now, by the ULLN and (11), then  $B_{2n} = o_P(1)$  holds. Hence,

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

The theorem follows from (12) 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.  $\square$

PROOF OF COROLLARY 2: The consistency of  $\hat{\rho}$  and  $\hat{V}$  follows from the ULLN of Jennrich (1969, Theorem 2) and (11). Giacomini and Komunjer (2005), on the other hand, proved

the consistency of the out-of-sample version of  $A_\tau$ . It also follows in this context that  $\widehat{A}_\tau = A + o_P(1)$ . Now, by Slutsky's Lemma the corollary is proved.  $\square$

PROOF OF THEOREM 2: The proof is similar to that of Theorem 1. Define the process

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

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

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

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

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

Whereas the arguments after (12) imply that

$$\left| C_{2n} - E \left[ \frac{\partial F_{Y_t}(\theta_0)}{\partial \theta'_0} h_{t-j,\alpha}(\theta_0) \right] \frac{1}{\sqrt{P-j}} \sum_{t=R+1+j}^n H(t-j-1) \right| = o_P(1).$$

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

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

and similarly, define  $\xi_{1n,j}$  and  $\xi_{2n,j}$  with  $\theta_0$  replacing  $\widehat{\theta}_{t-1}$ . Now, simple algebra shows that

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

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

$$\sqrt{P-j}(\widehat{\gamma}_{P,j} - \gamma_{P,j}) = \left(\widehat{\xi}_{P,j} - \xi_{P,j}\right) - \alpha\left(\widehat{\xi}_{1n,j} - \xi_{1n,j}\right) - \alpha\left(\widehat{\xi}_{2n,j} - \xi_{2n,j}\right). \quad (13)$$

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

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

The latter display, part i) and (13) prove condition ii).

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

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

Details are omitted to save space.  $\square$

**PROOF OF COROLLARY 3:** The consistency of  $\widehat{\eta}$  and  $\widehat{V}$  follows from the ULLN of Jennrich (1969, Theorem 2) and (11). The consistency of  $\widehat{B}_\tau$  follows from Giacomini and Komunjer (2005). Now, by Slutsky's Lemma the corollary follows.  $\square$

## References

- Abramowitz, M and I.A. Stegun (1965): Handbook of Mathematical Functions. New York: Dover.
- Alexander J. McNeil, Rudiger Frey and Paul Embrechts (2005): Quantitative Risk Management, Princeton University Press, Princeton and Oxford
- Christian Gouriéroux and Joann Jasiak (2002): Value-at-Risk, forthcoming Handbook of Financial Econometrics, ed. Ait-Sahalia, Hansen.
- E.W. Rengifo and J.V.K. Rombouts (2004): Dynamic Optimal Portfolio Selection in a VaR Framework
- F.Javier Mencia and Enrique Sentana (2005): Estimation and Testing of Dynamic models with generalized Hyperbolic innovations
- F.Javier Mencia and Enrique Sentana (2008): Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation
- Gustavo M. de Athayde, Renato G. Flores Jr. (2004): Finding a maximum skewness portfolio- a general solution to three-moments portfolio choice, Journal of Economic Dynamic and Control 28(2004)1335-1352
- J. Carlos Escanciano and Jose Olmo (2007a): Backtesting Value-at-Risk with Estimation Risk, forthcoming in...
- J. Carlos Escanciano and Jose Olmo (2007b): Backtesting VaR with Estimation Risk, working paper.
- Jeroen V.K. Rombouts and Marno Verbeek (2004): Evaluating Portfolio Value-at-Risk using Semi-Parametric GARCH Models
- Jorion, P. (2001): Value-at-Risk: the New Benchmark for Controlling Market Risk. McGraw-Hill, New York.
- Luc Bauwens and Sebastien Laurent (2004): A new class of multivariate skew density, with application to GARCH models

Luc Bauwens, Sebastien Laurent and Jeroen V.K. Rombouts (2006): Multivariate GARCH models: A survey, *Journal of Applied Economics*, 21: 79-109 (2006)

Martin Hellmich and Stefan Kassberger (2007): Efficient portfolio optimization in the multivariate Generalized Hyperbolic framework

Michael W. Brandt (2004): Portfolio Choice Problems

Pierre Giot and Sebastien Laurent (2003): Value-at-Risk for long and short trading positions, *Journal of Applied Economics* 18: 641-664

Wei-han Liu and Douglas J. Miller: A Study on Portfolio Value-at-Risk