

This is a postprint version of the following published document:

Escanciano, J. C., & Pei, P. (2012). Pitfalls in backtesting Historical Simulation VaR models. *Journal of Banking & Finance*, 36 (8), pp. 2233-2244.

DOI:[10.1016/j.jbankfin.2012.04.004](https://doi.org/10.1016/j.jbankfin.2012.04.004)

© Elsevier, 2012



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Pitfalls in backtesting Historical Simulation VaR models

Juan Carlos Escanciano^{a,1,*}, Pei Pei^{a,b}

^a*Department of Economics, Indiana University, 105 Wylie Hall, 100 South Woodlawn Avenue, Bloomington, IN 47405-7104, USA*

^b*Chinese Academy of Finance and Development, Central University of Finance and Economics, 39 South College Road, Beijing, 100081, China*

Abstract

Historical Simulation (HS) and its variant, the Filtered Historical Simulation (FHS), are the most popular Value-at-Risk forecast methods at commercial banks. These forecast methods are traditionally evaluated by means of the unconditional backtest. This paper formally shows that the unconditional backtest is always inconsistent for backtesting HS and FHS models, with a power function that can be even smaller than the nominal level in large samples. Our findings have fundamental implications in the determination of market risk capital requirements, and also explain Monte Carlo and empirical findings in previous studies. We also propose a data-driven weighted backtest with good power properties to evaluate HS and FHS forecasts. A Monte Carlo study and an empirical application with three U.S. stocks confirm our theoretical findings. The empirical application shows that multiplication factors computed under the current regulatory framework are downward biased, as they inherit the inconsistency of the unconditional backtest.

JEL classification: C52; C32; G21; G32

Keywords: Backtesting; Basel Accord; Risk management; Value at Risk; Conditional quantile; Market Risk Capital requirements

*Corresponding author. Tel.: +1 812 855 7925; fax: +1 812 855 3736.

E-mail addresses: jescanci@indiana.edu (Juan Carlos Escanciano), ppei@umail.iu.edu (Pei Pei)

¹Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

1. Introduction

Value-at-Risk (VaR) has become the standard tool for measuring market risk used by financial institutions and their regulators. VaR is defined as the maximum expected loss on an investment over a specified horizon at a particular confidence level; see Jorion (2006). VaR summarizes market risk into a single number which, in statistical terms, is a quantile of the conditional distribution of portfolio returns given investor's information set. Ever since the 1996 Amendment of the Basel Accord (see Basel Committee on Banking Supervision, 1996a and 1996b), VaR not only offers a measure of market risk, but it also forms the basis for the determination of market risk capital. Obviously, the foremost concern of VaR users is to forecast VaRs accurately in order to guide their decisions. According to a recent international survey by Pérignon and Smith (2010), the Historical Simulation (HS) forecast model and its variant, the Filtered HS (FHS), are the most currently used methods at commercial banks.² These methods use *unconditional* quantiles of raw data and standardized innovations, respectively, to forecast conditional quantiles of portfolio returns.

Within the framework of the Basel Accord, besides the VaR forecasts, other crucial ingredients in determining market risk capital requirements are the forecast evaluation or *backtesting* procedure and the multiplication factors based on the outcomes of the backtesting procedure.³ Therefore, backtesting, as a forecast evaluation method of a VaR model, is of vital importance for financial institutions and their regulators. There are several backtesting methods available. Among the most popular methods are the unconditional coverage test of Kupiec (1995, henceforth K-test), the conditional backtest of Christoffersen (1998, henceforth C-test), the CAViaR test of Engle and Manganelli (2004), the duration-based approach of Christoffersen and Pelletier (2004) and the density forecast evaluation approach of Berkowitz (2001). For an excellent review of these methods see Berkowitz, Christoffersen and Pelletier (2011). Arguably, the unconditional coverage test is the most widely employed backtesting procedure given its simplicity and also the one implicitly

²These authors find that 73% of the banks that disclose their VaR methodology in their 2005 annual reports use HS or related techniques.

³Currently, stressed VaR is also incorporated into the calculation of capital requirements, see Basel Committee on Banking Supervision (2011).

incorporated in the Basel Accord for determining market risk capital requirements (see Basel Committee on Banking Supervision, 1996b). It simply computes the frequency of the ex-post trading outcomes that are less than or equal to the ex-ante VaR forecasts, and compares this frequency with the specified VaR level, e.g., 1%. Large differences between the empirical frequency and the theoretical VaR level indicates that, on average, the forecasts are not optimal.⁴

This paper draws on the observation that the most widely used VaR models, HS or FHS models, are commonly evaluated with the K-test, and it formally shows that this combination is highly misleading. We prove that the K-test is *always* inconsistent in detecting non-optimal HS or FHS forecasts. That is, the rejection probability of the K-test does not converge to one as the sample size increases for non-optimal forecasts. This theoretical finding explains the low power associated with the K-test in simulations carried out in the literature, see e.g., Kupiec (1995), Pritsker (2006) and Pérignon and Smith (2008), among many others. For instance, Pritsker (2006) provides Monte Carlo evidence on the low power of the K-test for detecting lack of optimality of HS models in generalized autoregressive conditional heteroscedastic (GARCH) models. Based on simulations, he concludes that the K-test is not useful in detecting some deficiencies of HS models.⁵ We remark that we do not address in this paper the evaluation of HS or FHS VaR models, which Pritsker (2006) has already investigated, but rather the evaluation of the backtesting methods commonly used for assessing optimality of HS and FHS forecasts. Our results do provide a theoretical explanation for the aforementioned Monte Carlo findings.

As an alternative to the K-test we propose a new data-driven weighted backtest with good power properties for detecting lack of optimality of HS and FHS models. The new test statistic combines a well-known test, the independence test of Christoffersen (1998), with an adaptive backtest using a nonparametric weight. This combination results in a test with good power performance as demonstrated in our Monte Carlo simulations and empirical application, since the two

⁴As is standard in the VaR literature, optimality is with respect to the “tick” loss function; see next section for a formal definition of optimal forecast.

⁵Pritsker (2006) shows theoretically that HS models are both under-responsive to changes in conditional risk, and respond to changes in risk in an asymmetric fashion.

test statistics provide asymptotically independent information. The proposed test statistic is our preferred choice within a general class of weighted backtests discussed in the paper. We also show that this class contains many of the existing traditional backtests, and that within this class, our preferred choice possesses some optimal power properties for assessing optimality of HS and FHS forecasts.

As mentioned earlier, an important implication of backtesting results is the determination of the multiplication factors for market risk capital requirements. Our theoretical results imply that computing the multiplication factors based on the K-test, as suggested by the current regulatory framework, is highly misleading for HS and FHS models, because the lack of power of the K-test will result in underestimated multiplication factors, and then, in likely insufficient market risk capital reserves. An empirical application to U.S. stocks highlights this important implication of our theoretical results. To the best of our knowledge, this is the first paper that shows theoretically the important limitation of backtesting methods for HS and FHS models, and their implications on market risk capital requirements.

The remainder of the paper is organized as follows. Section 2 describes the forecast evaluation problem from a decision point of view and shows the inconsistency of the K-test for HS and FHS models. In Section 3 we propose a large class of weighted backtesting methods and a data-driven test within the class, which we term D-test, and we establish the asymptotic distribution theory for them. The finite-sample performances of the K-test, C-test and D-test for backtesting HS and FHS models are investigated using Monte Carlo methods in Section 4. Section 5 analyzes, in an application to several portfolios of U.S stock data, the testing results for the optimality of HS and FHS forecasts and the corresponding implications on the determination of the multiplication factors for market risk capital requirements. Our application illustrates the lack of power of the commonly used backtesting methods and the ability of the new backtesting method to uncover suboptimal VaR forecasts from HS and FHS models. Finally, section 6 concludes. Mathematical proofs are gathered in the Appendices at the end of the paper.

2. Backtesting Historical Simulation models

In this section, we introduce the forecast evaluation problem for VaR models, and we prove the inconsistency of the K-test in testing optimality of HS and FHS forecasts. We denote the real-valued time series of portfolio returns or Profit and Loss (P&L) account by Y_t , and assume that at time $t - 1$ the bank's information set is given by I_{t-1} , which contains past values of Y_t and other relevant economic and financial variables. We assume that VaR users face the following decision problem: with the information available at time $t - 1$, they aim to find a measurable function $q^*(I_{t-1})$ minimizing the conditional expected asymmetric or “tick” loss function, that is

$$q^*(I_{t-1}) := \arg \min_q E [L_\alpha(Y_t, q(I_{t-1})) \mid I_{t-1}], \quad (1)$$

where $L_\alpha(Y_t, q(I_{t-1})) := \{\alpha - 1(Y_t \leq q(I_{t-1}))\} (Y_t - q(I_{t-1}))$ and $1(A)$ is the indicator function of the event A , that is, $1(A) = 1$ if A occurs and 0 otherwise. The symbol $:=$ denotes definition. Henceforth, an optimal forecast according to this loss function is simply called optimal (see e.g., Christoffersen, 1998).

It is well-known that, provided the conditional distribution of Y_t given I_{t-1} is continuous, the optimal forecast $q^*(I_{t-1})$ in (1) is the α -th conditional quantile (i.e., VaR) of Y_t given I_{t-1} , say $q_\alpha(I_{t-1})$, which is implicitly defined by the restriction

$$P(Y_t \leq q_\alpha(I_{t-1}) \mid I_{t-1}) = \alpha, \text{ almost surely (a.s.), } \alpha \in (0, 1), \forall t \in \mathbb{Z}. \quad (2)$$

A common approach in empirical work is to assume a *parametric* VaR model for $q_\alpha(I_{t-1})$, say $\mathcal{M} = \{m_\alpha(I_{t-1}, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$, and to proceed to make VaR forecasts using the model \mathcal{M} . In parametric models the unknown parameter θ_0 belongs to a suitable parameter space $\Theta \subset \mathbb{R}^p$. The literature on parametric VaR modeling has been advancing rapidly, see e.g., Engle and Manganelli (2004), Koenker and Xiao (2006) and Gouriéroux and Jasiak (2008), among many others.⁶

⁶The reader should not be confused with a fully parametric approach in which the whole conditional distribution

The preferred model by financial institutions is the HS model or its variant, the FHS, see Barone-Adesi et al. (1998, 1999) and Barone-Adesi and Giannopoulos (2001). The HS model specifies a time-constant VaR model

$$m_\alpha(I_{t-1}, \theta_0) = \theta_0 \equiv F_Y^{-1}(\alpha), \quad (3)$$

where, henceforth, for a generic strictly stationary sequence $\{Z_t\}_{t \in \mathbb{Z}}$, $F_Z^{-1}(\alpha)$ denotes the unconditional quantile function of Z_t evaluated at α , i.e., $F_Z^{-1}(\alpha) := \inf\{z \in \mathbb{R} : F_Z(z) \geq \alpha\}$, where $F_Z(\cdot)$ is the cumulative distribution function (cdf) of Z_t . The main assumption underlying HS is the time-invariant conditional variance of returns, which is an unrealistic assumption and always empirically rejected.

In contrast, the FHS assumes an underlying time-variant location-scale structure, and hence it specifies a time-varying VaR model

$$m_\alpha(I_{t-1}, \theta_0) = \mu(I_{t-1}, \beta_0) + \sigma(I_{t-1}, \beta_0)F_\varepsilon^{-1}(\alpha), \quad \theta_0 = (\beta_0, F_\varepsilon^{-1}(\alpha)), \quad (4)$$

where $\mu_t \equiv \mu(I_{t-1}, \beta_0)$ and $\sigma_t \equiv \sigma(I_{t-1}, \beta_0)$ are the conditional mean and standard deviation of Y_t given I_{t-1} , respectively, and the standardized innovations $\{\varepsilon_t := (Y_t - \mu_t)/\sigma_t\}_{t \in \mathbb{Z}}$ are assumed to be independent and identically distributed (i.i.d.), and independent of I_{t-1} . The unknown parameter θ_0 comprises β_0 and $F_\varepsilon^{-1}(\alpha)$, which are the conditional mean and standard deviation model parameter and the α -th unconditional quantile of the standardized innovations, respectively. Among the most common models for μ_t and σ_t are the autoregressive moving average (ARMA) and GARCH models, respectively, see Berkowitz and O'Brien (2002) for motivation of this specification in financial risk management.⁷ For a review of ARMA and GARCH models see e.g., Li,

of Y_t given I_{t-1} is specified. We do not need such full parametric specification in VaR analysis, since our concern is just on the conditional quantile function at a particular quantile level α . In this sense and in the statistical jargon the specified conditional distribution is semiparametric.

⁷For example, the AR(1)-GARCH(1,1) model is one of the popular specifications for financial data, where $\mu(I_{t-1}, \beta_0) = \mu_{00}Y_{t-1}$ and $\sigma(I_{t-1}, \beta_0) = \omega_{00} + \omega_{01}(Y_{t-1} - \mu_{00}Y_{t-2})^2 + \omega_{02}\sigma^2(I_{t-2}, \beta_0)$ with $\beta_0 = (\mu_{00}, \omega_{00}, \omega_{01}, \omega_{02})'$, $|\mu_{00}| < 1$, $\omega_{00} > 0$, $\omega_{01} \geq 0$, $\omega_{02} \geq 0$ and $\omega_{01} + \omega_{02} < 1$.

Ling and McAleer (2002).

Once a VaR model has been specified for forecasting, a natural testing problem is to evaluate the optimality of the resulting VaR forecasts. This is the so-called backtesting procedure for VaR models. That is, we aim to test for the null hypothesis

$$H_0 : m_\alpha(I_{t-1}, \theta_0) \text{ is an optimal forecast,}$$

against the alternative hypothesis, say H_A , that $m_\alpha(I_{t-1}, \theta_0)$ is not optimal. From our previous discussion, the null entails that $m_\alpha(I_{t-1}, \theta_0)$ is a correctly specified VaR model, $m_\alpha(I_{t-1}, \theta_0) = q_\alpha(I_{t-1})$ a.s.. In this paper we are concerned with tests for H_0 when $m_\alpha(I_{t-1}, \theta_0)$ is given by a HS or FHS model in (3) or (4), respectively.

A necessary condition for an optimal forecast is that it is unbiased, i.e., (2) implies that

$$E[h_{t,\alpha}(\theta_0)] = \alpha,$$

where $h_{t,\alpha}(\theta_0) := 1(Y_t \leq m_\alpha(I_{t-1}, \theta_0))$ is the so-called *hit*. This implication suggests a test for H_0 based on rejecting for large absolute values of

$$K_P \equiv K(P, R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [h_{t,\alpha}(\theta_0) - \alpha],$$

where P and R are the out-of-sample and in-sample sample sizes, respectively, and $n := P + R$. This is the so-called unconditional backtest (Kupiec, 1995), in short K-test. The K-test is the reference backtest in the banking industry and the academia. Escanciano and Olmo (2010) has extended the K-test to a situation where the parameter θ_0 is unknown and estimated. Our main result formally shows that the K-test and its generalization to estimated parameter are inconsistent in detecting non-optimal forecasts from HS or FHS models. Intuitively, the test statistic K_P will remain “small” under the lack of forecast optimality for HS or FHS models, as the corresponding hit sequence has a mean equal to α under the alternative hypothesis. This intuition is formalized

below in our Theorem 1.

Under appropriate regularity conditions and the null hypothesis, $K_P/\sqrt{\alpha(1-\alpha)}$ converges to a standard normal random variable, see Kupiec (1995). The asymptotic power function of Kupiec's two sided test at $\tau\%$ nominal level is then defined as

$$\Pi(\tau, \alpha) := \lim_{P \rightarrow \infty} \Pr \left(\left| \frac{K_P}{\sqrt{\alpha(1-\alpha)}} \right| > z_{1-\tau/2} \mid H_A \right),$$

where z_τ is the corresponding τ -quantile from the standard normal distribution. That is, $\Pi(\tau, \alpha)$ is the limit probability of rejecting optimality of HS or FHS forecasts when in fact they are not optimal. It will be shown that the asymptotic power function depends on the long-run variance

$$\begin{aligned} \sigma_K^2 &:= \alpha(1-\alpha) + 2 \sum_{j=1}^{\infty} Cov(h_{t,\alpha}(\theta_0), h_{t-j,\alpha}(\theta_0)) \\ &:= \alpha(1-\alpha) + 2\sigma_{hK}, \end{aligned}$$

which is finite under Assumption A1 in Appendix A. The following Theorem shows the inconsistency of the K-test. Let Φ be the cdf of the standard normal random variable.

THEOREM 1: *Assume that $m_\alpha(I_{t-1}, \theta_0)$ is given by a HS or FHS model, and let Assumption A1 in Appendix A hold. Then,*

$$\Pi(\tau, \alpha) = 2\Phi \left(z_{\tau/2} \frac{\sqrt{\alpha(1-\alpha)}}{\sigma_K} \right).$$

A consequence of Theorem 1 is that the K-test is inconsistent, that is, $\Pi(\tau, \alpha) < 1$. The magnitude of the power function depends on the relative size of the long-run variance, σ_K^2 , and the asymptotic null variance, $\alpha(1-\alpha)$. If $\sigma_K^2 < \alpha(1-\alpha)$, or equivalently $\sigma_{hK} < 0$, the power function is smaller than the nominal level τ , otherwise it is larger than τ . Empirically, $\sigma_{hK} > 0$ is more consistent with actual financial data due to the well-known property of clustering of extremes. Therefore, our theoretical results suggest that the K-test will have low power, but still larger than the nominal level for large sample sizes. Also, note that the larger σ_{hK} , the higher the power of the K-test.

Thus, the inconsistency of the K-test is less problematic in periods of high persistence in the (left) tail dependence (e.g. during a financial crisis).

Theorem 1 is proved under a mixing condition, see Assumption A1 in Appendix A. We remark, however, that any other weak dependence concept leading to a Central Limit Theorem (CLT) for the K-test under the alternative hypothesis could be used here. Particularly in the extreme case of $\sigma_{hK} = \infty$, the CLT does not hold (at the usual rate of \sqrt{P}), and the K-test diverges to infinity under the alternative hypothesis. Thus, the assumption $\sigma_{hK} < \infty$ is necessary and sufficient for the inconsistency of the K-test. Strictly speaking, Theorem 1 is not applicable when $\sigma_{hK} = \infty$, but it still provides useful information and explains simulations results obtained with models such as the JP Morgan's (1996) RiskMetric model, see our simulations below.

Theorem 1 sheds light on simulations and applications carried out in the literature. We remark on some conclusions that the literature has drawn from simulations. First, our Theorem 1 shows that the low power of the K-test cannot be in general improved by considering small VaR confidence levels $1 - \alpha$, large sample sizes P or different approximations to its finite-sample distributions, e.g., bootstrap. Second, the lack of power affects both HS and FHS models. Third, Theorem 1 also holds when θ_0 is estimated. Most Monte Carlo studies in the literature have evaluated the performance of the classical K -test in a setting where θ_0 is estimated and P/R is moderate or large. As shown in Escanciano and Olmo (2010), these existing Monte Carlo results have to be taken with caution, since the asymptotic variance of $K_P/\sqrt{\alpha(1-\alpha)}$ evaluated at the estimated θ_0 is not one unless P/R converges to zero. Hence, it is of fundamental importance to control the Type-I error, i.e., the probability of rejecting an optimal forecast, before evaluating the power performance of the K-test. This is done by replacing the critical value $z_{1-\tau/2}$ by an appropriate one, so that under the null the rejection probability of the test is τ , see Escanciano and Olmo (2010) for details. With the suitable choice of the critical value, the same conclusion of Theorem 1 holds when θ_0 is estimated.

3. A class of weighted backtesting procedures

Motivated by the lack of power of the K-test, we propose a general class of weighted backtesting procedures based on the test statistic

$$K_{P,w} := \frac{1}{\sqrt{\hat{B}_w P}} \sum_{t=R+1}^n w(I_{t-1})(h_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha), \quad (5)$$

where $\hat{B}_w := P^{-1} \sum_{t=R+1}^n w^2(I_{t-1})$ is a consistent estimate of $B_w := E[w^2(I_{t-1})]$, $\hat{\theta}_{t-1}$ is an estimator for θ_0 , see examples below, and $w(I_{t-1})$ is a bounded, non-zero weight function, whose choice will be based on power considerations of the resulting test, see our discussion after Theorem 2. First, we will develop the asymptotic theory of $K_{P,w}$ for a generic w . Large absolute values of $K_{P,w}$ will indicate non-optimality of the forecasts. Note that the K-test statistic corresponds to the choice $w(I_{t-1}) \equiv 1$.

For the sake of completeness we discuss the asymptotic distribution theory under three different forecasting schemes, namely, the recursive, rolling and fixed forecasting schemes. They differ in how the parameter θ_0 of the HS or FHS model is estimated. In the recursive scheme, the estimator $\hat{\theta}_t$ is computed with all the sample available up to time t . In the rolling scheme, only the last R values of the series are used to estimate $\hat{\theta}_t$, that is, $\hat{\theta}_t$ is constructed from the sample at periods $s = t - R + 1, \dots, t$. Finally, in the fixed scheme, the parameter is not updated when new observations become available, i.e., $\hat{\theta}_t = \hat{\theta}_R$, for all $t, R \leq t \leq n$.

Examples of estimate $\hat{\theta}_t$ for HS and FHS models are as follows. In the HS model, the natural estimate of θ_0 is the empirical quantile estimator $\hat{\theta}_t = F_{1,t-1,Y}^{-1}(\alpha)$, $\hat{\theta}_t = F_{t-R,t-1,Y}^{-1}(\alpha)$ and $\hat{\theta}_t = F_{1,R,Y}^{-1}(\alpha)$ for the recursive, rolling and fixed forecasting schemes, respectively, where $F_{r,s,Y}^{-1}(\alpha)$ denotes the α -th empirical quantile function of $\{Y_t\}_{t=r}^s$, $r < s$, i.e., $F_{r,s,Y}^{-1}(\alpha) := \inf\{z \in \mathbb{R} : F_{r,s,Y}(z) \geq \alpha\}$ with the empirical cdf $F_{r,s,Y}(z) := (s - r + 1)^{-1} \sum_{t=r}^s 1\{Y_t \leq z\}$. For example, if $R = 100$ and $\alpha = 5\%$, then $\hat{\theta}_t$ under the fixed forecasting scheme is simply the 5th order statistic of the sample $\{Y_t\}_{t=1}^R$.

In the FHS model, the parameter β_0 is usually estimated by the Quasi-Maximum Likelihood

Estimator (QMLE) that is routinely used for estimating GARCH models, say $\widehat{\beta}_t$, see (A.2), and estimators for $F_\varepsilon^{-1}(\alpha)$ are constructed as for the HS but replacing $\{Y_t\}_{t=r}^s$ by residuals $\{\widehat{\varepsilon}_t\}_{t=r}^s$, $r < s$, where

$$\widehat{\varepsilon}_t := \frac{Y_t - \mu(I_{t-1}, \widehat{\beta}_t)}{\sigma(I_{t-1}, \widehat{\beta}_t)}, \quad t = 1, \dots, n.$$

Now we establish the asymptotic theory of $K_{P,w}$ for HS and FHS models.

THEOREM 2: *Let Assumptions A1-A5 in Appendix A hold. Then, for HS and FHS models:*

- (i) *Under H_0 , $K_{P,w} \xrightarrow{d} N(0, \alpha(1 - \alpha))$.*
- (ii) *Under H_A , $P^{-1/2}K_{P,w} \xrightarrow{P} E[w(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)]/\sqrt{B_w}$.*

Theorem 2 establishes the asymptotic null distribution of $K_{P,w}$ and also shows the probability limit of $P^{-1/2}K_{P,w}$ under H_A for both HS and FHS models. Notice that under H_A , if $E[w(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)] \neq 0$, $K_{P,w}$ will diverge to $\pm\infty$, and the test based on $K_{P,w}$ will be consistent. On the contrary, if $E[w(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)] = 0$, then the test based on $K_{P,w}$ will be inconsistent, and an analogous result to Theorem 1 will hold.

Assumptions A1-A5 required in Theorem 2 are standard in the financial econometrics literature, see Appendix A for detailed explanations. Here we particularly remark on Assumption A4, which requires that $\pi := \lim_{n \rightarrow \infty} P/R = 0$. This is an assumption systematically employed in the literature, see e.g., Engle and Manganelli (2004), with the only exception of Escanciano and Olmo (2010, 2011) that concentrate on $\pi > 0$. The motivation for the latter was that financial institutions often choose $R = P = 250$. However, there are three important reasons to maintain the assumption $\pi = 0$ in our context. First, with the assumption $\pi = 0$, an in-sample size R that is much larger than 250 could be used, which guarantees a good precision of the extreme quantile estimators used in VaR inferences. Note that long spans of in-sample historical data are available in practice for VaR analysis carried out on a daily basis. Second, it is convenient to assume $\pi = 0$, as it substantially simplifies the asymptotic null variance of the test (cf. Escanciano and Olmo, 2011). Therefore, from a practical point of view, assuming $\pi = 0$ is highly desirable for VaR inferences to be simple and meaningful at low quantile levels such as 5% or 1%. Our simulations

also show that a choice of π of 0.1, e.g. $(R, P) = (2500, 250)$, provides satisfactory finite sample performances of our proposed test. Finally, we stress that the choice of π is irrelevant to the main theoretical findings on the inconsistency of the K-test.

Now we investigate how to choose the weight function w in $K_{P,w}$. The choice of w can be based on computational convenience and/or power considerations of the resulting test. It is interesting to notice that the class of weighted backtests we propose includes as particular cases traditional tests such as the test of independence of Christoffersen (1998), and its extensions to higher order autocorrelations (see Berkowitz, Christoffersen and Pelletier, 2011) by allowing for multivariate weights. To see this, take

$$w(I_{t-1}) \equiv w_{aj}(I_{t-1}) := h_{t-j,\alpha}(\theta_0) - \alpha, \quad j \geq 1,$$

and note that

$$E[w_{aj}(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)] = Cov(h_{t-j,\alpha}(\theta_0), h_{t,\alpha}(\theta_0)), \quad j \geq 1,$$

which is the j -th order autocovariance of the hit sequence. Tests based on autocorrelations of $\{h_{t,\alpha}(\theta_0)\}$ are computationally convenient. The power properties of the resulting tests depend on whether $E[w_{aj}(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)]$ is zero or not under the alternative. Given the well-known clustering effect of financial data we expect tests based on autocorrelations to be powerful in detecting non-optimal forecasts of the HS model. However, for FHS,

$$E[w_{aj}(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)] = Cov(1(\varepsilon_{t-j} \leq F_\varepsilon^{-1}(\alpha)), 1(\varepsilon_t \leq F_\varepsilon^{-1}(\alpha))), \quad j \geq 1, \quad (6)$$

so in this case the consistency of tests based on autocorrelations will depend on the dependence in the pseudo-innovations (here ε_t is the probabilistic limit of $\varepsilon_t(\hat{\beta}_t)$ as $n \rightarrow \infty$) at the left tail. Our simulations suggest that (6) was “small” for various commonly used data generating processes, leading to low power for autocorrelation-based tests.

Therefore, the above theoretical results suggest that the choice of w has an influential effect on the power properties of the resulting test. A natural question arises: does there exist a choice of w that leads to a test with omnibus power properties? We address this question in the next section.

3.1. A powerful data-driven weighted backtest

In order to construct a powerful backtest, we must choose w such $E[w(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)]$ is non-zero under the alternative (see Theorem 2). The following Lemma shows that there exists a choice of w that guarantees consistency of the test based on $K_{P,w}$. Let $F_{I_{t-1}}(\cdot)$ denote the conditional distribution of Y_t given I_{t-1} .

LEMMA 1: *Let the assumptions of Theorem 2 hold. Then, with $w^*(I_{t-1}) := F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0))$, the test based on K_{P,w^*} is consistent.*

Lemma 1 shows that our class of weighted backtests includes a consistent test. However, $w^*(I_{t-1}) = F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0))$ is in general unknown, and its nonparametric estimation is generally complicated because w^* will in general depend on the infinite-dimensional information set I_{t-1} . If a nonparametric approach is entertained, we suggest to reduce the dimensionality of the problem and consider the choice

$$w_1^*(Y_{t-1}) = F_{Y_{t-1}}(m_\alpha(I_{t-1}, \theta_0)), \quad (7)$$

where $F_{Y_{t-1}}(\cdot)$ denotes the conditional distribution of Y_t given Y_{t-1} . The resulting test will be consistent against the class of alternatives where $F_{Y_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) \neq \alpha$. This represents a large class of alternatives, while the so-called curse of dimensionality problem of nonparametric estimation is avoided. Hence, the proposed approach seems to be a good compromise between feasibility and power.

To estimate w_1^* in (7) nonparametrically, note that

$$\begin{aligned} w_1^*(Y_{t-1}) &= E[h_{t,\alpha}(\theta_0)|Y_{t-1} = y] \\ &= E[h_{t,\alpha}(\theta_0)|F_Y(Y_{t-1}) = F_Y(y)] \\ &:= E[h_{t,\alpha}(\theta_0)|U_{t-1} = F_Y(y)], \end{aligned}$$

where the conditioning variable Y_{t-1} is normalized by taking an integral transformation, provided that F_Y is strictly increasing. That is, we define

$$U_t := F_Y(Y_t) \text{ and } U_{n,t-1} := F_n(Y_{t-1}) := \frac{1}{n} \sum_{s=1}^n 1\{Y_{s-1} \leq Y_{t-1}\}, \quad (8)$$

where the empirical cdf F_n is implicitly defined. Note that U_t is uniformly distributed over the interval $[0, 1]$. There are many nonparametric estimators available for the regression function w_1^* , see Li and Racine (2007). The popular kernel estimator is not recommended in our current setting, as it is known to be inaccurate at the tails of the conditional distribution. For our particular situation, one appealing estimator for w_1^* is a global estimator obtained by series methods.

We introduce the series estimator for w_1^* as follows. First, we consider a vector of series base functions:

$$p^S(u) := (p_0(u), \dots, p_S(u))', \quad u \in [0, 1], \quad S \geq 1, \quad (9)$$

where $p_0(u) \equiv 1$ and $p^S(u)$ is a truncation of an orthogonal polynomial sequence $\{p_S(u)\}_{S=0}^\infty$. Using these base functions, we approximate the regression function $g(u) := E[h_{t,\alpha}(\theta_0)|U_{t-1} = u]$ by $p^S(u)'\gamma$ with an appropriate vector γ . The parameter γ can be estimated by the ordinary least squares (OLS) regression of $H_{n,\alpha} := [h_{1,\alpha}(\hat{\theta}_n), \dots, h_{n,\alpha}(\hat{\theta}_n)]'$ on $P_n := [p^S(U_{n,0}), \dots, p^S(U_{n,n-1})]'$. We define the series estimator for $g(u)$ as

$$\hat{g}(u) := p^S(u)'\gamma_n, \quad (10)$$

where $\gamma_n = [P'_n P_n]^{-1} P'_n H_{n,\alpha}$ is the OLS estimator for γ . Then, the final estimator for $w_1^*(Y_{t-1})$ is

$$\hat{w}_1^*(Y_{t-1}) := \hat{g}(F_n(Y_{t-1})), \quad 1 \leq t \leq n. \quad (11)$$

Note that $\hat{W}_1^* := (\hat{w}_1^*(Y_0), \dots, \hat{w}_1^*(Y_{n-1}))'$ are simply the fitted values of an OLS regression, that is, $\hat{W}_1^* = P_n [P'_n P_n]^{-1} P'_n H_{n,\alpha}$. Then, we construct the weighted test statistic

$$K_{P, \hat{w}_1^*} := \frac{1}{\sqrt{\hat{B}_{\hat{w}_1^*} P}} \sum_{t=R+1}^n \hat{w}_1^*(Y_{t-1}) (h_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha), \quad (12)$$

where $\hat{B}_{\hat{w}_1^*} := P^{-1} \sum_{t=R+1}^n [\hat{w}_1^*(Y_{t-1})]^2$. Note that K_{P, \hat{w}_1^*} will depend on the number of components S in the series estimator \hat{w}_1^* , so we make this dependence explicit and write $K_{P, \hat{w}_1^*}(S)$. Next, we address the challenging and important practical issue of “bandwidth” choice of S for testing purposes. Let S_{\max} denote an upper bound for S , and $\gamma_n = (\gamma_{n0}, \dots, \gamma_{nS_{\max}})'$ the corresponding vector of OLS estimates. Note that under the null $F_{Y_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) = \alpha$, hence the coefficients $\gamma_{n1}, \dots, \gamma_{nS_{\max}}$ are expected to be close to zero, but significantly different from zero under the alternative if S_{\max} is sufficiently large.

Then, we construct an aggregate measure of significance as

$$Q_S = n \sum_{j=1}^S \gamma_{nj}^2,$$

and choose S^* such that

$$S^* = \min\{S : 1 \leq S \leq S_{\max}; L_S \geq L_h, h = 1, 2, \dots, S_{\max}\}$$

where $L_S = Q_S - S \log n$. The aggregate measure of significance Q_S is constructed as the sum of the squares of the normalized estimated coefficients of the series estimator, similar to Neyman’s (1937) smooth test. Here, we follow Ledwina (1994) and employ the Schwarz’s (1978) Bayesian Information Criterion (BIC) to select the number of component S^* . Therefore, the selection of the

number of components S in the series estimator is fully data-driven.

Our final recommended test combines the nonparametric weighted backtest with the test of independence of Christoffersen (1998). That is, we define

$$D_n := \frac{(K_{P, \hat{w}_1^*}(S^*))^2 + (K_{P, \hat{w}_{a1}})^2}{\alpha(1 - \alpha)},$$

where $\hat{w}_{a1} := h_{t-1, \alpha}(\hat{\theta}_{t-2}) - \alpha$. We call the new test based on D_n the D-test, as it is fully Data-Driven. The motivation for this combination is that the two test statistics $K_{P, \hat{w}_1^*}(S^*)$ and $K_{P, \hat{w}_{a1}}$ are asymptotically independent under the null, so they provide independent pieces of information. In our simulations, we show that the D-test has good finite-sample size and power properties.

To investigate the consistency properties of D_n , we introduce a sequence of coordinates $\{\gamma_j\}_{j=0}^{\infty}$ in the series expansion of $w_1^*(\cdot)$,

$$w_1^*(Y_{t-1}) = \sum_{j=0}^{\infty} \gamma_j p_j(F_Y(Y_{t-1})),$$

where $\gamma_j = E[h_{t, \alpha}(\theta_0) p_j(F_Y(Y_{t-1}))]$, $j \in \mathbb{Z}$, $\gamma_0 = \alpha$ and $p_0(F_Y(Y_{t-1})) \equiv 1$. Our next result establishes the limiting distribution of D_n .

THEOREM 3: *Let Assumptions 1-6 in Appendix A hold. Then, for HS and FHS models:*

(i) *Under H_0 ,*

$$D_n \xrightarrow{d} \chi_2^2,$$

where χ_2^2 is the chi-squared random variable with two degrees of freedom.

(ii) *Under H_A , $D_n \rightarrow +\infty$, provided $\gamma_j \neq 0$ for some $0 < j \leq S_{\max}$.*

The proposed data-driven test rejects H_0 if $D_n > \chi_{2, 1-\tau}^2$, where $\chi_{2, 1-\tau}^2$ denotes the $(1 - \tau)$ -th quantile of the chi-squared distribution with two degrees of freedom. The assumption for consistency is quite general, as there is always such j when S_{\max} is sufficiently large, as long as $F_{Y_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) \neq \alpha$. Allowing $S_{\max} \rightarrow \infty$ with the sample size will lead to a consistent test

for any alternative for which $F_{Y_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) \neq \alpha$. In finite samples, S_{\max} is always finite, so for all practical matters a theory with finite S_{\max} suffices. Unreported simulations suggest that the test is not sensitive to S_{\max} , as long as it is moderate relative to the sample size P . A rule of thumb that we choose is $S_{\max} = \lceil P^{2/5} \rceil$, where $\lceil \cdot \rceil$ represents the integer part, which seems to work well with the data generating processes considered in the simulations.

4. Monte Carlo experiments

In this section, the traditional unconditional backtest of Kupiec (K-test), the conditional backtest of Christoffersen (C-test) and the proposed new data-driven test (D-test) are compared in a set of Monte Carlo experiments. As Theorem 1 proves, the K-test is inconsistent for evaluating optimality of the HS or FHS model forecasts. Therefore, the K-test is expected to have poor power in finite samples. To see the impact of the data-driven selection of S^* on D_n , we also compare the finite sample performances of the proposed data-driven D-test with the corresponding one using $S = 5$ (denoted as $D5$). To implement our test, we employ a power series estimator using the Legendre polynomials defined on the unit interval (after taking the integral transform of Y_{t-1}) to estimate the nonparametric weights w_1^* .⁸ The upper bound for the number of the base functions included in the series estimator is set to be $S_{\max} = 9$.

Under the null hypothesis, the data generating process (DGP) for the HS model is $Y_t = \varepsilon_t$, where $\{\varepsilon_t\}_{t=1}^n$ are independent and identically distributed (*i.i.d.*) standard normal, in short $\varepsilon_t \sim i.i.d.N(0, 1)$, denoted as $A0_{HS}$. For the FHS model, we define the null model as a GARCH(1,1) model, denoted as $A0_{FHS}$:

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 0.05 + 0.1Y_{t-1}^2 + 0.85\sigma_{t-1}^2,$$

⁸The Legendre polynomials $p_s(u)$ are defined as

$$p_s(u) = \frac{1}{s!} \frac{d^s}{du^s} \left[(u^2 - u)^s \right],$$

where s is a nonnegative integer and $u \in [0, 1]$. The first few Legendre polynomials are $p_0(u) = 1$, $p_1(u) = 2u - 1$, $p_2(u) = 6u^2 - 6u + 1$, and $p_3(u) = 20u^3 - 30u^2 + 12u - 1$.

where $\{\varepsilon_t\}$ are *i.i.d.* standardized student-t disturbances with $\nu = 5$ degrees of freedom. The GARCH(1,1) model is the most widely used specification for daily financial data, see e.g., Andersen et. al. (2006). Additionally, the parameter values we choose represent typical parameter values in empirical applications. To investigate power performances, we choose the following alternatives:

ALT1: RiskMetrics model: $Y_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = 0.06Y_{t-1}^2 + 0.94\sigma_{t-1}^2$.

ALT2: AR(1)-GARCH(1,1) model: $Y_t = 0.3Y_{t-1} + u_t$, $u_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = 0.05 + 0.1u_{t-1}^2 + 0.85\sigma_{t-1}^2$.

ALT3: EGARCH(1,1) model: $Y_t = h_t \varepsilon_t$, $\ln h_t^2 = 0.01 + 0.9 \ln h_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - (2/\pi)^{1/2}) - 0.8\varepsilon_{t-1}$.

ALT4: TAR model: $Y_t = a_t Y_{t-1} + \varepsilon_t$, $a_t = 0.7 \cdot 1(\varepsilon_{t-1} < -0.5) - 0.7 \cdot 1(\varepsilon_{t-1} \geq 0.5)$.

ALT5: Bilinear model (BIL): $Y_t = 0.7Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$.

ALT6: Exponential Autoregressive model (EXP): $Y_t = 0.6Y_{t-1} \exp(-0.5Y_{t-1}^2) + \varepsilon_t$.

ALT7: AR(1)-GARCH(1,1) model with Hansen's (1994) skewed t innovations denoted by $\varepsilon_t \sim GT(\cdot|\eta_t, \lambda_t)$ (AR-GARCH-GT):

$$\begin{aligned}
 Y_t &= 0.5Y_{t-1} + v_t, & v_t &= \sigma_t \varepsilon_t, & (13) \\
 \sigma_t^2 &= 0.025 + 0.25v_{t-1}^2 + 0.5\sigma_{t-1}^2, \\
 \varepsilon_t &\sim GT(\cdot|\eta_t, \lambda_t), & \eta_t &= g_{(2.1,30)}(\tilde{\eta}_t), & \lambda_t &= g_{(-0.9,0.9)}(\tilde{\lambda}_t), \\
 \tilde{\eta}_t &= -1 - 0.5Y_{t-1} - 0.1Y_{t-1}^2, & \tilde{\lambda}_t &= -0.1 - 0.15Y_{t-1} - 0.1Y_{t-1}^2,
 \end{aligned}$$

where g is the logistic function, i.e., $g_{(L,U)}(x) = L + (U - L)/(1 + \exp(-x))$.

In models ALT1-ALT6, $\varepsilon_t \sim i.i.d.N(0, 1)$. Model ALT1 is the JP Morgan's (1996) RiskMetric model. Model ALT3-ALT5 are studied in Escanciano and Olmo (2010). Model ALT6 is a standard nonlinear model and model ALT7 is from Hansen (1994).

For all the tests implemented the nominal level is fixed at 5%, the asymptotic critical values are used, and the rejection rates are calculated over 1000 Monte Carlo trials. We consider two VaR levels, $\alpha = 5\%$ and 1% , respectively. For simplicity, we employ the fixed forecast scheme for

both the HS and FHS models. In the simulations we consider two different combinations of the in-sample and out-of-sample size. They are $(R, P) = (2500, 250)$ and $(R, P) = (5000, 500)$. Note that $P/R = 0.1$, which allows us to use the asymptotic theory with $\pi = 0$ as an approximation for the finite sample distribution. The large in-sample sizes R considered also guarantee a good precision in the estimated quantiles and our nonparametric estimates of $F_{Y_{t-1}}(m_\alpha(I_{t-1}, \theta_0))$.

Table 1 and Table 2 contain the simulation results for HS and FHS models, respectively. The main conclusions from these tables are as follows. First, all the tests appear to have satisfactory size performances, with rejection probabilities close to the nominal level of 5%. The only exceptional case is the C-test, which is heavily undersized when backtesting 1% VaRs.

TABLES 1 AND 2 ABOUT HERE

Second, the K-test generally displays very poor empirical power properties in detecting the lack of optimality of HS and FHS models for the alternatives considered, which supports our theoretical results. For HS models, the empirical power of the K-test is high for ALT1, moderate for ALT3 and ALT7, and low for the rest of the alternatives. For FHS models, the power is very low for all, but for ALT1. The high power against ALT1 is theoretically explained by the very large values of σ_K , see our Theorem 1. In cases where the K-test has low power, we do not see a significant increase of the power when the out-of-sample size P increases, which is also in accordance with our theoretical results.

Third, the C-test has reasonably good power for backtesting HS models, but it performs poorly for backtesting FHS models. This lack of power suggests that (6) takes on small values for these alternatives. Thus, this test is not recommended for detecting lack of optimality of FHS models.

Fourth, the D-test has satisfactory good power performances for backtesting both HS and FHS models. A comparison with D_5 reveals that there are not substantial benefits from choosing S from the data in these simulations. It is unknown if this conclusion can be extrapolated to other DGPs.

Finally, we observe significant differences between the two VaR levels considered, with better performances at 5% VaR level. For example, the C-test has poor power performances in backtest-

ing 1% VaRs, especially for FHS models. However, it should be stressed that the results at 1% have to be taken with caution given the relatively small out-of-sample sizes considered here; see Escanciano and Olmo (2010) for a discussion and illustration of misleading inferences at 1% VaR level.

To conclude, these simulations confirm our theoretical results. It is shown that the K-test is inconsistent for detecting non-optimality of HS and FHS forecasts, and that the new proposed data-driven test (D-test) has excellent empirical power properties, while being simple to implement.

5. Application

In this application we illustrate the practical importance of our theoretical results in financial risk management, particularly in the evaluation of HS and FHS models and in the subsequent determination of the multiplication factors for market risk capital requirements. To that end, we compare backtesting results from the aforementioned three tests, the K-test, C-test and the proposed D-test, using portfolios based on three representative U.S. stocks traded on the New York Stock Exchange (NYSE). The stocks are Walt Disney (DIS), General Electric (GE) and Merck & Company (MRK). Daily data on their market closure prices⁹ are collected over the period of 01/04/1999 - 12/31/2009, and then the daily returns are calculated as 100 times the difference of the log prices. The compositions of the three portfolios considered are (0.4, 0.1, 0.5), (0.1, 0.1, 0.8) and (0.3, 0.1, 0.6), respectively, where the numbers in each parentheses from left to right represent the portfolio weights on DIS, GE and MRK, respectively. These weights are chosen for illustrative purposes but not based on any portfolio optimizing behavior. We would expect to obtain similar conclusions for optimal portfolios according to our theoretical results.

In order to get a preliminary view of the data set, we present some summary statistics in Table 3 and the graph of daily returns in Figure 1 for the three portfolios. Overall, the daily returns are highly volatile, slightly autocorrelated, negatively skewed, leptokurtic and exhibit volatility clustering.

⁹Data source is Yahoo! Finance.

TABLE 3 AND FIGURE 1 ABOUT HERE

For each portfolio, we forecast daily VaRs for the last year of the chosen sample, 01/02/2009-12/31/2009, with a corresponding sample size of $P = 252$. For this application, following the standard practice we choose a rolling forecast scheme and backtest those VaR forecasts with the K-test, C-test and D-test, respectively. The in-sample period is 01/04/1999 - 12/31/2008, yielding an in-sample size of $R = 2515$. This large in-sample size, relative to the out-of-sample size, justifies the asymptotic approximations used. Both HS and FHS models at 5% and 1% VaR levels are considered in the application. For the FHS model, an AR(1)-GARCH(1,1) model is used. For each test, we report the p-value, denoted as p_K , p_C and p_D , where the subscripts, K, C and D, represent the K-test, C-test and D-test, respectively. Under the current regulatory framework, the computation of the multiplication factors for market risk capital requirements, mf , is based on the number of VaR violations in the out-of-sample, say $N := \sum_{t=R+1}^n h_{t,\alpha}(\hat{\theta}_{t-1})$, so we also report N and the corresponding mf for each case.¹⁰

TABLE 4 ABOUT HERE

Table 4 collects the application results. The top panel reports the application results for backtesting HS models. As expected, HS is not the optimal model for the three portfolios, as it ignores some important properties of our data such as time-varying volatility and volatility clustering. We see that the K-test fails to reject the optimality of HS for the three portfolios with $p_K \geq 0.1629$. This result is in line with our theoretical findings on the inconsistency of the K-test. On the contrary, the proposed D-test is able to detect the non-optimality of the HS model. We observe that

¹⁰According to Basel Committee on Banking Supervision (1996b), the multiplication factors are determined by classifying the number of VaR violations in the out-of-sample into three zones, distinguished by colors into a hierarchy of responses. The green zone is the zone where the backtesting results are consistent with an accurate model; the yellow zone corresponds to the backtesting results that can be consistent with either accurate or inaccurate models; and the red zone is the range where the backtesting results almost certainly indicate an inaccurate model. For instance, for $\alpha = 0.01$ and $P = 250$, see Campbell (2005), the multiplication factors mf are computed as

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

the D-test strongly rejects the optimality of HS for Portfolio 1 at both 5% and 1% VaR levels with $p_D = 0.0127$ and $p_D = 0.0000$, respectively, for Portfolio 2 at 5% VaR level with $p_D = 0.0185$, and for Portfolio 3 at 5% VaR level with $p_D = 0.0190$. Hence, a researcher evaluating VaR forecasts based on the K-test would erroneously accept a non-optimal HS model as a good model for measuring market risk in these portfolios.

The results for backtesting FHS models are reported in the bottom panel of Table 4. FHS is expected to provide better VaR forecasts in the application, as it takes time-varying volatility into account. However, based on the backtesting results from the D-test, the optimality of FHS is still rejected for Portfolio 1 at both 5% and 1% VaR levels with $p_D = 0.0183$ and $p_D = 0.0000$, respectively, for Portfolio 2 at 5% VaR level with $p_D = 0.0085$, and for Portfolio 3 at 5% VaR level with $p_D = 0.0001$, respectively. Again, we observe the inability of the K-test to reject the optimality of FHS for the three portfolios, with a p-value of $p_K \geq 0.2439$.

As for the determination of the multiplication factors, we see that they are all set at the lowest level of 3. The backtesting results that cause the lowest level of the multiplication factors do not suggest a problem with the VaR model used, which should not be the case in our application. For example, for Portfolio 1, HS and FHS models are rejected by the proposed D-test, so the multiplication factors should also be increased accordingly in order to compensate for the inefficient forecasts. Hence, our theoretical and empirical results suggest that in the current regulatory framework the multiplication factors are underestimated for the cases where HS or FHS is non-optimal. It is the inconsistency of the K-test that leads to such underestimation. In practice, the underestimated multiplication factors would make banks lose the incentives to improve their models, and also might result in insufficient market risk capital reserves, which expose banks to high risks. Therefore, a major implication of our results is that the multiplication factors determined by the current regulatory framework are highly misleading when HS and FHS provide inefficient forecasts.

6. Conclusion

In this paper, we prove that the traditional unconditional backtest is inconsistent in testing the optimality of HS or FHS models. We show that its asymptotic power function can be higher or lower than the nominal level, depending on the relative size of certain long run variance and the asymptotic null variance of the test, but it is *always* smaller than one. To overcome the inconsistency problem of the most commonly used unconditional backtest, we propose a general class of weighted backtesting procedures, which includes consistent tests and well known tests such as the independence test as special cases, and we select within this class a data-driven test with good power properties. Our test combines nonparametric weights estimated by a series method, with the classical backtest of independence. Our simulation experiments confirm our theoretical findings. An application to portfolios of some U.S stocks reveals the practical importance of our theoretical results in financial risk management. In practice, the use of the K-test for backtesting HS or FHS models might result in the erroneous acceptance of a non-optimal model, and in the underestimation of the multiplication factors derived from the current regulatory framework. Our findings suggest that a modification of the current computation of multiplication factors is needed, to account for the inconsistency of the K-test. A promising approach would be to base the computation of multiplication factors on the proposed D-test, given its excellent finite-sample performance. This topic is deferred for future research.

Appendix A. Assumptions

To derive our theoretical results, we first need to introduce some notations and assumptions. Assume that $I_{t-1} = (X_{t-1}, X_{t-2}, \dots)$, where X_t includes Y_t , and define the α -mixing coefficients as

$$\alpha(m) = \sup_{B \in \mathcal{F}_n, A \in \mathcal{P}_{n+m}} |P(A \cap B) - P(A)P(B)|, \quad m \geq 1, \quad (\text{A.1})$$

where the σ -fields \mathcal{F}_n and \mathcal{P}_n are $\mathcal{F}_n := \sigma(\varepsilon_t, t \leq n)$ and $\mathcal{P}_n := \sigma(\varepsilon_t, t \geq n)$, respectively. Henceforth, we take the convention that $\varepsilon_t = Y_t$ for HS models.

Define the family of the conditional distributions $F_x(y) := P(Y_t \leq y \mid I_{t-1} = x)$ and let $f_x(y)$ be the associated conditional densities. The following regularity conditions are used in the main results.

Assumption A1: $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is strictly stationary and strong mixing process with mixing coefficients satisfying $\sum_{j=1}^{\infty} \alpha(j) < \infty$. $F_\varepsilon(y)$ is strictly increasing and continuous in $y \in \mathbb{R}$. The conditional variance $\sigma^2(I_{t-1}, \beta_0) > 0$ a.s. and $\sigma_K^2 > 0$.

Assumption A2: The family of distributions functions $\{F_x : x \in \mathbb{R}^\infty\}$ has Lebesgue densities $\{f_x : x \in \mathbb{R}^\infty\}$ that are uniformly bounded from above and below, i.e., $\inf_{x \in \mathbb{R}^\infty, y \in \mathbb{R}} |f_x(y)| > C^{-1}$ and $\sup_{x \in \mathbb{R}^\infty, y \in \mathbb{R}} |f_x(y)| \leq C$, for a positive constant C , and equicontinuous: for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\sup_{x \in \mathbb{R}^\infty, |y-z| \leq \delta} |f_x(y) - f_x(z)| \leq \epsilon$.

Assumption A3: The parameter space Θ is compact in \mathbb{R}^p . The true parameter θ_0 belongs to the interior of Θ . The estimator $\hat{\theta}_t$ is such that $\max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t - \theta_0) = O_P(1)$ under the null and $\max_{R \leq t \leq n} (\hat{\theta}_t - \theta_0) = o_P(1)$ under the alternative.

Assumption A4: $R, P \rightarrow \infty$ as $n \rightarrow \infty$, and $\pi := \lim_{n \rightarrow \infty} P/R = 0$.

Assumption A5: Assume $\mu(I_{t-1}, \beta)$ and $\sigma(I_{t-1}, \beta)$ are continuously differentiable in β (a.s.) such that $E[\sup_{\beta \in B_0} |\dot{\mu}_t(\beta)|^2] < C$ and $E[\sup_{\beta \in B_0} |\dot{\sigma}_t(\beta)|^2] < C$, for a neighborhood B_0 of β_0 .

Assumption A6: The functions $p_j(u)$ for $j = 0, \dots, S$ are mutually orthogonal polynomials on $[0, 1]$ and continuously differentiable with bounded derivatives.

Assumptions A1 to A6 are standard in the literature. Assumption A1 suffices for Theorem 1. For sufficient conditions for the α -mixing ARMA-GARCH processes, see e.g., Doukhan (1994), Carrasco and Chen (2002), Francq and Zakoïan (2006) and references therein. Assumptions A1 to A5 are needed for Theorem 2. Assumption A2 is required as in Escanciano and Olmo (2010). Assumption A3 can be easily verified for the estimators involved in HS and FHS models such as the empirical quantile estimator in both HS and FHS models and the QMLE in FHS models. For example, in an AR(1)-GARCH(1,1) model under the fixed forecasting scheme where $\theta_0 = (\beta_0, F_\varepsilon^{-1}(\alpha))$, the parameter β_0 can be estimated by the QMLE

$$\widehat{\beta}_t \equiv \widehat{\beta}_R = \arg \min_{\beta \in \Theta_\beta} R^{-1} \sum_{s=1}^R \left[\frac{(Y_s - \mu_0 Y_{s-1})^2}{\widetilde{\sigma}_s^2(\beta)} + \log \widetilde{\sigma}_s^2(\beta) \right], \quad (\text{A.2})$$

where $\widetilde{\sigma}_s^2$ is defined recursively, for $s \geq 1$, by

$$\widetilde{\sigma}_s^2(\beta) = \omega_0 + \omega_1(Y_{s-1} - \mu_0 Y_{s-2})^2 + \omega_2 \widetilde{\sigma}_{s-1}^2(\beta)$$

and $\beta = (\mu_0, \omega_0, \omega_1, \omega_2)' \in \Theta_\beta \equiv \{\beta \in \mathbb{R}^4 : |\mu_0| < 1, \omega_0 > 0, \omega_1 \geq 0, \omega_2 \geq 0 \text{ and } \omega_1 + \omega_2 < 1\}$. Francq and Zakoïan (2004) provide further details on the QMLE and sufficient conditions for Assumption A3 to hold. Among many others, Yoshihara (1995) provides sufficient conditions for Assumption 3 to hold with the empirical quantile estimator under α -mixing conditions. We discuss Assumption A4 in the paper. Assumption A5 is standard in the literature and satisfied by ARMA-GARCH models under mild moment conditions. Assumption A6 is satisfied, for instance, by the Legendre polynomials used in the simulations.

Appendix B. Mathematical proofs

Proof of Theorem 1: Note that the hit sequence $\{h_{t,\alpha}(\theta_0)\}_{t \in \mathbb{Z}}$ is a function of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, and hence, is strictly stationary and strong mixing with mixing coefficients $\alpha_h(j)$ bounded by $\alpha(j)$ in (A.1). Hence, by Assumption A1, the mixing coefficients satisfy $\sum_{j=1}^{\infty} \alpha_h(j) < \infty$. In addition, by

definition, for the HS model,

$$E[h_{t,\alpha}(\theta_0) - \alpha] = E[1(Y_t \leq F_Y^{-1}(\alpha))] - \alpha = F_Y(F_Y^{-1}(\alpha)) - \alpha = 0,$$

and for the FHS model,

$$E[h_{t,\alpha}(\theta_0) - \alpha] = E[1(\varepsilon_t \leq F_\varepsilon^{-1}(\alpha))] - \alpha = F_\varepsilon(F_\varepsilon^{-1}(\alpha)) - \alpha = 0.$$

Therefore, we have that the time series $\{h_{t,\alpha}(\theta_0) - \alpha\}_{t=1+R}^n$ is strictly stationary and strong mixing, has zero mean and is bounded for both the HS and FHS models. Thus, a strong mixing CLT applies to K_P under the alternative hypothesis H_A , and then it holds that

$$K_P \xrightarrow{d} N(0, \sigma_K^2).$$

Hence, the power function of K_P at $\tau\%$ nominal level is

$$\begin{aligned} \Pi(\tau, \alpha) &:= \lim_{P \rightarrow \infty} \Pr \left(\left| \frac{K_P}{\sigma_K} \right| > z_{1-\tau/2} \frac{\sqrt{\alpha(1-\alpha)}}{\sigma_K} \mid H_A \right) \\ &= 2 \left(1 - \Phi \left(z_{1-\tau/2} \frac{\sqrt{\alpha(1-\alpha)}}{\sigma_K} \right) \right) \\ &= 2\Phi \left(z_{\tau/2} \frac{\sqrt{\alpha(1-\alpha)}}{\sigma_K} \right). \end{aligned}$$

□

Proof of Theorem 2: The proof follows from Theorem 1 and Corollary 1 in Escanciano and Olmo (2010), but replacing K_n there by the weighted version

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

where we define, for $\theta \in \Theta$, $F_{t-1}(\theta) := F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta))$ to simplify notation. Notice that all the

arguments used in Escanciano and Olmo (2010) apply to the current setting without changes. \square

Proof of Lemma 1: By the law of iterated expectations

$$E[w^*(I_{t-1})(h_{t,\alpha}(\theta_0) - \alpha)] = \text{Var}[F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0))] \geq 0, \quad (\text{B.1})$$

with equality to zero only when $F_{I_{t-1}}(m_\alpha(I_{t-1}, \theta_0)) = \alpha$ a.s., that is, only under the null. \square

Proof of Theorem 3:

We first prove (i). Define $S^* = \min\{S : 1 \leq S \leq S_{\max}; L_S \geq L_h, h = 1, 2, \dots, S_{\max}\}$, where $L_S = Q_S - S \log n$. We prove that, under Assumptions A1-A6 and H_0 ,

$$\lim_{n \rightarrow \infty} P(S^* = 1) = 1. \quad (\text{B.2})$$

Notice that

$$P(S^* = 1) = 1 - \sum_{s=2}^{S_{\max}} P(S^* = s) \geq 1 - \sum_{s=2}^{S_{\max}} P(L_s \geq L_1).$$

Now, for $1 < s \leq S_{\max}$,

$$P(L_s \geq L_1) \leq P(Q_s \geq (s-1) \log n).$$

We shall show that under the null $Q_s = O_p(1)$. To that end, let 1_n be a $n \times 1$ vector of ones, and write

$$\begin{aligned} \gamma_n &= [n^{-1}P'_nP_n]^{-1} (n^{-1}P'_n(H_{n,\alpha} - \alpha 1_n) + \alpha n^{-1}P'_n 1_n) \\ &\equiv [n^{-1}P'_nP_n]^{-1} \{A_{1n} + \alpha A_{2n}\}. \end{aligned}$$

Assumption A6 and the Glivenko-Cantelli's theorem imply that

$$[n^{-1}P'_nP_n]^{-1} = [P'P]^{-1} + o_P(1),$$

where $P := [p^s(U_0), \dots, p^s(U_{n-1})]'$. On the other hand,

$$\begin{aligned}\sqrt{n}A_{1n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n p^s(U_{t-1}) \left(h_{t,\alpha}(\widehat{\theta}_n) - \alpha \right) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n \{p^s(U_{n,t-1}) - p^s(U_{t-1})\} \left(h_{t,\alpha}(\widehat{\theta}_n) - \alpha \right) \\ &=: A_{11n} + A_{12n}.\end{aligned}$$

An application of Theorem A1 in Delgado and Escanciano (2007) with (using their notation)

$$w(\varepsilon_{n,t}, I_{n,t}, x) = p_j(F(Y_{t-1})) [h_{t,\alpha}(\theta) - F_{t-1}(\theta)], \quad j = 0, \dots, s,$$

where $x = (F, \theta) \in \Pi := \mathcal{H} \times \Theta$, with \mathcal{H} the class of non-decreasing functions of bounded variation, yields that $A_{12n} = o_p(1)$ by a stochastic equicontinuity argument. Similarly, by Escanciano and Olmo (2010) it can be shown that

$$\begin{aligned}A_{11n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n p^s(U_{t-1}) (h_{t,\alpha}(\theta_0) - \alpha) + \frac{1}{\sqrt{n}} \sum_{t=1}^n p^s(U_{t-1}) \left(F_{t-1}(\widehat{\theta}_n) - F_{t-1}(\theta_0) \right) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n p^s(U_{t-1}) (h_{t,\alpha}(\theta_0) - \alpha) + E[p^s(U_{t-1}) \frac{\partial F_{t-1}(\theta_0)}{\partial \theta'}] \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1) \\ &= O_p(1).\end{aligned}$$

We write each component of $\sqrt{n}A_{2n}$ as

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n p_j(U_{n,t-1}) = \sqrt{n} \int p_j(F_n(x)) F_n(dx),$$

where F_n is the empirical cdf of $\{Y_{t-1}\}_{t=1}^n$. Lemma 20.10 in van der Vaart (1998) and the weak convergence of $\sqrt{n}(F_n - F_Y)$ imply that $\sqrt{n}(A_{2n} - e_1) = O_p(1)$, where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{s+1}$. From the previous arguments and the definition of Q_s we conclude that under the null $Q_s = O_p(1)$.

It then follows that

$$P(Q_s \geq (s-1) \log n) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Therefore, (B.2) holds.

To simplify notations, we write $\widehat{p}_{1,t-1} = p_1(U_{n,t-1})$ and $p_{1,t-1} = p_1(U_{t-1})$. With $S^* = 1$, the series estimators for the weights are

$$\widehat{w}_1^*(Y_{t-1}) = \gamma_{n0} + \gamma_{n1} p_1(U_{n,t-1}), \quad t = R+1, \dots, n,$$

where $(\gamma_{n0}, \gamma_{n1})'$ is the OLS estimator γ_n with $s = 1$, i.e.,

$$\gamma_{n0} = \overline{h_{t,\alpha}} - \gamma_{n1} \overline{\widehat{p}_{1,t-1}}$$

and

$$\gamma_{n1} = \frac{\frac{1}{n} \sum_{t=1}^n [h_{t,\alpha}(\widehat{\theta}_n) - \overline{h_{t,\alpha}}] [\widehat{p}_{1,t-1} - \overline{\widehat{p}_{1,t-1}}]}{\frac{1}{n} \sum_{t=1}^n [\widehat{p}_{1,t-1} - \overline{\widehat{p}_{1,t-1}}]^2} := \frac{a_n}{b_n}$$

with $\overline{h_{t,\alpha}} = n^{-1} \sum_{t=1}^n h_{t,\alpha}(\widehat{\theta}_n)$ and $\overline{\widehat{p}_{1,t-1}} = n^{-1} \sum_{t=1}^n \widehat{p}_{1,t-1}$. We show that, under Assumptions A1-A6 and H_0 , uniformly in $y \in \mathbb{R}$,

$$\widehat{w}_1^*(y) \xrightarrow{p} w_1^*(y) \equiv \alpha. \quad (\text{B.3})$$

Notice that

$$\begin{aligned} a_n &= \frac{1}{n} \sum_{t=1}^n [h_{t,\alpha}(\widehat{\theta}_n) - \overline{h_{t,\alpha}}] \widehat{p}_{1,t-1} \\ &= \frac{1}{n} \sum_{t=1}^n [h_{t,\alpha}(\widehat{\theta}_n) - \alpha] [\widehat{p}_{1,t-1} - p_{1,t-1}] + \frac{1}{n} \sum_{t=1}^n [h_{t,\alpha}(\widehat{\theta}_n) - \alpha] p_{1,t-1} + (\alpha - \overline{h_{t,\alpha}}) \frac{1}{n} \sum_{t=1}^n \widehat{p}_{1,t-1} \\ &:= a_{1n} + a_{2n} + a_{3n}, \end{aligned}$$

Since the polynomial function $p_1(\cdot)$ is equicontinuous and $\sup_{y \in \mathbb{R}} |F_n(y) - F(y)| \xrightarrow{p} 0$, then

$$\max_{1 \leq t \leq n} |\widehat{p}_{1,t-1} - p_{1,t-1}| := \max_{1 \leq t \leq n} |p_1(F_n(Y_{t-1})) - p_1(F(Y_{t-1}))| \xrightarrow{p} 0,$$

and hence, $a_{1n} = o_p(1)$. By Lemma A1 in Escanciano and Olmo (2010) and the uniform law of large numbers (ULLN),

$$a_{2n} = \frac{1}{n} \sum_{t=1}^n [h_{t,\alpha}(\widehat{\theta}_n) - \alpha] p_{1,t-1} \xrightarrow{p} \text{cov}(h_{t,\alpha}(\theta_0), p_{1,t-1}) = 0.$$

Similarly, $\overline{h_{t,\alpha}} \xrightarrow{p} E[h_{t,\alpha}(\theta_0)] = \alpha$ and $\overline{p_{1,t-1}} = O_p(1)$, so $a_{3n} = o_p(1)$. Hence, $a_n = o_p(1)$, and since $b_n = O_p(1)$, then $\gamma_{n1} \xrightarrow{p} 0$ and $\gamma_{n0} \xrightarrow{p} \alpha$. Therefore, (B.3) is proved. An implication of (B.3) is that $\widehat{B}_{\widehat{w}_1^*} := P^{-1} \sum_{t=R+1}^n [\widehat{w}_1^*(Y_{t-1})]^2$ is a consistent estimator of $B_{w_1^*} := \alpha^2$.

Based on the above, we can write

$$\begin{aligned} K_{P, \widehat{w}_1^*}(S^*) &= \frac{1}{\sqrt{\widehat{B}_{\widehat{w}_1^*} P}} \sum_{t=R+1}^n [\widehat{w}_1^*(Y_{t-1}) - w_1^*(Y_{t-1})][h_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha] \\ &\quad + \frac{1}{\sqrt{\widehat{B}_{\widehat{w}_1^*} P}} \sum_{t=R+1}^n w_1^*(Y_{t-1})[h_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha] \\ &:= K_{1n} + K_{2n}. \end{aligned}$$

An application of Theorem A1 in Delgado and Escanciano (2007) yields that $K_{1n} = o_P(1)$, and Slutsky's Theorem and Theorem 2 imply that $K_{P, \widehat{w}_1^*}(S^*) \xrightarrow{d} N(0, \alpha(1 - \alpha))$.

Theorem 2 in Escanciano and Olmo (2010) implies, with $\pi = 0$, that $K_{P, \widehat{w}_{a1}} \xrightarrow{d} N(0, \alpha(1 - \alpha))$. It remains to prove that $K_{P, \widehat{w}_1^*}(S^*)$ and $K_{P, \widehat{w}_{a1}}$ are asymptotically independent under the null.

This follows from the fact that

$$\begin{aligned}
Cov(K_{P,\hat{w}_1^*}(1), K_{P,\hat{w}_{a1}}) &= Cov(\alpha[h_{t,\alpha}(\theta_0) - \alpha], [h_{t-1,\alpha}(\theta_0) - \alpha][h_{t,\alpha}(\theta_0) - \alpha]) + o(1) \\
&= \alpha^2(1 - \alpha)E[h_{t-1,\alpha}(\theta_0) - \alpha] + o(1) \\
&= o(1).
\end{aligned}$$

Thus, by Cramer-Wold Device,

$$D_n = \frac{(K_{P,\hat{w}_1^*}(S^*))^2 + (K_{P,\hat{w}_{a1}})^2}{\alpha(1 - \alpha)} \xrightarrow{d} \chi_2^2,$$

where χ_2^2 is the chi-squared random variable with two degrees of freedom.

We now prove (ii). To that end, we need to prove that, under H_A ,

$$P(S^* \geq j) \rightarrow 1, \tag{B.4}$$

provided that $\gamma_j \neq 0$ for some $0 < j \leq S_{\max}$, and $\gamma_1 = \gamma_2 = \dots = \gamma_{j-1} = 0$.

For $s = 1, \dots, j - 1$,

$$P(S^* = s) \leq P(L_s \geq L_j).$$

and

$$P(L_s \geq L_j) = P(Q_s \geq (s - j) \log n + Q_j) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

The last convergence follows from the law of large numbers, since

$$n^{-1}Q_s \rightarrow 0 \text{ and } n^{-1}Q_j \rightarrow \gamma_j^2 > 0.$$

Therefore, for each $M > 0$,

$$\begin{aligned}
P(K_{P, \hat{w}_1^*}(S^*) \leq M) &= P(K_{P, \hat{w}_1^*}(S^*) \leq M, S^* \geq j) + o(1) \\
&\leq P(K_{P, \hat{w}_1^*}(j) \leq M) + o(1) \\
&= o(1),
\end{aligned}$$

where the last equality follows from the law of large numbers and the fact that

$$\begin{aligned}
E \left[\left(\sum_{k=0}^j \gamma_k p_k(F_Y(Y_{t-1})) \right) (h_{t, \alpha}(\theta_0) - \alpha) \right] &= E \left[\left(\sum_{k=0}^j \gamma_k p_k(F_Y(Y_{t-1})) \right) \left(\sum_{h=1}^{\infty} \gamma_h p_h(F_Y(Y_{t-1})) \right) \right] \\
&= \gamma_j^2 E[p_j^2(F_Y(Y_{t-1}))] > 0,
\end{aligned}$$

by the law of iterated expectations and the orthogonality of the polynomials. Then, $K_{P, \hat{w}_1^*}(S^*) \rightarrow \infty$, and hence $D_n \rightarrow \infty$ as $n \rightarrow \infty$. \square

References

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. In: Elliot, G., Granger, C.W.J., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Vol.1, 1st edition. North-Holland, Amsterdam.

Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don't look back. *Risk* 11, 100-104.

Barone-Adesi, G., Giannopoulos, K., Vosper L., 1999. VaR without correlations for nonlinear portfolios. *Journal of Futures Markets* 19, 583-602.

Barone-Adesi, G., Giannopoulos, K., 2001. Non-parametric VaR techniques: Myths and realities. *Economic Notes* 30, 167-181.

Basel Committee on Banking Supervision, 1996a. Amendment to the capital accord to incorporate market risks. Bank for International Settlements.

Basel Committee on Banking Supervision, 1996b. Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements. Bank for International Settlements.

Basel Committee on Banking Supervision, 2011. Revision to the Basel II market risk framework. Bank for International Settlements.

Berkowitz, J., 2001. Testing density forecasts with applications to risk management. *Journal of Business and Economic Statistics* 19, 4, 465-474.

Berkowitz, J., Christoffersen, P.F., Pelletier, D., 2011. Evaluating Value-at-Risk models with desk-level data. *Management Science* 57, 12, 2213-2227.

Berkowitz, J., O'Brien, J., 2002. How accurate are Value at Risk models at commercial banks? *Journal of Finance* 57, 3, 1093-1111.

Campbell, S.D., 2005. A review of backtesting and backtesting procedures. Working paper, Federal Reserve Board.

Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.

Christoffersen, P.F., 1998. Evaluating interval forecasts. *International Economic Review* 39,

841-862.

Christoffersen, P.F., Pelletier, D., 2004. Backtesting Value-at-Risk: A duration-based approach. *Journal of Financial Econometrics* 2, 84-108.

Delgado, M.A., Escanciano, J.C., 2007. Nonparametric tests for conditional symmetry in dynamic models. *Journal of Econometrics* 141, 652-682.

Doukhan, P., 1994. *Mixing: Properties and Examples*. Lecture Notes in Statistics, Vol. 85. Springer-Verlag, New York.

Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business and Economic Statistics* 22, 4, 367-381.

Escanciano, J.C., Olmo, J., 2010. Backtesting parametric Value-at-Risk with estimation risk. *Journal of Business and Economics Statistics* 28, 36-51.

Escanciano, J.C., Olmo, J., 2011. Robust backtesting tests for Value-at-Risk models. *Journal of Financial Econometrics* 9, 132-161.

Francq, C., Zakoïan, J.-M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 4, 605-637.

Francq, C., Zakoïan, J.-M., 2006. Mixing properties of a general class of GARCH(1,1) models without moment assumptions. *Econometric Theory* 22, 815-834.

Gourieroux, C., Jasiak, J., 2008. Dynamic quantile models. *Journal of Econometrics* 147, 198-205.

Hansen, B.E., 1994. Autoregressive conditional density estimation. *International Economic Review* 35, 705-730.

Koenker, R., Xiao, Z., 2006. Quantile autoregression. *Journal of the American Statistical Association* 101, 980-990.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3, 73-84.

Ledwina, T., 1994. Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association* 89, 1000-1005.

Li, W.K., Ling, S., McAleer, M., 2002. Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* 16, 245-269.

Li, Q., Racine, J.S., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, New Jersey.

Neyman, J., 1937. Smooth test for goodness of fit. *Scandinavian Aktuarietidskr* 20, 150-199.

Jorion, P., 2006. *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd edition. McGraw-Hill, New York.

JP Morgan, 1996. *RiskMetrics-Technical Document*, 4th edition. Morgan Guaranty Trust Company, New York.

Pérignon, C., Smith, D.R., 2008. A new approach to comparing VaR estimation method. *The Journal of Derivatives* 16, 2, 54-66.

Pérignon, C., Smith, D.R., 2010. The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking and Finance* 34, 2, 362-377.

Pritsker, M., 2006. The hidden dangers of historical simulation. *Journal of Banking and Finance* 30, 2, 561-582.

Yoshihara, K., 1995. The Bahadur representation of sample quantile for sequences of strongly mixing random variables. *Statistics and Probability Letters* 24, 299-304.

Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics* 6, 461-464.

van der Vaart, A.W., 1998. *Asymptotic Statistics*. Cambridge University Press, New York.

Table 1: Simulation Results for Backtesting HS Models at 5% Nominal Level

	<i>K</i> -test		<i>D</i> -test		<i>D</i> 5-test		<i>C</i> -Test	
	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR
R=2500, P=250								
$A0_{HS}$	0.047	0.040	0.069	0.044	0.066	0.061	0.050	0.008
ALT1	0.987	0.010	0.569	0.009	0.702	0.009	0.051	0.009
ALT2	0.276	0.137	0.709	0.325	0.678	0.377	0.721	0.222
ALT3	0.409	0.186	0.692	0.341	0.697	0.389	0.811	0.265
ALT4	0.076	0.055	0.623	0.206	0.574	0.326	0.363	0.084
ALT5	0.072	0.089	0.841	0.315	0.844	0.590	0.166	0.131
ALT6	0.151	0.081	0.791	0.286	0.730	0.385	0.627	0.140
ALT7	0.290	0.160	0.699	0.341	0.671	0.361	0.737	0.228
R=5000, P=500								
$A0_{HS}$	0.063	0.037	0.062	0.060	0.063	0.067	0.049	0.011
ALT1	1.000	0.998	0.724	0.002	0.883	0.002	0.003	0.001
ALT2	0.315	0.200	0.903	0.533	0.889	0.568	0.930	0.356
ALT3	0.421	0.289	0.881	0.535	0.888	0.557	0.951	0.400
ALT4	0.120	0.080	0.850	0.363	0.788	0.497	0.591	0.144
ALT5	0.107	0.095	0.980	0.795	0.982	0.794	0.193	0.200
ALT6	0.161	0.105	0.957	0.504	0.927	0.607	0.893	0.284
ALT7	0.326	0.220	0.864	0.500	0.860	0.511	0.937	0.334

Notes: This table compares the size and power properties of Kupiec's test, the proposed data-driven test with data-driven selection of S^* , the corresponding test with the chosen value of $S^* = 5$ and Christoffersen's likelihood ratio test for HS models, denoted as *K*-test, *D*-test, *D*5-test and *C*-test, respectively. 1000 Monte Carlo replications.

Table 2: Simulation Results for Backtesting FHS Models at 5% Nominal Level

	<i>K</i> -test		<i>D</i> -test		<i>D</i> 5-test		<i>C</i> -Test	
	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR
R=2500, P=250								
$A0_{FHS}$	0.041	0.048	0.065	0.040	0.073	0.042	0.050	0.010
ALT1	0.353	0.011	0.157	0.020	0.212	0.015	0.249	0.002
ALT2	0.097	0.080	0.392	0.179	0.361	0.167	0.195	0.042
ALT3	0.132	0.138	0.387	0.208	0.341	0.180	0.209	0.073
ALT4	0.049	0.028	0.236	0.123	0.213	0.115	0.063	0.008
ALT5	0.118	0.212	0.507	0.321	0.513	0.330	0.159	0.043
ALT6	0.055	0.060	0.291	0.133	0.226	0.122	0.090	0.017
ALT7	0.111	0.083	0.347	0.129	0.319	0.125	0.203	0.039
R=5000, P=500								
$A0_{FHS}$	0.063	0.049	0.071	0.068	0.066	0.066	0.054	0.015
ALT1	0.889	0.817	0.705	0.007	0.784	0.004	0.192	0.002
ALT2	0.120	0.069	0.594	0.248	0.525	0.216	0.308	0.040
ALT3	0.173	0.172	0.626	0.368	0.527	0.314	0.323	0.090
ALT4	0.072	0.046	0.357	0.146	0.305	0.141	0.084	0.014
ALT5	0.199	0.264	0.743	0.469	0.749	0.486	0.352	0.102
ALT6	0.085	0.052	0.418	0.180	0.325	0.159	0.110	0.018
ALT7	0.138	0.057	0.532	0.182	0.471	0.168	0.305	0.029

Notes: This table compares the size and power properties of Kupiec's test, the proposed data-driven test with data-driven selection of S^* , the corresponding test with the chosen value of $S^* = 5$ and Christoffersen's likelihood ratio test for FHS models, denoted as *K*-test, *D*-test, *D*5-test and *C*-test, respectively. 1000 Monte Carlo replications.

Table 3: Summary Statistics

	Portfolio 1 (0.4,0.1,0.5)	Portfolio 2 (0.1,0.1,0.8)	Portfolio 3 (0.3,0.1,0.6)
Mean	-0.0311	-0.0470	-0.0364
Variance	3.3390	4.5504	3.5718
Kurtosis	55.0084	182.9456	96.2160
Skewness	-2.8894	-7.3010	-4.4703
Minimum	-35.3413	-56.2868	-42.3231
Maximum	11.4305	10.4282	11.0964
Autocorr.	-0.0317	-0.0208	-0.0285
LB(10)	9.0524	0.2049	1.9774

Notes: This table presents some summary statistics for the daily returns of the three portfolios. For each portfolio, the three numbers in the parentheses represent the corresponding portfolio weights on DIS, GE and MRK, respectively. The summary statistics include the first four moments, minimum, and maximum of each series, the first-order autocorrelation coefficient and the Ljung-Box Q-statistic of order 10 on the squared returns.

Table 4: Application Results

	Portfolio 1 (0.4,0.1,0.5)		Portfolio 2 (0.1,0.1,0.8)		Portfolio 3 (0.3,0.1,0.6)	
	5% VaR	1% VaR	5% VaR	1% VaR	5% VaR	1% VaR
Application Results for Backtesting HS Models						
N	14	4	16	1	16	2
mf	3	3	3	3	3	3
p_K	0.3429	0.1744	0.1629	0.8321	0.1629	0.6290
p_C	0.0992	0.0872	0.1353	0.5497	0.1353	0.9302
p_D	0.0127	0.0000	0.0185	0.6290	0.0190	0.9430
Application Results for Backtesting FHS Models						
N	15	3	13	1	14	2
mf	3	3	3	3	3	3
p_K	0.2439	0.3806	0.4540	0.8321	0.3429	0.6290
p_C	0.1226	0.0629	0.0709	0.5497	0.0138	0.9302
p_D	0.0183	0.0000	0.0085	0.6290	0.0001	0.9430

Notes: N represents the number of violations, and mf the multiplication factor determined under the current regulatory framework. p_K , p_C and p_D represent the p-value for the K-test, C-test and D-test, respectively.

Figure 1: Daily Returns of the Three Portfolio

