Backtesting Expected Shortfall: Accounting for Tail Risk*

Zaichao Du†

Southwestern University of Finance and Economics, China

Juan Carlos Escanciano‡

Indiana University, United States of America

August 24, 2015

Abstract

The Basel Committee on Banking Supervision (BIS) has recently sanctioned Expected Shortfall (ES) as the market risk measure to be used for banking regulatory purposes, replacing the well-known Value-at-Risk (VaR). This change is motivated by the appealing theoretical properties of ES as a measure of risk and the poor properties of VaR. In particular, VaR fails to control for "tail risk". In this transition, the major challenge faced by financial institutions is the unavailability of simple tools for evaluation of ES forecasts (i.e. backtesting ES). The main purpose of this article is to propose such tools. Specifically, we propose backtests for ES based on cumulative violations, which are the natural analogue of the commonly used backtests for VaR. We establish the asymptotic properties of the tests, and investigate their finite sample performance through some Monte Carlo simulations. An empirical application to three major stock indexes shows that VaR is generally unresponsive to extreme events such as

*We thank three anonymous referees and the Editor, Neng Wang, for helpful comments. Also, we would like to thank Yan Dong for providing helpful computational resources. A preliminary version of the ideas of this paper appear in the unpublished paper "Simple Automatic Portmanteau Tests for Conditional Dynamic Models", dated April 4th, 2012.

†RIEM, SWUFE, 55 GuangHuaCun St., Chengdu, China 610074. Email: duzc@swufe.edu.cn. Research funded by the National Natural Science Foundation of China, 71401140.

‡Corresponding author: Department of Economics, Indiana University, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA, e-mail: jescanci@indiana.edu. Research funded by the Spanish Plan Nacional de I+D+I, reference number ECO2012-33053.

those experienced during the recent financial crisis, while ES provides a more accurate description of the risk involved.

Keywords: risk management; expected shortfall; backtesting; tail risk; Value-at-Risk.

1. INTRODUCTION

The quantification of market risk for derivative pricing, for portfolio choice and for risk management purposes has long been of interest to researchers and financial institutions alike. Ever since the early 1990s, the leading tool for measuring market risk has been the Value at Risk (VaR), see Jorion (2006) and Christoffersen (2009) for comprehensive reviews. VaR summarizes the worst loss over a target horizon that will not be exceeded at a given level of confidence called coverage level. Despite its universality, conceptual simplicity and easy evaluation, VaR has been criticized because of its fundamental deficiencies. VaR does not account for "tail risk". It only tells us the most we can lose if a tail event does not occur; if a tail event does occur, we can expect to lose more than the VaR, but the VaR itself gives us no indication of how much that might be. Other deficiencies of the VaR are its lack of sub–additivity (see Artzner et al. (1997, 1999) and Acerbi and Tasche (2002)) or of convexity (Basak and Shapiro (2001)). These limitations have prompted the implementation of an alternative, coherent, measure of risk — the Expected Shortfall (ES).[1] ES is the expected value of losses beyond a given level of confidence. In its consultative document on the Third Basel Accord, dated May 3, 2012, the Basel Committee explicitly raised the prospect of phasing out VaR and replacing it with the ES (Basel Committee, 2012). The major challenge in the implementation of the ES as the leading measure of market risk is the unavailability of simple tools for its evaluation (see Yamai and Yoshiba (2002, 2005) and Kerkhof and Melenberg (2004)). The main purpose of this article is to propose such tools.

Our proposal is based on the following observation. It is well-known that for each coverage level, violations —the days on which portfolio losses exceed the VaR— should be unpredictable if the risk model is appropriate, i.e. centered violations should be a martingale difference sequence (mds) (see e.g. Berkowitz, Christoffersen and Pelletier (2011)). Indeed, rather than just one mds, centered violations form a class of mds indexed by the coverage level. The integral of the violations over the coverage level in the left tail, which we refer to as *cumulative violations*, also form a mds. The cumulative violation process accumulates all violations in the left tail, just like the ES accumulates the VaR in the left tail. We can therefore use existing methods to check for the mds property (see Escanciano and Lobato

[1] Other names for ES are Conditional VaR, Average VaR, tail VaR or expected tail loss.

(2009a) for a survey of these methods). In particular, we suggest a Box-Pierce test (cf. Box and Pierce (1970)). Our Box-Pierce test is the analogue for ES of the conditional backtests proposed by Christoffersen (1998) and Berkowitz, Christoffersen and Pelletier (2011) for VaR. There are also unconditional implications of the mds property of cumulative violations that can be checked to evaluate ES measures. This leads to the analogue for ES of the unconditional backtest for VaR proposed by Kupiec (1995). See McNeil and Frey (2000), Berkowitz (2001), Kerkhof and Melenberg (2004), Wong (2008, 2010) and Acerbi and Szekely (2014) for other unconditional backtests for ES. In this article, our main focus is on conditional backtests for ES, which, to the best of our knowledge, are not yet available in the literature.[2] However, for completeness we also consider unconditional backtests.

We investigate the asymptotic behavior of our backtests for ES accounting for the presence of estimation risk. Estimation error in VaR/ES forecasts has been studied, among others, in Christoffersen and Goncalves (2005) and more recently in Gourieroux and Zakoian (2013) and Francq and Zakoian (2015). Escanciano and Olmo (2010) investigated estimation risk in backtesting VaR. In the context of backtesting ES, the only study of estimation risk that we are aware of is that of Kerkhof and Melenberg (2004), who considered Historical Simulation ES forecasts for their unconditional backtest. In analogy to what Escanciano and Olmo (2010) do for VaR, we show theoretically and by simulations that not accounting for estimation effects leads to noticeable size distortions in backtesting ES when the in-sample estimation period is not large relative to the out-of-sample evaluation period, which is often the case in practice. To address this limitation of the basic backtests we propose and study modified versions that are robust to the presence of estimation risk, and confirm their robustness through some Monte Carlo simulations.

The main message of this article is that, in contrast with most sentiments expressed in the academic and non-academic literatures, backtesting ES is not more difficult than backtesting VaR.[3] The proposed tests are very easy to implement, they are the natural analogues of those for VaR, and they can be used as part of the toolkit for the internal

[2] Conditional backtests are well-known to be generally more powerful than their unconditional counterparts for commonly used models such as the Filtered Historical Simulation model; see Escanciano and Pei (2012) for a formal explanation in the context of VaR.

[3] Our assessment agrees with that of Kerkhof and Melenberg (2004) and Acerbi and Szekely (2014), among others. See the latter reference for discussion on the possibility of backtesting ES and the concept of "elicitability".

model-based approach suggested by the Basel Committee, thereby leading to a measurement and evaluation of market risk that better captures tail risk.

The remainder of this article is organized as follows. Section 2 introduces some notations used throughout the paper and the building blocks for our backtests: the cumulative violation process. In Section 3 we propose the new unconditional and conditional backtests, and derive their asymptotic properties. Section 4 investigates the finite-sample performance of the proposed backtests through a set of Monte Carlo experiments. In Section 5 we apply our tests to three major stock indexes, the S&P500, the Deutsche German Stock Index (DAX) and the Hang Seng Index, using daily data from the 2008 financial crisis. This empirical application shows that VaR is unresponsive to extreme events such as those experienced during the financial crisis, while ES provides a more accurate description of the risk involved. In Section 6 we conclude. An appendix contains the mathematical proofs of our results.

2. THE CUMULATIVE VIOLATION PROCESS

Let Y_t denote the revenue of a bank at time t, and let Ω_{t-1} denote the risk manager's information at time $t-1$, which contains lagged values of Y_t and possibly lagged values of other variables, say X_t. That is, $\Omega_{t-1} = \{X_{t-1}, X_{t-2}, ...; Y_{t-1}, Y_{t-2}, ...\}$. We assume $\{Y_t, X_t\}_{t=\infty}^{\infty}$ is a strictly stationary and ergodic process, but we note that our results can be extended to some non-stationary and non-ergodic sequences following the results in e.g. Escanciano (2007). Let $G(\cdot, \Omega_{t-1})$ denote the conditional cumulative distribution function (cdf) of Y_t given Ω_{t-1}, i.e. $G(\cdot, \Omega_{t-1}) = \Pr(Y_t \leq \cdot | \Omega_{t-1})$. For simplicity of notation, we drop "almost surely" in all equalities involving random variables. Assume $G(\cdot, \Omega_{t-1})$ is continuous. Let $\alpha \in [0, 1]$ denote the coverage level. The α-level VaR is defined as the quantity $VaR_t(\alpha)$ such that

$$\Pr(Y_t \leq -VaR_t(\alpha) | \Omega_{t-1}) = \alpha. \tag{1}$$

That is, the $-VaR_t(\alpha)$ is the $\alpha - th$ percentile of the distribution G,

$$VaR_t(\alpha) = -G^{-1}(\alpha, \Omega_{t-1}) = -\inf\{y : G(y, \Omega_{t-1}) \geq \alpha\}.$$

Define the α-violation or hit at time t as

$$h_t(\alpha) = 1(Y_t \leq -VaR_t(\alpha)),$$

where $1(\cdot)$ denotes the indicator function. That is, the violation takes the value one if the loss at time t is larger than or equal to $VaR_t(\alpha)$, and it is zero otherwise. An implication of (1) is that violations are Bernoulli variables with mean α, and moreover, centered violations are a *mds* for each $\alpha \in [0, 1]$, i.e.

$$E\left[h_t(\alpha) - \alpha | \Omega_{t-1}\right] = 0 \text{ for each } \alpha \in [0, 1].$$

This restriction has been the basis for the extensive literature on backtesting VaR. Two of its main implications, the zero mean property of the hit sequence $\{h_t(\alpha) - \alpha\}_{t=1}^{\infty}$ and its uncorrelation led to the unconditional and conditional backtests of Kupiec (1995) and Christoffersen (1998), respectively, which are the most widely used backtests.

The VaR has been criticized for its inability to capture "tail risk". This can be seen from the hit sequence $\{h_t(\alpha) - \alpha\}_{t=1}^{\infty}$ itself, which contains information on whether losses are larger than VaR, but not on the actual size of the loss when a violation occurs. This and other limitations of VaR have motivated a move to ES, which, unlike VaR, measures the riskiness of a position by considering both the size and the likelihood of losses beyond a confidence level. ES is defined as the conditional expected loss given that the loss is larger than $VaR_t(\alpha)$, that is,

$$ES_t(\alpha) = E\left[-Y_t | \Omega_{t-1}, -Y_t > VaR_t(\alpha)\right]. \tag{2}$$

Definition of a conditional probability and a change of variables yield a useful representation of $ES_t(\alpha)$ in terms of $VaR_t(\alpha)$,

$$ES_t(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} VaR_t(u) du. \tag{3}$$

Unlike $VaR_t(\alpha)$, which only contains information on one quantile level α, $ES_t(\alpha)$ contains information from the whole left tail, by integrating all VaRs from 0 to α. To test the correct specification of $ES_t(\alpha)$, it seems natural to consider the integral of $h_t(\alpha)$, or the cumulative violation process,

$$H_t(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} h_t(u) du.$$

Since $h_t(u)$ has mean u, by Fubini's Theorem $H_t(\alpha)$ has mean $1/\alpha \int_0^{\alpha} u du = \alpha/2$. Moreover, again by Fubini's Theorem, the *mds* property of the class $\{h_t(\alpha) - \alpha : \alpha \in [0, 1]\}_{t=1}^{\infty}$

is preserved by integration, which means that $\{H_t(\alpha) - \alpha/2\}_{t=1}^{\infty}$ is also a mds. This is the key observation of this article. For computational purposes, it is convenient to define $u_t = G(Y_t, \Omega_{t-1})$. Using that $h_t(u) = 1(Y_t \leq -VaR_t(u)) = 1(u_t \leq u)$, we obtain

$$\begin{aligned} H_t(\alpha) &= \frac{1}{\alpha}\int_0^{\alpha} 1(u_t \leq u)du \\ &= \frac{1}{\alpha}(\alpha - u_t)1(u_t \leq \alpha). \end{aligned} \tag{4}$$

Like violations, cumulative violations are distribution-free, since $\{u_t\}_{t=1}^{\infty}$ comprises a sample of independent and identically distributed (*iid*) $U[0,1]$ variables (see Rosenblatt (1952) for an early use of this property and see also Berkowitz (2001) and Hong and Li (2005) for applications in finance). Acerbi and Tasche (2002) and Emmer, Kratz and Tasche (2014) used the representation in (3) to approximate the integral with a Riemann sum with four terms.[4] Working with violations avoids approximations, as the integral in (4) can be computed exactly. Unlike violations, cumulative violations contain information on the tail risk: when violations are zero, cumulative violations are also zero, but when a violation occurs, the cumulative violation measures how far is the actual value of Y_t from its quantile, through the term $\alpha - u_t = G(G^{-1}(\alpha, \Omega_{t-1}), \Omega_{t-1}) - G(Y_t, \Omega_{t-1})$.[5]

The variables $\{u_t\}_{t=1}^{\infty}$ necessary to construct $\{H_t(\alpha)\}_{t=1}^{\infty}$ are generally unknown, since the distribution of the data G is unknown. In practice, researchers and risk managers specify a parametric conditional distribution $G(\cdot, \Omega_{t-1}, \theta_0)$, where θ_0 is some unknown parameter in $\Theta \subset \mathbb{R}^p$, and proceed to estimate θ_0 before producing VaR/ES forecasts. Popular choices for distributions $G(\cdot, \Omega_{t-1}, \theta_0)$ are those derived from location-scale models with Student's t distributions, but other choices can be certainly entertained in our setting.[6] With the parametric model at hand, we can define the "generalized errors"

$$u_t(\theta_0) = G(Y_t, \Omega_{t-1}, \theta_0)$$

and the associated cumulative violations

$$H_t(\alpha, \theta_0) = \frac{1}{\alpha}(\alpha - u_t(\theta_0))1(u_t(\theta_0) \leq \alpha).$$

[4] See also the related literature that proposes backtesting VaR over a subset of risk levels (see Hurlin and Tokpavi (2006), Perignon and Smith (2008) and Colletaz, Hurlin and Perignon (2013)).

[5] In fact, $d(y,x) = |G(y, \Omega_{t-1}) - G(x, \Omega_{t-1})|$ is a distance function.

[6] We could also extend our methods to semiparametric specifications where θ_0 includes an infinite-dimensional component. We leave this extension for future research.

Very much like for VaRs, the arguments above provide a theoretical justification for backtesting ES by checking whether $\{H_t(\alpha,\theta_0)-\alpha/2\}_{t=1}^{\infty}$ have zero mean (unconditional ES backtest) and whether $\{H_t(\alpha,\theta_0)-\alpha/2\}_{t=1}^{\infty}$ are uncorrelated (conditional ES backtest). We propose test statistics for these hypotheses in the next section.

3. BACKTESTING ES

In this section we propose our backtests for ES. The unconditional backtest is simply a t-test for the hypothesis $E\left[H_t(\alpha,\theta_0)\right]=\alpha/2$, and it is the analogue for the ES of the unconditional VaR backtest proposed in Kupiec (1995). The conditional backtest is a Portmanteau Box-Pierce test applied to sample versions of $H_t(\alpha,\theta_0)$. Our conditional backtest is the analogue for ES of the conditional backtests proposed in Christoffersen (1998) and Berkowitz, Christoffersen and Pelletier (2011) for VaR. We first investigate the asymptotic distributions for our basic backtests allowing for estimation risk, and show that not accounting for it can potentially lead to size distortions. We then propose modified backtests that are robust to estimation risk.

3.1 Basic Unconditional and Conditional Backtests

In practice, the parameters of the model θ_0 are unknown, and they need to be estimated to construct forecasts for ES. For simplicity of presentation we follow here a fixed forecasting scheme, although our theory can be trivially extended to other forecasting schemes (rolling and recursive); see, e.g., Escanciano and Olmo (2010) and references therein for details. That is, the in-sample period $\{Y_{-T+1},\hat{\Omega}_{-T},...,Y_0,\hat{\Omega}_{-1}\}$ of size T is used to estimate θ_0, say by $\widehat{\theta}_T$, where $\widehat{\theta}_T$ is a consistent estimator for θ_0, for example the conditional maximum likelihood estimator (CMLE), and $\hat{\Omega}_{t-1}$ is the observed information set that approximates the infeasible information set Ω_{t-1} (for example by using some initial values for the unobserved infinite past history of the data). With $\widehat{\theta}_T$ we construct residuals

$$\widehat{u}_t = G(Y_t,\hat{\Omega}_{t-1},\widehat{\theta}_T),$$

and estimated cumulative violations

$$\widehat{H}_t(\alpha)=\frac{1}{\alpha}(\alpha-\widehat{u}_t)1(\widehat{u}_t\leq\alpha).$$

Then, an out-of-sample period $\{Y_1, \hat{\Omega}_0, ..., Y_n, \hat{\Omega}_{n-1}\}$ of size n is used to evaluate (backtest) the ES model. Our backtests are based on the estimated cumulative violations.[7]

The unconditional backtest for ES is a standard t-test for the null hypothesis

$$H_{0u} : E(H_t(\alpha, \theta_0)) = \alpha/2.$$

Note that simple calculations show that $E\left[H_t^2(\alpha)\right] = \alpha/3$, and hence, $Var(H_t(\alpha)) = \alpha(1/3 - \alpha/4)$. Therefore, a simple t-test statistic is as follows

$$U_{ES} = \frac{\sqrt{n}\left(\overline{H}(\alpha) - \alpha/2\right)}{\sqrt{\alpha(1/3 - \alpha/4)}}, \tag{5}$$

where $\overline{H}(\alpha)$ denotes the sample mean of $\{\widehat{H}_t(\alpha)\}_{t=1}^n$, i.e.

$$\overline{H}(\alpha) = \frac{1}{n}\sum_{t=1}^{n}\widehat{H}_t(\alpha).$$

Due to the parameter estimation effect, the asymptotic distribution of U_{ES} is generally not a standard normal and depends on the asymptotic relative magnitude of the in-sample (estimation) size T and the out-of-sample (evaluation) size n. Assume both $T \to \infty$ and $n \to \infty$, such that $n/T \to \lambda < \infty$. The next theorem gives the null limit distribution of U_{ES}. Two quantities that appear in this and other asymptotic distributions are the conditional *cdf* of the parametric errors $u_t(\theta)$, i.e.

$$F_t(\theta, x) \equiv \Pr[u_t(\theta) \leq x | \Omega_{t-1}],$$

and the influence function of the estimator $\widehat{\theta}_T$, denoted by l_t; see Assumption A2 in Appendix A for a definition of the influence function. The symbol $\longrightarrow^d$ denotes convergence in distribution, and B' denotes the transpose of the vector or matrix B.

Theorem 1 *Under Assumptions A0-A4 in Appendix A,*

$$U_{ES} \longrightarrow^d N(0, \sigma_\lambda^2(\alpha)),$$

[7] A referee pointed out that a "drawback" of backtests using cumulative violations is that they require more than the ex-ante ES and the ex-post returns series. There are (unconditional) backtests using only these inputs, but they are based on quantities that, unlike cumulative violations, are not distribution-free and require bootstrap methods; see, e.g., McNeil and Frey (2000).

with the asymptotic variance $\sigma^2_\lambda(\alpha)$ *given by*

$$\sigma^2_\lambda(\alpha) = \frac{\alpha(1/3-\alpha/4) + \lambda R'_{ES} E[l_t l_t'] R_{ES}}{\alpha(1/3-\alpha/4)},$$

where $R_{ES} = 1/\alpha \cdot E\left[\int_0^\alpha \left(\partial F_t(\theta_0, x)/\partial\theta\right) dx\right]$.

Remark 1: In the proof of Theorem 1, we actually quantify the parameter estimation effect as follows

$$\sqrt{n}\left(\overline{H}(\alpha) - \frac{\alpha}{2}\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[H_t(\alpha,\theta_0) - \frac{\alpha}{2}\right] + \underbrace{\frac{\sqrt{\lambda}}{\sqrt{n}} R'_{ES} \sum_{t=-T+1}^{0} l_t}_{\text{Estimation Effect}} + o_P(1),$$

which is analogous to what Escanciano and Olmo (2010) do for VaR in their Theorem 1.

In Appendix B, we give an explicit expression for R_{ES} as well as its estimate for a general location-scale model. A special case of Theorem 1 under which knowledge of R_{ES} and l_t is not required is when $\lambda = 0$. Theorem 1 shows that U_{ES} has a standard normal limit distribution when $\lambda = 0$, i.e. when the estimation period is much larger than the evaluation period. In this case, inference with the unconditional backtest is substantially simplified.

Corollary 1 *Under the assumptions of Theorem 1 and* $\lambda = 0$,

$$U_{ES} \longrightarrow^d N(0,1).$$

Next, we consider the conditional backtest with the null hypothesis

$$H_{0c} : E[H_t(\alpha,\theta_0) - \alpha/2 | \Omega_{t-1}] = 0,$$

which is the analogue of the null of conditional backtest for VaR, see e.g. Christoffersen (1998). We need the following notations. Define the lag-j autocovariance and autocorrelation of $H_t(\alpha)$ for $j \geq 0$ by

$$\gamma_j = Cov(H_t(\alpha), H_{t-j}(\alpha)) \text{ and } \rho_j = \frac{\gamma_j}{\gamma_0},$$

respectively. We drop the dependence of γ_j and other related quantities on α for simplicity of notation. The sample counterparts of γ_j and ρ_j based on a sample $\{H_t(\alpha)\}_{t=1}^n$ are

$$\gamma_{nj} = \frac{1}{n-j}\sum_{t=1+j}^{n}(H_t(\alpha) - \alpha/2)(H_{t-j}(\alpha) - \alpha/2) \text{ and } \rho_{nj} = \frac{\gamma_{nj}}{\gamma_{n0}},$$

respectively. Notice that we use the unconditional mean restriction in the definition of autocorrelations. As a result, tests based on γ_{nj} are expected to have power against deviations from H_{0c} where $\{H_t(\alpha)\}$ are uncorrelated but have mean different from $\alpha/2$.[8] In our present context $\{H_t(\alpha)\}_{t=1}^n$ is, however, unobservable, as θ_0 is unknown and Ω_{t-1} is not completely observed. Then, we substitute $\widehat{H}_t(\alpha)$ for $H_t(\alpha)$ in γ_{nj} and obtain

$$\widehat{\gamma}_{nj} = \frac{1}{n-j}\sum_{t=1+j}^{n}(\widehat{H}_t(\alpha)-\alpha/2)(\widehat{H}_{t-j}(\alpha)-\alpha/2) \text{ and } \widehat{\rho}_{nj} = \frac{\widehat{\gamma}_{nj}}{\widehat{\gamma}_{n0}}.$$

Notice that $\rho_j = 0$ for $j \geq 1$ under H_{0c}. Simple conditional tests can be constructed using $\widehat{\rho}_{nj}$, for example the Box-Pierce test statistic

$$C_{ES}(m) = n\sum_{j=1}^{m}\widehat{\rho}_{nj}^2. \tag{6}$$

As with the unconditional backtest, the asymptotic distribution of $C_{ES}(m)$ depends on λ, as well as on other unknown quantities.

Theorem 2 *Under Assumptions A0-A4 in Appendix A,*

$$C_{ES}(m) \longrightarrow^d \sum_{j=1}^{m}\pi_j Z_j^2,$$

where $\{\pi_j\}_{j=1}^m$ are the m eigenvalues of the matrix Σ with the ij–th element given by

$$\Sigma_{ij} = \delta_{ij} + \lambda R_i' E[l_t l_t'] R_j, \tag{7}$$

$$R_j = \frac{1}{\alpha(1/3-\alpha/4)}E\left[(H_{t-j}(\alpha)-\alpha/2)\int_0^{\alpha}\frac{\partial F_t(\theta_0,x)}{\partial\theta}dx\right], \tag{8}$$

δ_{ij} is the Kronecker delta function, which takes value 1 if $i=j$, and 0 otherwise, and $\{Z_j\}$ are independent standard normal variables.

One implication of Theorem 2 is that $C_{ES}(m)$ generally has a weighted chi-squared limit distribution that depends on R_j, and hence on the model and data generating process, in a

[8] Some algebra shows that γ_{nj} roughly equals $1/(n-j)\sum_{t=1+j}^{n}(H_t(\alpha)-\widetilde{H}_t)(H_{t-j}(\alpha)-\widetilde{H}_t)]+(\widetilde{H}_t-\alpha/2)^2$, with $\widetilde{H}_t$ the sample mean of $H_t(\alpha)$. The first term brings power against deviations from zero autocorrelations of $H_t(\alpha)$, while the second term brings power against deviations from H_{0u}. One could also consider tests that do not use the unconditional mean restriction in the definition of autocorrelations (i.e. using the sample mean of cumulative violations). However, Monte Carlo simulations suggest that these tests are less powerful than the ones we use here.

complicated way. However, this limit distribution becomes standard when $\lambda = 0$, as in this case $\pi_j = 1$, for $j = 1, ..., m$.

Corollary 2 *Under the assumptions of Theorem 2 and $\lambda = 0$,*

$$C_{ES}(m) \longrightarrow^d \chi^2_m,$$

where χ^2_m is a chi-square distribution with m degrees of freedom.

Summarizing, our basic unconditional and conditional backtests have standard null limiting distributions when $\lambda = 0$, i.e. when the estimation period is much larger than the evaluation period, and they have nonstandard null limiting distributions when $\lambda > 0$. This implies that these basic tests implemented with critical values from the standard distributions may not be able to properly control for the Type-I error due to estimation effects, unless T is much larger than n (e.g. $T = 2500$ and $n = 250$). To overcome this limitation, we propose in the next section modifications of the basic tests that are robust to estimation risk.

3.2 Robust Unconditional and Conditional Backtests

In this subsection, we propose backtests that explicitly take into account the estimation effects and have standard limit distributions for any λ, $0 \leq \lambda < \infty$. Theorem 1 already gives an expression for the estimation effect of the unconditional backtest (5), which suggests the following modified test statistic

$$MU_{ES} = \frac{\sqrt{n}\left(\overline{H}(\alpha) - \alpha/2\right)}{\sqrt{\alpha(1/3 - \alpha/4) + \frac{n}{T}\widehat{R}'_{ES} W_T \widehat{R}_{ES}}},$$

where

$$\widehat{R}_{ES} = \frac{1}{\alpha n} \sum_{t=1}^{n} \int_0^{\alpha} \frac{\partial \widehat{F}_t(\widehat{\theta}_T, x)}{\partial \theta} dx$$

with $\widehat{F}_t(\theta, x)$ a consistent estimator for $F_t(\theta, x)$, so that $\widehat{R}_{ES} = R_{ES} + o_P(1)$ (the symbol $o_P(1)$ denotes a term that converges in probability to zero), and

$$W_T = \frac{1}{T} \sum_{t=-T+1}^{0} \hat{l}_t \hat{l}'_t,$$

with $\hat{l}_t$ a consistent estimator of l_t, so that $W_T = E[l_t l'_t] + o_P(1)$. That is, W_T is a consistent estimator for the asymptotic variance of the estimator $\widehat{\theta}_T$, which is already implemented in many statistical packages. See Section 4 for expressions for $\widehat{F}_t(\theta, x)$ and $\hat{l}_t$, respectively, in the context of GARCH models estimated by the CMLE method.

Similarly, by Theorem 2 one can modify the conditional backtest (6) as follows

$$MC_{ES}(m) = n\widehat{\rho}_n^{(m)\prime}\widehat{\Sigma}^{-1}\widehat{\rho}_n^{(m)},$$

where $\widehat{\rho}_n^{(m)} = (\widehat{\rho}_{n1}, \widehat{\rho}_{n2}...\widehat{\rho}_{nm})'$ and $\widehat{\Sigma}$ is a consistent estimator for Σ, i.e. $\widehat{\Sigma} = \Sigma + o_P(1)$, with the ij$-th$ element

$$\widehat{\Sigma}_{ij} = \delta_{ij} + \frac{n}{T}\widehat{R}_i' W_T \widehat{R}_j,$$

and

$$\widehat{R}_j = \frac{1}{\alpha(1/3 - \alpha/4)} \frac{1}{n-j} \sum_{t=j+1}^{n} \left\{ \left(\widehat{H}_{t-j}(\alpha) - \frac{\alpha}{2} \right) \int_0^\alpha \frac{\partial \widehat{F}_t(\widehat{\theta}_T, x)}{\partial \theta} dx \right\}.$$

In Appendix B, we give explicit expressions for $\widehat{R}_{ES}$ and $\widehat{R}_j$ for location-scale models. See also Section 4 below.

From Theorem 1 and Theorem 2, it can be shown that the modified test statistics MU_{ES} and $MC_{ES}(m)$ have standard limit distributions regardless of the value of λ.

Corollary 3 *Under the assumptions of Theorem 1 and the consistency of* $\widehat{R}_{ES}$ and W_T,

$$MU_{ES} \longrightarrow^d N(0, 1).$$

Corollary 4 *Under the assumptions of Theorem 2 and the consistency of* $\widehat{\Sigma}$,

$$MC_{ES}(m) \longrightarrow^d \chi_m^2.$$

Unlike in Corollary 1 and Corollary 2 where the limiting distributions are standard only under $\lambda = 0$, in Corollary 3 and Corollary 4 this holds more generally for $0 \leq \lambda < \infty$.

4. MONTE CARLO SIMULATIONS

To assess the finite sample performance of our proposed tests, we carry out some Monte Carlo studies. For comparison purposes, we report the tests results for both ES and VaR. Following Kerkhof and Melenberg (2004) and others, we choose a larger coverage level α for ES than for VaR. Specifically, we consider the following simple rule-of-thumb: choose the coverage level for ES twice (or approximately twice) that of VaR, so that the expected value of violations and cumulative violations coincide (or approximately coincide). Following these arguments, we consider in the simulations $\alpha = 0.1$, 0.05 and 0.025 for ES, corresponding to $\alpha = 0.05$, 0.025 and 0.01 for VaR, respectively. We compare the new unconditional and conditional backtest for ES with the classical ones for VaR, namely

$$U_{VaR} = \frac{\sqrt{n}\left(\overline{h}(\alpha) - \alpha\right)}{\sqrt{\alpha(1-\alpha)}},$$

with $\overline{h}(\alpha)$ the sample average of $\{\widehat{h}_t(\alpha) = 1(\widehat{u}_t \leq \alpha)\}_{t=1}^{n}$; the Box-Pierce-type test for VaR

$$C_{VaR}(m) = n \sum_{j=1}^{m} \widetilde{\rho}_{nj}^{2},$$

with $\widetilde{\rho}_{nj} = \widetilde{\gamma}_{nj}/\widetilde{\gamma}_{n0}$ and $\widetilde{\gamma}_{nj} = 1/(n-j)\sum_{t=1+j}^{n}(\widehat{h}_t(\alpha) - \alpha)(\widehat{h}_{t-j}(\alpha) - \alpha)$. We also report the tests for VaR that are robust to the estimation effects,

$$MU_{VaR} = \frac{\sqrt{n}\left(\overline{h}(\alpha) - \alpha\right)}{\sqrt{\alpha(1-\alpha) + \frac{n}{T}\widetilde{R}_{VaR}' W_T \widetilde{R}_{VaR}}},$$

and

$$MC_{VaR}(m) = n\widetilde{\rho}_{n}^{(m)\prime}\widetilde{\Sigma}^{-1}\widetilde{\rho}_{n}^{(m)},$$

with $\widetilde{\rho}_{n}^{(m)} = (\widetilde{\rho}_{n1}, \widetilde{\rho}_{n2}, ..., \widetilde{\rho}_{nm})'$ and $\widetilde{\Sigma}_{ij} = \delta_{ij} + n/T\widetilde{R}_i' W_T \widetilde{R}_j$. Expressions for $\widetilde{R}_{VaR}$ and $\widetilde{R}_j$ are given in Appendix B (see also Escanciano and Olmo (2010)).

We use the popular AR(1)-GARCH(1,1) specification as our null model for Y_t, under which the VaR and ES are given by

$$\begin{aligned} VaR_t(\alpha) &= -a_0 Y_{t-1} - \sigma_t F_v^{-1}(\alpha),\ \sigma_t^2 = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \text{ and} \\ ES_t(\alpha) &= -a_0 Y_{t-1} - \sigma_t m(\alpha),\ m(\alpha) = E[\varepsilon_t | \varepsilon_t \leq F_v^{-1}(\alpha)], \end{aligned} \tag{9}$$

respectively, where $\varepsilon_t \sim t_v$, a Student's t distribution with unknown degrees of freedom v, with α-quantile denoted by $F_v^{-1}(\alpha)$. The true parameter is set to $\theta_0 = (a_0, \omega_0, \alpha_0, \beta_0) = (0.05, 0.05, 0.1, 0.85)$ and $v = 5$, which are some typical parameter values in empirical applications.

In each simulation, using the in-sample data, we estimate θ_0 and v by the CMLE method. We then obtain $\widehat{u}_t = F_{\widehat{v}}(\varepsilon_t(\widehat{\theta}_T))$, where $F_{\widehat{v}}(\cdot)$ denotes the *cdf* of a Student's t with $\widehat{v}$ degrees of freedom and $f_{\widehat{v}}(\cdot)$ denotes the *pdf*; $\varepsilon_t(\theta) = \varepsilon_t(a, \omega, \alpha, \beta) = (Y_t - aY_{t-1})/\sigma_t(\theta)$ with $\sigma_t^2(\theta) = \omega + \alpha(Y_{t-1} - aY_{t-2})^2 + \beta\sigma_{t-1}^2(\theta)$. Expressions for the quantities needed to compute our modified backtests are given as follows: $\widehat{F}_t(\theta, x) = F_{\widehat{v}}\left((aY_{t-1} - \widehat{a}Y_{t-1} + x\sigma_t(\theta))/\sigma_t(\widehat{\theta}_T)\right)$, and $\hat{l}_t = S^2\partial(\ln f_{\widehat{v}}(\varepsilon_t(\widehat{\theta}_T)) - \ln \sigma_t(\widehat{\theta}_T))/\partial\theta$, with S^2 a consistent estimator for the asymptotic variance of $\sqrt{T}(\widehat{\theta}_T - \theta_0)$. We calculate the test statistics $U_{ES}, U_{VaR}, C_{ES}(m), C_{VaR}(m)$, as well as their modified versions $MU_{ES}, MU_{VaR}, MC_{ES}(m)$ and $MC_{VaR}(m)$ for $m = 1, 3$ and 5. Here we only report the results for $m = 5$.[9] We simply use a $N(0, 1)$ to approximate the limit distributions of U_{ES} and U_{VaR}, and a χ^2_m to approximate the limit distributions of $C_{ES}(m)$ and $C_{VaR}(m)$, although these are good approximations only for small n/T according to our theory in the previous section. We repeat the experiment 1000 times with in-sample sizes $T = 250, 500, 2500$ and out-of-sample sizes $n = 250, 500$.

Our null and alternative data generating processes for Y_t are as follows:

H_0: AR(1)-GARCH(1,1) model:

$$\begin{aligned} Y_t &= 0.05Y_{t-1} + v_t,\ v_t = \sigma_t\varepsilon_t,\ \varepsilon_t \sim t_5 \\ \sigma_t^2 &= 0.05 + 0.1Y_{t-1}^2 + 0.85\sigma_{t-1}^2, \end{aligned} \tag{10}$$

A_1: TAR model: $Y_t = a_tY_{t-1} + \nu_t$, $\nu_t = \sigma_t\varepsilon_t$, $\sigma_t^2 = 0.04 + 0.1\nu_{t-1}^2 + 0.89\sigma_{t-1}^2$, $a_t = 0.7 \cdot 1(\nu_{t-1} \leq -2)$.

A_2: GARCH in Mean model: $Y_t = 2.5\sigma_t^2 + \nu_t$, $\nu_t = \sigma_t\varepsilon_t$, $\sigma_t^2 = 0.01 + 0.29\nu_{t-1}^2 + 0.7\sigma_{t-1}^2$.

A_3: AR(1)-ARCH(2) model: $Y_t = 0.05Y_{t-1} + v_t$, $v_t = \sigma_t\varepsilon_t$, $\sigma_t^2 = 0.1 + 0.1v_{t-1}^2 + 0.8v_{t-2}^2$.

A_4: AR(1)-EGARCH(1,1) model: $Y_t = 0.05Y_{t-1} + v_t$, $v_t = h_t\varepsilon_t$, $lnh_t^2 = 0.01 + 0.9lnh_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$.

A_5: AR(1)-Stochastic Volatility model: $Y_t = 0.05Y_{t-1} + v_t$, $v_t = h_t\varepsilon_t$, $h_t^2 = 0.1Y_{t-1}^2 + \exp(0.98lnh_{t-1}^2 + e_t)$, $e_t \sim iid\ N(0, 1)$.

A_6: AR(1)-GARCH(1,1) model with mixed normal innovations: Y_t is as in (10), with $\varepsilon_t \sim [0.6 \cdot N(1, \sqrt{2}) + 0.4 \cdot N(-1.5, \sqrt{0.75})]/\sqrt{3}$.

In these models $\{\varepsilon_t\}$ is generally iid t_5, unless otherwise specified. Similar models are studied in Escanciano and Velasco (2010) and Escanciano and Olmo (2010). In A_1 and A_2,

[9] The simulation results for $m = 1$ and 3 are available from the authors upon request.

only the conditional mean is incorrectly specified while the other aspects of the distribution are correctly specified. In A_3, A_4 and A_5, only the conditional variance is incorrectly specified. In A_6, only the distribution of the innovations $\{\varepsilon_t\}$ is incorrectly specified.

Tables 1-9 give the empirical sizes and size-corrected powers of the tests at 5% nominal level. Consistent with our theory in the previous section, the basic tests have severe size distortions for small T due to the estimation effects, but these size distortions reduce significantly for large T, such as $T = 2500$. The modified tests have satisfactory sizes even for small T, with the exception of the conditional backtest for VaR, $MC_{VaR}(5)$, which shows some over-rejections. The size distortions are bigger for backtesting $ES_t(0.025)$ and $VaR_t(0.01)$, as there are fewer observations in the extreme tail. For a given in-sample size T, the size performance of our modified backtests for ES, MU_{ES} and $MC_{ES}(5)$, generally becomes better as the out-of-sample size n increases from 250 to 500, in contrast to what happens with the basic tests (as expected, since the ratio n/T is larger).

Due to the size distortions for small T, our reported powers are all size-corrected. The unconditional backtest for ES detects well alternatives A_2 and A_6, while the conditional backtest for ES has moderate power against alternatives $A_2 - A_5$ and high power against A_1 and A_6. In most cases, especially A_2 and A_6, $C_{ES}(5)$ and $MC_{ES}(5)$ have higher power than $C_{VaR}(5)$ and $MC_{VaR}(5)$. Finally, we observe that the power for MU_{ES} and $MC_{ES}(5)$ increases with the out-of-sample size n, suggesting that these tests are consistent for these alternatives. Unreported simulations confirm that these conclusions are also valid for other choices of innovations' distributions, including Hansen Skewed t distribution (see Hansen (1994)) with time-varying higher order moments.[10]

TABLES 1-9 ABOUT HERE

The main conclusions from our Monte Carlo analysis are the following. There are significant size distortions for the basic tests for small values of T, such as $T = 250$, but these distortions decrease substantially for large values of T, such as $T = 2500$ (see, e.g., Table 7). In contrast, robust tests have uniformly good empirical size. However, this improvement in size comes at a computational cost, since the test statistic needs to be modified accordingly with the model and estimator used. For this reason, we recommend practitioners to use the basic tests with large values of the in-sample size T (e.g. $T = 2500$, $n = 250$), and use the robust

[10] These simulations are available from the authors upon request.

backtests if their in-sample size is small or moderate (e.g. $T = 250$, $n = 250$).

The simulations also suggest that unconditional and conditional tests are complementary rather than substitutes. Therefore, we recommend to use both in practice. Also, in these simulations backtests for ES compare favorably to those for VaR. The next section provides further empirical evidence for this comparison in the context of real data on three major stock indexes during the recent financial crisis.

5. EMPIRICAL APPLICATION

In this section we illustrate with an empirical application to three major stock indexes the advantages of using ES as a measure of market risk in periods of financial turmoil, such as those experienced during the recent financial crisis. Based on our new tools, we provide empirical evidence showing that VaR is not responsive to extreme events during the financial crisis, as measured by traditional VaR backtests with regulatory coverage levels, while the new ES backtests are able to reject the validity of forecasts for one of the most commonly used models of risk, an AR(1)-GARCH(1,1) model with Student's t innovations. Our empirical results complement those of Kourouma, Dupre, Sanfilippo and Taramasco (2011) and O'Brien and Szerszen (2014), who focussed on the evaluation of the performance of classical backtests during the financial crisis for stock indexes and five major US banks, respectively. In contrast, here we confront these classical VaR backtests with the newly proposed ES backtests.[11]

We consider the daily S&P500 Index, the DAX and the Hang Seng Index (HS), three of the major stock indexes in the world. Our data are obtained from *finance.yahoo.com* over the period January 1, 1997 - June 30, 2009. Table 10 presents the descriptive statistics for the series for the in-sample and out-of-sample periods. The in-sample period in our analysis is from January 1, 1997 to June 30, 2007, and the out-of-sample period is July 1st, 2007 - June 30, 2009, the financial crisis period. Generally, the returns are leptokurtic and very volatile with big losses, especially during the crisis. Excess kurtosis is evident in all three indexes and dramatically so in the Hang Seng due to the turmoil right after the return of Hong Kong to China in 1997. The data are plotted in Figure 1. One can

[11] The R code to replicate the empirical application of this paper is available at the portal Run My Code (http://www.runmycode.org/).

observe the volatility clustering feature of the data. Moreover, the returns are more volatile during the financial crisis period, particularly for the S&P500 which lost almost half of its value between July 2008 and the market bottom in March 2009. Besides, there are some volatile periods for S&P500 and DAX in 2001-2003 during the September 11 Attack and stock market downturn; Hong Kong, on the other hand, experienced a volatile period in late 1997.

TABLE 10 ABOUT HERE

FIGURE 1 ABOUT HERE

We fit an AR(1)-GARCH(1,1) model with Student's t innovations to the log-return Y_t. The implied VaR and ES at level α are given by

$$\begin{aligned} VaR_t(\alpha) &= -a_0 Y_{t-1} - \sigma_t F_v^{-1}(\alpha),\ \sigma_t^2 = \omega_0 + \alpha_0 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \text{ and} \\ ES_t(\alpha) &= -a_0 Y_{t-1} - \sigma_t m(\alpha),\ m(\alpha) = E[\varepsilon_t | \varepsilon_t \leq F_v^{-1}(\alpha)], \end{aligned}$$

respectively, as defined in (9).

We estimate the parameters by CMLE using the in-sample data, and perform backtesting with the out-of-sample data. Table 11 reports the CMLE estimates, including estimates for the terms $F_v^{-1}(\alpha)$ and $m(\alpha)$ for the levels of α considered.[12] We observe a similar high level of volatility persistence for the three indexes. Hang Seng has the smallest Student's t degree of freedom v of the innovation distribution, and hence a fatter tail, in agreement with the high kurtosis of Hang Seng observed in Table 10.

TABLE 11 ABOUT HERE

Figures 2-4 plot the estimated $VaR_t(0.05)$ and $ES_t(0.1)$ for the three series, respectively. One can see that $-ES_t(0.1)$ is smaller than $-VaR_t(0.05)$. When Y_t falls below $-VaR_t(0.05)$, $-ES_t(0.1)$ is closer to the true Y_t compared with $-VaR_t(0.05)$. Take September 15, 2008 for example, when Lehman Brothers filed bankruptcy. S&P500 fell by 4.83% on that day, and the estimated $VaR_t(0.05)$ is 1.82%, while $ES_t(0.1)$ is 2.65%, closer to the actual loss.

[12] Here we treat the degree of freedom parameter v as discrete and unknown. Then the estimation effect only stems from the uncertainty of estimating the location and scale parameters. See Escanciano and Olmo (2010) for similar findings.

There are 41 cases out of 504 observations over the period July 1st, 2007 - June 30, 2009 where S&P500 returns fall below their $-VaR_t(0.05)$. The average loss of S&P500 for those cases is 3.82%, and the average of $VaR_t(0.05)$ is 2.79% while that of $ES_t(0.1)$ is 3.07%. There are 11 cases where S&P500 returns fall below their $-VaR_t(0.01)$. The average loss of S&P500 for those cases is 3.76%, and the average of $VaR_t(0.01)$ is 3.13% while that of $ES_t(0.025)$ is 3.20%. Therefore, ES better describes the extreme losses than VaR. The results for DAX and Hang Seng tell similar stories.

FIGURES 2-4 ABOUT HERE

Table 12 reports the number of violations $\left(V(\alpha) = \sum_{t=1}^{n} \widehat{h}_t(\alpha)\right)$, cumulative violations $\left(CV(\alpha) = \sum_{t=1}^{n} \widehat{H}_t(\alpha)\right)$ and the expected value of violations $(n\alpha)$ for the three indexes in the pre-crisis and crisis periods. The pre-crisis period is here defined as the end of in-sample period with the same number of observations as the out-of-sample crisis period. Comparing $V(\alpha)$ and $CV(\alpha)$ before and after the crisis, one can see a significant increase of risk in the crisis in general. One exception is $V(0.01)$ of DAX, which actually drops in the crisis, while $CV(0.025)$ of DAX does increase in the crisis. If we take a further look, although the number of violations $V(0.01)$ drops, the losses are much larger in the crisis period than the pre-crisis period. This explains the increase of $CV(0.025)$ and the decrease of $V(0.01)$ in the crisis period. Table 12 shows significant discrepancies between violations and cumulative violations at the coverage level suggested by the Basel committee ($\alpha = 0.01$ for VaR).

TABLE 12 ABOUT HERE

Figure 5 plots the cumulative violations $\left\{\widehat{H}_t(0.1)\right\}$ of the three indexes in the out-of-sample crisis period. We observe large values of $\widehat{H}_t(0.1)$, which indicates a large loss on that day. As we can see there are more such cases for S&P500 than DAX and Hang Seng. For the three indexes there is substantial clustering of cumulative violations, which suggests deviations from the *mds* hypothesis implied by an appropriate ES forecast. To formally assess this hypothesis we apply our conditional backtest.

FIGURE 5 ABOUT HERE

Table 13 reports the p-values of the unconditional tests U_{ES}, U_{VaR}, MU_{ES}, MU_{VaR} and the conditional tests $C_{ES}(5)$, $C_{VaR}(5)$, $MC_{ES}(5)$, $MC_{VaR}(5)$ for the three indexes, respectively. Our conditional backtests for ES, $C_{ES}(5)$ and $MC_{ES}(5)$, generally reject the null model, while $C_{VaR}(5)$ and $MC_{VaR}(5)$ do not. Figures 6 and 7 plot the sample autocorrelations of $\widehat{H}_t(0.025)$ and $\widehat{h}_t(0.01)$, respectively, from which we can also clearly see that the model for $ES_t(0.025)$ is rejected at 5% level, while the model for $VaR_t(0.01)$ is not.

Figure 7 shows insignificant autocorrelations of $\left\{\widehat{h}_t(0.01)\right\}$ for all three series. The autocorrelations of $\widehat{h}_t(0.01)$ of DAX and Hang Seng are actually very close to 0 for the first twelve lags, as there are only five Y_t's falling below $-VaR_t(0.01)$ for those two indexes. The corresponding number for S&P500 is eleven. The crisis originated and had a bigger impact in the US, which brought about more extreme losses in the stock market in the US than in Germany and Hong Kong. This also may explain why the unconditional test U_{VaR} in Table 13 has a small p-value for S&P500, and a big p-value for the other two indexes.

The cumulative violations $\left\{\widehat{H}_t(0.025)\right\}$, on the other hand, have significant autocorrelations for all three series. The number of extreme losses may not be big, but the average losses can be big and highly correlated. The cumulative violations series $\left\{\widehat{H}_t(0.025)\right\}$ take both of those two pieces of information into account. On the contrary, conditional VaR backtests only look at clustering of tail events, and not to their magnitude. Therefore, as reported in Table 13, our test based on $\left\{\widehat{H}_t(0.025)\right\}$ is able to better detect the problems of one of the most commonly used risk models during the 2008 financial crisis.

TABLE 13 ABOUT HERE

FIGURES 6-7 ABOUT HERE

In summary, based on VaR backtesting, one cannot find unambiguous empirical evidence against the AR-GARCH model with Student's t innovations at the regulatory coverage level, and hence adjust the way the reserved capital is calculated during the financial crisis period. Instead, if one uses ES as the risk measure, our proposed backtesting procedure clearly rejects this model. Our empirical analysis here confirms that the theoretical advantages of ES over VaR documented in Artzner et al. (1997, 1999) also have empirical manifestations in the context of backtesting market risk. We have provided in this article a set of tools based on cumulative violations that help assess not only the likelihood of financial losses but also the size of such losses.

6. CONCLUSIONS

Despite the substantial theoretical evidence documenting the superiority of ES over VaR as a measure of risk, only recently ES has been embraced by financial institutions and regulators as an alternative to VaR for financial risk management. Arguably, one of the major obstacles in this transition has been the unavailability of simple tools for the evaluation of ES forecasts (backtests). In this article, we have introduced cumulative violations as the building blocks for constructing unconditional and conditional backtests for ES, much like violations are the building blocks for the most commonly used backtests for VaR. Unlike violations, cumulative violations contain information on the tail risk and, therefore, provide a more complete description of the risk involved.

We have proposed basic unconditional and conditional backtests as well as modified backtests that are robust to the presence of estimation risk. The unconditional backtests check for the mean of cumulative violations with a t-test. The proposed conditional backtests are Portmanteau tests applied to estimated cumulative violations. We also recommend to complement the information provided by formal tests with graphical tools such as the plot of cumulative violations and autocorrelograms of cumulative violations. Our conditional backtest involves two choices that practitioners need to make: the coverage level α and the number of autocorrelations considered m. We have suggested choices for α such as $\alpha = 0.1$, 0.05 and 0.025 for ES. Smaller values are not recommended, as they would require very large out-of-sample sizes to achieve a satisfactory approximation of the finite sample distribution by the asymptotic distribution of tests. Regarding the choice of the number of correlations, we have suggested to apply the test with $m = 5$. A sensible alternative, however, is to consider a data-driven choice of m similar to that proposed in Escanciano and Lobato (2009b). This combined procedure has been shown to deliver simple and reliable inferences in other contexts, and it can be certainly used here to provide a fully data-driven backtests for ES at a small computational price.

7. APPENDIX

Appendix A: Assumptions

This section introduces the assumptions and some formulae needed for our results and tests. We first introduce some notations. Let $\| \cdot \|$ denote the Euclidean norm, and let C be a generic constant that may change from expression to expression.

For completeness, we shall present a more general version of our results where a generic transformation $\varphi(u_t)$ is considered, where $\varphi \in \Psi$, and Ψ is the class of measurable functions $\varphi : [0,1] \to \mathbb{R}$, which are right continuous with left limits (*cadlag*), of bounded variation or non-decreasing. The case of cumulative violations corresponds to the special case

$$\varphi(u_t) = \frac{1}{\alpha}(\alpha - u_t)1(u_t \leq \alpha). \tag{11}$$

In the results of the main text we refer to the assumptions below holding for this specific choice of φ, but the results in this appendix are shown for a general $\varphi \in \Psi$. In the sequel, we simplify the notations as follows: $u_t(\theta) = G(Y_t, \Omega_{t-1}, \theta)$, $u_t \equiv u_t(\theta_0)$, $c_\varphi = E[\varphi(u_t)]$ and $v_\varphi = var(\varphi(u_t))$. In particular, for φ in (11) let $c_{ES} = \alpha/2$ and $v_{ES} = \alpha(1/3 - \alpha/4)$ denote the corresponding mean and variance, respectively. Let Θ_0 be an arbitrary neighborhood of $\theta_0 \in \Theta$. Consider the following assumptions.

Assumption A0: The conditional distribution of Y_t given Ω_{t-1} is given by $G(\cdot, \Omega_{t-1}, \theta_0)$.

Assumption A1: $\{Y_t, X_t\}_{t=-T+1}^{n}$ is strictly stationary and ergodic.

Assumption A2: The estimator $\widehat{\theta}_T$ is $\sqrt{T}$-consistent for θ_0, where θ_0 is in the interior of Θ. Moreover, $\widehat{\theta}_T$ satisfies the following asymptotic (Bahadur) expansion,

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=-T+1}^{0} l_t + o_p(1),$$

where l_t is such that $E\left[l_t \mid \Omega_{t-1}\right] = 0$ and $E[l_t l_t']$ exists and is positive definite.

Assumption A3: The effect of information truncation satisfies

$$\sup_{\theta \in \Theta_0} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| \varphi(G(Y_t, \hat{\Omega}_{t-1}, \theta)) - \varphi(G(Y_t, \Omega_{t-1}, \theta)) \right| = o_P(1).$$

Assumption A4: $F_t(\theta, x)$ is continuously differentiable in θ and $x \in [0,1]$ a.s. Moreover, $v_\varphi < \infty$,

$$E\left[\sup_{\theta\in\Theta_0, 0\le x\le 1}\left\|\frac{\partial F_t(\theta,x)}{\partial x}\right\|\right] < C \text{ and } E\left[\int_0^1 \sup_{\theta\in\Theta_0}\left\|\frac{\partial F_t(\theta,x)}{\partial \theta}\right\| d\varphi(x)\right] < C.$$

Assumption A0 is standard in the literature, and it assumes the model is correctly specified. It can be relaxed to the condition

$$P\left(Y_t \le y|\Omega_{t-1}\right) = G(y, \Omega_{t-1}, \theta_0) \text{ for all } y \le G^{-1}(\alpha, \Omega_{t-1}, \theta_0),$$

without changing the theory of this article. Assumption A1 is made here for easy exposition. Our results can be extended to some non-stationary and non-ergodic sequences, see e.g. Escanciano (2007). Assumption A2 is satisfied by most commonly used estimators, such as the (quasi-)maximum likelihood estimator and the generalized method of moments estimator, see e.g. Bose (1998) and Wu (2007). Assumption A3 is on the effect of information truncation due to the unavailability of the infinite history of observations, and it holds for many time series models with finite second moment, including stationary and invertible ARMA processes, GARCH processes etc., see e.g. the discussions in Bai (2003) and Hong and Lee (2003). This assumption is not needed when the process is Markovian. Assumption A4 is required for the asymptotic equicontinuity[13] of certain empirical processes and the uniform law of large numbers.

Appendix B: Expressions for the Estimation Effects

Here we give explicit expressions for $\widehat{R}_{ES}$, $\widehat{R}_j$, $\widetilde{R}_{VaR}$ and $\widetilde{R}_j$ for the general location-scale model

$$Y_t = \mu_t + \sigma_t \varepsilon_t, \tag{12}$$

where $\mu_t = \mu(\Omega_{t-1}, \theta_0) = E[Y_t \mid \Omega_{t-1}]$, $\sigma_t^2 = \sigma^2(\Omega_{t-1}, \theta_0) = Var[Y_t \mid \Omega_{t-1}]$; and ε_t follows a distribution with *cdf* $G_\varepsilon(\cdot)$ and density function $g_\varepsilon(\cdot)$.

We have

$$\frac{\partial F_t(\theta_0, x)}{\partial \theta} = g_\varepsilon(G_\varepsilon^{-1}(x))\frac{\dot{\mu}_t + G_\varepsilon^{-1}(x)\dot{\sigma}_t}{\sigma_t}, \tag{13}$$

[13]For definition of asymptotic (uniform) equicontinuity see Chapter 1.5 in van der Vaart and Wellner (1996).

with $\dot{\mu}_t = \partial\mu(\Omega_{t-1}, \theta_0)/\partial\theta$ and $\dot{\sigma}_t = \partial\sigma(\Omega_{t-1}, \theta_0)/\partial\theta$. Therefore,

$$
\begin{aligned}
R_{ES} &= \frac{1}{\alpha}E\left\{\int_0^\alpha \frac{\partial F_t(\theta_0, x)}{\partial\theta}dx\right\} \\
&= \frac{1}{\alpha}E\left\{\int_0^\alpha g_\varepsilon(G_\varepsilon^{-1}(x))\frac{\dot{\mu}_t + G_\varepsilon^{-1}(x)\dot{\sigma}_t}{\sigma_t}dx\right\} \\
&= \frac{1}{\alpha}E\left\{\int_{G_\varepsilon^{-1}(0)}^{G_\varepsilon^{-1}(\alpha)} g_\varepsilon(z)\frac{\dot{\mu}_t + z\dot{\sigma}_t}{\sigma_t}dG_\varepsilon(z)\right\} \\
&= \frac{1}{\alpha}E\left\{g_\varepsilon(\varepsilon_t)1(\varepsilon_t \le G_\varepsilon^{-1}(\alpha))\frac{\dot{\mu}_t + \varepsilon_t\dot{\sigma}_t}{\sigma_t}\right\},
\end{aligned}
$$

and

$$
\widehat{R}_{ES} = \frac{1}{\alpha n}\sum_{t=1}^n g_\varepsilon(\widehat{\varepsilon}_t)1(\widehat{\varepsilon}_t \le G_\varepsilon^{-1}(\alpha))\frac{\dot{\widehat{\mu}}_t + \widehat{\varepsilon}_t\dot{\widehat{\sigma}}_t}{\widehat{\sigma}_t},
$$

with $\dot{\widehat{\mu}}_t = \partial\mu(\Omega_{t-1}, \widehat{\theta}_T)/\partial\theta$, $\dot{\widehat{\sigma}}_t = \partial\sigma(\Omega_{t-1}, \widehat{\theta}_T)/\partial\theta$ and $\widehat{\sigma}_t = \sigma(\Omega_{t-1}, \widehat{\theta}_T)$. Moreover,

$$
\begin{aligned}
R_j &= \frac{1}{v_{ES}(\alpha)}E\left\{(H_{t-j}(\alpha) - \alpha/2)\int_0^\alpha g_\varepsilon(G_\varepsilon^{-1}(x))\frac{\dot{\mu}_t + G_\varepsilon^{-1}(x)\dot{\sigma}_t}{\sigma_t}dx\right\} \\
&= \frac{1}{v_{ES}(\alpha)}E\left\{(H_{t-j}(\alpha) - \alpha/2)\int_{G_\varepsilon^{-1}(0)}^{G_\varepsilon^{-1}(\alpha)} g_\varepsilon(z)\frac{\dot{\mu}_t + z\dot{\sigma}_t}{\sigma_t}dG_\varepsilon(z)\right\} \\
&= \frac{1}{v_{ES}(\alpha)}E\left\{(H_{t-j}(\alpha) - \alpha/2)g_\varepsilon(\varepsilon_t)1(\varepsilon_t \le G_\varepsilon^{-1}(\alpha))\frac{\dot{\mu}_t + \varepsilon_t\dot{\sigma}_t}{\sigma_t}\right\},
\end{aligned}
$$

whose feasible counterpart is given by

$$
\widehat{R}_j = \frac{1}{v_{ES}(\alpha)}\frac{1}{n-j}\sum_{t=j+1}^n (\widehat{H}_{t-j}(\alpha) - \alpha/2)g_\varepsilon(\widehat{\varepsilon}_t)1(\widehat{\varepsilon}_t \le G_\varepsilon^{-1}(\alpha))\frac{\dot{\widehat{\mu}}_t + \widehat{\varepsilon}_t\dot{\widehat{\sigma}}_t}{\widehat{\sigma}_t}.
$$

Similarly, for the modified backtests for VaR, we have

$$
\widetilde{R}_{VaR} = \frac{g_\varepsilon(G_\varepsilon^{-1}(\alpha))}{n}\sum_{t=1}^n \frac{\dot{\widehat{\mu}}_t + G_\varepsilon^{-1}(\alpha)\dot{\widehat{\sigma}}_t}{\widehat{\sigma}_t},
$$

and

$$
\widetilde{R}_j = \frac{-g_\varepsilon(G_\varepsilon^{-1}(\alpha))}{(n-j)\alpha(1-\alpha)}\sum_{t=j+1}^n \frac{\dot{\widehat{\mu}}_t + G_\varepsilon^{-1}(\alpha)\dot{\widehat{\sigma}}_t}{\widehat{\sigma}_t}(1(\widehat{u}_{t-j} \le \alpha) - \alpha).
$$

Appendix C: Proofs

Here we prove more general versions of Theorems 1 and 2 for a generic transformation $\varphi(u_t)$, for $\varphi \in \Psi$, which includes $H(\cdot)$ and $h(\cdot)$ as special cases. These more general results are of independent interest, and they can be used to develop backtests for other coherent risk measures different from ES.

THEOREM A1: *Under Assumptions A0-A4,*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(\widehat{u}_t)-c_\varphi)\longrightarrow^{d} N(0,\sigma_\varphi^2),$$

where

$$\begin{aligned}\sigma_\varphi^2 &= v_\varphi+\lambda R_\varphi^{'}E[l_t l_t^{'}]R_\varphi,\\ R_\varphi &= -E\left\{\int_0^1\frac{\partial F_t(\theta_0,x)}{\partial\theta}d\varphi(x)\right\}.\end{aligned}$$

PROOF OF THEOREM A1: We first consider the case of no information truncation. This occurs if G depends only on a finite number of lagged Y_t and X_t.

Similar arguments as the proof for Theorem 1 in Escanciano and Olmo (2010) show that under Assumptions A0-A2 and A4

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\varphi(\widehat{u}_t)-E[\varphi(\widehat{u}_t)|\Omega_{t-1}]\right]-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\varphi(u_t)-E[\varphi(u_t)|\Omega_{t-1}]\right]=o_P(1).$$

Therefore,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(\widehat{u}_t)-c_\varphi) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(u_t)-c_\varphi)+\\ &\quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{\theta}_T-\theta_0)'\frac{\partial E[\varphi(u_t(\widetilde{\theta}_T))|\Omega_{t-1}]}{\partial\theta}+o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(u_t)-c_\varphi)+\\ &\quad \sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T-\theta_0)'\frac{1}{n}\sum_{t=1}^{n}\frac{\partial E[\varphi(u_t(\widetilde{\theta}_T))|\Omega_{t-1}]}{\partial\theta}+o_P(1),\end{aligned}$$

with $\widetilde{\theta}_T$ an intermediate point between $\widehat{\theta}_T$ and θ_0.

Notice that

$$\begin{aligned}\frac{\partial E[\varphi(u_t(\theta))|\Omega_{t-1}]}{\partial\theta} &= \frac{\partial\left(\int_0^1\varphi(x)dF_t(\theta,x)\right)}{\partial\theta}\\ &= -\frac{\partial\left(\int_0^1 F_t(\theta,x)d\varphi(x)\right)}{\partial\theta}.\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(\widehat{u}_t)-c_\varphi) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(u_t)-c_\varphi)+\sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T-\theta_0)'R_\varphi+o_P(1)\\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(u_t)-c_\varphi)+\sqrt{\lambda}R'_\varphi\frac{1}{\sqrt{T}}\sum_{t=-T+1}^{0}l_t+o_P(1).
\end{aligned}$$

Notice that the first term on the right hand side converges in distribution to $N(0,\nu_\varphi)$, and the covariance between the first two terms are 0 as the summand in the first term is for out-of-sample observations and the second term is for in-sample observations. These, together with the above display, imply Theorem A1.

Next we consider the case of information truncation. Define $\widetilde{u}_t = G(Y_t,\Omega_{t-1},\widehat{\theta}_T)$, and then we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(\widehat{u}_t)-c_\varphi)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(\widehat{u}_t)-\varphi(\widetilde{u}_t))+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi(\widetilde{u}_t)-c_\varphi).$$

Assumption A3 implies that the first term on the right hand side is $o_p(1)$. Then notice that the arguments above for $1/\sqrt{n}\sum_{t=1}^{n}(\varphi(\widehat{u}_t)-c_\varphi)$ without information truncation can be applied directly to $1/\sqrt{n}\sum_{t=1}^{n}(\varphi(\widetilde{u}_t)-c_\varphi)$, which completes the proof of Theorem A1.

■

Proofs of Theorem 1, Corollary 1 and Corollary 3: The proofs follow directly from Theorem A1. ■

Next we prove a more general version of Theorem 2, for which we need the following lemmas. Define the processes

$$\begin{aligned}
R_{nj}(x,y) &= \frac{1}{n-j}\sum_{t=1+j}^{n}\{1(u_t\le x)-x\}\{1(u_{t-j}\le y)-y\},\\
\widehat{R}_{nj}(x,y) &= \frac{1}{n-j}\sum_{t=1+j}^{n}\{1(\widehat{u}_t\le x)-x\}\{1(\widehat{u}_{t-j}\le y)-y\}.
\end{aligned}$$

Lemma A1: *Under Assumptions A0-A4, we have*

$$\sup_{0\le x\le 1,0\le y\le 1}\left|\sqrt{n-j}[\widehat{R}_{nj}(x,y)-R_{nj}(x,y)]-\sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T-\theta_0)'E_j(x,y)\right|=o_p(1),$$

where

$$E_j(x,y)=E\left\{\frac{\partial F_t(\theta_0,x)}{\partial\theta}[I(u_{t-j}\le y)-y]\right\}.$$

Lemma A1 is a special case of Theorem 1 in Du (2015), and hence, its proof is omitted.

LEMMA A2: *Let $R(x, y)$ be a function defined on $[0,1]^2$ such that $R(\cdot, y) \in \Psi$ for $0 \le y \le 1$, $R(x, \cdot) \in \Psi$ for $0 \le x \le 1$ and $R = 0$ on the boundaries. Denote by $\ell([0,1]^2)$ the metric space of all such functions endowed with the supremum norm. Then the mapping*

$$R \to \int_0^1 \int_0^1 \varphi(x)\varphi(y)R(dx, dy)$$

is continuous in R for any $\varphi \in \Psi$.

PROOF OF LEMMA A2: By the Integration by Parts Theorem (Theorem 11, Shiryaev 1996, pp. 206) and the definition of R, we have

$$\int_0^1 \int_0^1 \varphi(x)\varphi(y)R(dx, dy) = \int_0^1 \int_0^1 R(x, y)\varphi(dx)\varphi(dy).$$

Noticing that

$$\begin{aligned} & \left| \int_0^1 \int_0^1 R_1(x, y)\varphi(dx)\varphi(dy) - \int_0^1 \int_0^1 R_2(x, y)\varphi(dx)\varphi(dy) \right| \\ \le \quad & \sup |R_1(x, y) - R_2(x, y)| \int_0^1 \int_0^1 |\varphi(dx)\varphi(dy)|, \end{aligned}$$

for any $R_1, R_2 \in \ell([0,1]^2)$, and $\int |\varphi(dx)| < \infty$ as $\varphi \in \Psi$, the proof is complete. ■

With the above two lemmas in place, we are ready to prove a more general version of Theorem 2. Define the lag-j autocovariance and autocorrelation of $\varphi(u_t)$ for $j \ge 0$ by

$$\gamma_j = Cov(\varphi(u_t), \varphi(u_{t-j})) \text{ and } \rho_j = \frac{\gamma_j}{\gamma_0},$$

respectively. The sample counterparts of γ_j and ρ_j based on a sample $\{u_t\}_{t=1}^n$ are

$$\gamma_{nj} = \frac{1}{n-j} \sum_{t=1+j}^{n} (\varphi(u_t) - c_\varphi)(\varphi(u_{t-j}) - c_\varphi) \text{ and } \rho_{nj} = \frac{\gamma_{nj}}{\gamma_{n0}},$$

respectively. As $\{u_t\}_{t=1}^n$ is unobservable, we substitute $\widehat{u}_t$ for u_t in γ_{nj} and obtain

$$\widehat{\gamma}_{nj} = \frac{1}{n-j} \sum_{t=1+j}^{n} (\varphi(\widehat{u}_t) - c_\varphi)(\varphi(\widehat{u}_{t-j}) - c_\varphi) \text{ and } \widehat{\rho}_{nj} = \frac{\widehat{\gamma}_{nj}}{\widehat{\gamma}_{n0}}.$$

THEOREM A2: *Under Assumptions A0-A4,*

$$\sqrt{n}\widehat{\rho}_n^{(m)} \longrightarrow^d N(0, \Sigma)$$

with the ij−th element of Σ *given by*

$$\Sigma_{ij} = \delta_{ij} + \lambda R_i' E[l_t l_t'] R_j,$$

where

$$R_j = \frac{-1}{v_\varphi} E\left\{ (\varphi(u_{t-j}) - c_\varphi) \int_0^1 \frac{\partial F_t(\theta_0, x)}{\partial \theta} d\varphi(x) \right\}, \tag{14}$$

and δ_{ij} *is the Kronecker delta function, which takes value 1 if* $i = j$, *and 0 otherwise.*

Proof of Theorem A2: We first consider the case of no information truncation. This occurs if G depends only on a finite number of lagged Y_t and X_t.

Notice that

$$\begin{aligned} \sqrt{n-j}\widehat{\gamma}_{nj} &= \sqrt{n-j} \int_0^1 \int_0^1 \varphi(x)\varphi(y)\widehat{R}_{nj}(dx, dy), \\ \sqrt{n-j}\gamma_{nj} &= \sqrt{n-j} \int_0^1 \int_0^1 \varphi(x)\varphi(y)R_{nj}(dx, dy). \end{aligned}$$

By Lemma A1 and A2, we have

$$\sqrt{n-j}\widehat{\gamma}_{nj} = \sqrt{n-j}\gamma_{nj} + \sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T - \theta_0)' \int_0^1 \int_0^1 \varphi(x)\varphi(y)E_j(dx, dy) + o_p(1), \tag{15}$$

where

$$\begin{aligned} \int_0^1 \int_0^1 \varphi(x)\varphi(y)E_j(dx, dy) &= \int_0^1 \int_0^1 E\left\{ \varphi(x) \frac{\partial^2 F_t(\theta_0, x)}{\partial x \partial \theta} dx \varphi(y)[I(u_{t-j} \le dy) - dy] \right\} \\ &= \int_0^1 E\left\{ \varphi(x) \frac{\partial^2 F_t(\theta_0, x)}{\partial x \partial \theta} dx (\varphi(u_{t-j}) - c_\varphi) \right\} \\ &= E\left\{ \int_0^1 \varphi(x) \frac{\partial^2 F_t(\theta_0, x)}{\partial x \partial \theta} dx \ (\varphi(u_{t-j}) - c_\varphi) \right\} \\ &= E\left\{ \left[\varphi(x) \frac{\partial F_t(\theta_0, x)}{\partial \theta} \Big|_0^1 - \int_0^1 \frac{\partial F_t(\theta_0, x)}{\partial \theta} d\varphi(x) \right] \right. \\ &\qquad \left. (\varphi(u_{t-j}) - c_\varphi) \right\} \\ &= -E\left\{ \int_0^1 \frac{\partial F_t(\theta_0, x)}{\partial \theta} d\varphi(x) \ (\varphi(u_{t-j}) - c_\varphi) \right\} \\ &= v_\varphi R_j, \end{aligned}$$

with R_j defined in (14). The interchange of expectation and integral above follows from Assumption A4, and the integration by parts follows from Theorem 11 of Shiryaev (1996, pp. 206).

Hence, we proved that

$$\sqrt{n-j}(\widehat{\gamma}_{nj} - \gamma_{nj}) = \sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T - \theta_0)' v_\varphi R_j + o_p(1).$$

We then have

$$\begin{aligned}\sqrt{n-j}(\widehat{\rho}_{nj}-\rho_{nj}) &= \frac{\sqrt{n-j}(\widehat{\gamma}_{nj}-\gamma_{nj})}{v_\varphi}+o_p(1) \\ &= R_j'\sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T-\theta_0)+o_p(1).\end{aligned}$$

Therefore,

$$\begin{aligned}\sqrt{n-j}\widehat{\rho}_{nj} &= \sqrt{n-j}\rho_{nj}+R_j'\sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T-\theta_0)+o_p(1) \\ &= \frac{1}{\sqrt{n-j}v_\varphi}\sum_{t=1+j}^{n}(\varphi(u_t)-c_\varphi)(\varphi(u_{t-j})-c_\varphi)+R_j'\sqrt{\lambda}\frac{1}{\sqrt{T}}\sum_{t=-T+1}^{0} l_t+o_p(1).\end{aligned}$$

Notice that $\sqrt{n}(\rho_{n1},\rho_{n2}\dots\rho_{nm})' \longrightarrow^d N(0,I_m)$, and the covariance between the first two terms are 0 as the summand in the first term is for out-of-sample observations and the second term is for in-sample observations. These, together with the above display, imply Theorem A2.

Next we consider the case of information truncation. Define $\widetilde{u}_t = G(Y_t,\Omega_{t-1},\widehat{\theta}_T)$ and $\widetilde{\gamma}_{nj} = 1/(n-j)\sum_{t=1+j}^{n}(\varphi(\widetilde{u}_t)-c_\varphi)(\varphi(\widetilde{u}_{t-j})-c_\varphi)$, and then we have

$$\sqrt{n-j}(\widehat{\gamma}_{nj}-\gamma_{nj}) = \sqrt{n-j}(\widehat{\gamma}_{nj}-\widetilde{\gamma}_{nj})+\sqrt{n-j}(\widetilde{\gamma}_{nj}-\gamma_{nj}).$$

We show that the first term on the right hand side is $o_p(1)$. Since

$$\begin{aligned}\sqrt{n-j}(\widehat{\gamma}_{nj}-\widetilde{\gamma}_{nj}) &= \frac{1}{\sqrt{n-j}}\sum_{t=1+j}^{n}[(\varphi(\widehat{u}_t)-c_\varphi)(\varphi(\widehat{u}_{t-j})-c_\varphi) \\ &\qquad -(\varphi(\widetilde{u}_t)-c_\varphi)(\varphi(\widetilde{u}_{t-j})-c_\varphi)] \\ &= \frac{1}{\sqrt{n-j}}\sum_{t=1+j}^{n}(\varphi(\widehat{u}_t)-\varphi(\widetilde{u}_t))(\varphi(\widehat{u}_{t-j})-c_\varphi) \\ &\quad +\frac{1}{\sqrt{n-j}}\sum_{t=1+j}^{n}(\varphi(\widetilde{u}_t)-c_\varphi)(\varphi(\widehat{u}_{t-j})-\varphi(\widetilde{u}_{t-j})),\end{aligned}$$

and $\varphi(\widetilde{u}_t)=O_p(1)$ as $\varphi\in\Psi$, it follows from Assumption A3 that $\sqrt{n-j}(\widehat{\gamma}_{nj}-\widetilde{\gamma}_{nj})=o_p(1)$. Then notice that the arguments above for $\sqrt{n-j}(\widehat{\gamma}_{nj}-\gamma_{nj})$ without information truncation can be applied directly to $\sqrt{n-j}(\widetilde{\gamma}_{nj}-\gamma_{nj})$, which completes the proof of Theorem A2. ■

Proof of Theorem 2: One can write $\Sigma = Q\Lambda Q'$, where Q is an orthogonal matrix, and Λ is a diagonal matrix with diagonal elements $\{\pi_j\}_{j=1}^m$. By Theorem A2, $Q'\sqrt{n}\widehat{\rho}_n^{(m)} \longrightarrow^d N(0,Q'\Sigma Q)=N(0,\Lambda)$. Theorem 2 then follows from the observation that $C_{ES}(m) = (Q'\sqrt{n}\widehat{\rho}_n^{(m)})'(Q'\sqrt{n}\widehat{\rho}_n^{(m)})$. ■

Proofs of Corollary 2 and Corollary 4: They follow from Theorem 2. ■

TABLES AND FIGURES

Table 1. Empirical rejection rates for backtesting $ES_t(.1)$ and $VaR_t(.05)$ at 5% significance level, $T = 250$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
				$n = 250$, Size and Power (size-corrected)				
H_0	0.169	0.150	0.118	0.103	0.043	0.039	0.053	0.075
A_1	0.080	0.096	0.218	0.164	0.077	0.086	0.207	0.235
A_2	0.549	0.578	0.229	0.197	0.478	0.209	0.212	0.052
A_3	0.035	0.052	0.130	0.096	0.006	0.011	0.155	0.137
A_4	0.063	0.076	0.206	0.142	0.012	0.022	0.199	0.169
A_5	0.072	0.084	0.101	0.076	0.075	0.050	0.114	0.085
A_6	0.351	0.509	0.230	0.227	0.441	0.557	0.236	0.203
				$n = 500$, Size and Power (size-corrected)				
H_0	0.230	0.214	0.125	0.107	0.044	0.043	0.057	0.064
A_1	0.086	0.088	0.356	0.298	0.079	0.098	0.249	0.393
A_2	0.651	0.596	0.283	0.216	0.564	0.229	0.219	0.056
A_3	0.044	0.043	0.190	0.154	0.004	0.004	0.244	0.237
A_4	0.063	0.062	0.284	0.197	0.019	0.020	0.248	0.237
A_5	0.076	0.076	0.114	0.110	0.091	0.062	0.130	0.086
A_6	0.409	0.531	0.294	0.314	0.518	0.667	0.295	0.295

Table 2. Empirical rejection rates for backtesting $ES_t(.05)$ and $VaR_t(.025)$ at 5% significance level, $T = 250$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
				$n = 250$, Size and Power (size-corrected)				
H_0	0.176	0.114	0.140	0.155	0.051	0.042	0.073	0.120
A_1	0.052	0.043	0.131	0.117	0.071	0.082	0.089	0.102
A_2	0.358	0.193	0.180	0.048	0.418	0.138	0.151	0.022
A_3	0.041	0.029	0.094	0.081	0.014	0.008	0.081	0.080
A_4	0.048	0.037	0.123	0.115	0.013	0.012	0.104	0.101
A_5	0.038	0.024	0.055	0.055	0.069	0.064	0.059	0.035
A_6	0.673	0.314	0.492	0.066	0.776	0.786	0.483	0.045
				$n = 500$, Size and Power (size-corrected)				
H_0	0.234	0.222	0.129	0.123	0.048	0.044	0.060	0.103
A_1	0.065	0.081	0.225	0.177	0.074	0.097	0.145	0.205
A_2	0.479	0.468	0.222	0.186	0.519	0.136	0.148	0.029
A_3	0.038	0.050	0.147	0.123	0.010	0.018	0.165	0.148
A_4	0.048	0.053	0.174	0.141	0.006	0.023	0.144	0.123
A_5	0.053	0.057	0.072	0.063	0.084	0.085	0.069	0.041
A_6	0.795	0.655	0.614	0.381	0.824	0.837	0.630	0.275

Table 3. Empirical rejection rates for backtesting $ES_t(.025)$ and $VaR_t(.01)$ at 5% significance level, $T = 250$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
				$n = 250$, Size and Power (size-corrected)				
H_0	0.179	0.034	0.140	0.090	0.071	0.078	0.064	0.075
A_1	0.028	0.086	0.078	0.107	0.040	0.053	0.077	0.064
A_2	0.120	0.001	0.130	0.046	0.180	0.116	0.115	0.009
A_3	0.028	0.030	0.067	0.093	0.010	0.008	0.076	0.065
A_4	0.032	0.102	0.076	0.102	0.006	0.003	0.088	0.095
A_5	0.018	0.031	0.082	0.101	0.024	0.028	0.073	0.079
A_6	0.265	0.000	0.275	0.067	0.418	0.812	0.294	0.004
				$n = 500$, Size and Power (size-corrected)				
H_0	0.239	0.258	0.140	0.133	0.068	0.071	0.059	0.096
A_1	0.048	0.073	0.136	0.093	0.059	0.058	0.097	0.098
A_2	0.312	0.254	0.216	0.003	0.421	0.068	0.140	0.003
A_3	0.036	0.045	0.112	0.083	0.011	0.015	0.115	0.098
A_4	0.030	0.057	0.127	0.095	0.017	0.008	0.117	0.119
A_5	0.019	0.037	0.051	0.047	0.059	0.065	0.051	0.036
A_6	0.640	0.139	0.544	0.007	0.897	0.864	0.562	0.008

Table 4. Empirical rejection rates for backtesting $ES_t(.1)$ and $VaR_t(.05)$ at 5% significance level, $T = 500$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
				$n = 250$, Size and Power (size-corrected)				
H_0	0.101	0.098	0.081	0.093	0.041	0.032	0.057	0.085
A_1	0.072	0.090	0.230	0.173	0.075	0.083	0.203	0.210
A_2	0.686	0.698	0.237	0.178	0.641	0.345	0.203	0.063
A_3	0.028	0.043	0.157	0.124	0.011	0.008	0.152	0.140
A_4	0.059	0.082	0.204	0.172	0.022	0.020	0.184	0.162
A_5	0.078	0.098	0.113	0.108	0.059	0.050	0.082	0.058
A_6	0.488	0.657	0.235	0.265	0.588	0.650	0.249	0.238
				$n = 500$, Size and Power (size-corrected)				
H_0	0.132	0.129	0.085	0.093	0.038	0.032	0.052	0.083
A_1	0.087	0.073	0.401	0.300	0.106	0.119	0.351	0.382
A_2	0.906	0.803	0.275	0.213	0.775	0.481	0.230	0.058
A_3	0.027	0.024	0.264	0.180	0.007	0.007	0.280	0.225
A_4	0.072	0.053	0.328	0.235	0.025	0.031	0.268	0.235
A_5	0.115	0.105	0.143	0.123	0.062	0.058	0.104	0.074
A_6	0.733	0.758	0.287	0.311	0.793	0.865	0.325	0.315

Table 5. Empirical rejection rates for backtesting $ES_t(.05)$ and $VaR_t(.025)$ at 5% significance level, $T = 500$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
$n = 250$, Size and Power (size-corrected)								
H_0	0.112	0.082	0.098	0.109	0.067	0.065	0.071	0.091
A_1	0.042	0.041	0.145	0.113	0.052	0.045	0.104	0.125
A_2	0.490	0.356	0.216	0.067	0.493	0.207	0.175	0.010
A_3	0.027	0.022	0.120	0.097	0.012	0.006	0.099	0.091
A_4	0.031	0.020	0.128	0.114	0.007	0.003	0.109	0.101
A_5	0.038	0.052	0.072	0.070	0.043	0.047	0.052	0.043
A_6	0.837	0.492	0.623	0.083	0.900	0.901	0.628	0.024
$n = 500$, Size and Power (size-corrected)								
H_0	0.151	0.143	0.100	0.098	0.054	0.055	0.063	0.094
A_1	0.058	0.057	0.253	0.165	0.064	0.070	0.201	0.191
A_2	0.721	0.694	0.257	0.167	0.693	0.313	0.182	0.035
A_3	0.024	0.033	0.173	0.127	0.013	0.012	0.140	0.114
A_4	0.036	0.039	0.195	0.146	0.018	0.021	0.185	0.141
A_5	0.081	0.080	0.102	0.088	0.061	0.063	0.065	0.038
A_6	0.956	0.970	0.789	0.467	0.959	0.964	0.804	0.476

Table 6. Empirical rejection rates for backtesting $ES_t(.025)$ and $VaR_t(.01)$ at 5% significance level, $T = 500$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
				$n = 250$, Size and Power (size-corrected)				
H_0	0.146	0.022	0.130	0.079	0.085	0.089	0.087	0.078
A_1	0.036	0.072	0.085	0.097	0.032	0.033	0.070	0.080
A_2	0.275	0.104	0.243	0.026	0.267	0.079	0.175	0.010
A_3	0.027	0.032	0.082	0.090	0.018	0.004	0.078	0.088
A_4	0.022	0.042	0.081	0.089	0.006	0.004	0.073	0.080
A_5	0.009	0.017	0.032	0.027	0.011	0.010	0.036	0.035
A_6	0.355	0.097	0.321	0.063	0.428	0.365	0.335	0.004
				$n = 500$, Size and Power (size-corrected)				
H_0	0.151	0.183	0.119	0.127	0.081	0.080	0.077	0.109
A_1	0.036	0.063	0.101	0.084	0.038	0.032	0.073	0.83
A_2	0.399	0.322	0.186	0.002	0.482	0.145	0.183	0.003
A_3	0.027	0.066	0.120	0.082	0.018	0.005	0.107	0.100
A_4	0.021	0.055	0.116	0.097	0.010	0.005	0.089	0.092
A_5	0.022	0.023	0.050	0.029	0.032	0.041	0.034	0.030
A_6	0.673	0.097	0.610	0.004	0.977	0.960	0.617	0.004

Table 7. Empirical rejection rates for backtesting $ES_t(.1)$ and $VaR_t(.05)$ at 5% significance level, $T = 2500$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
$n = 250$, Size and Power (size-corrected)								
H_0	0.069	0.072	0.051	0.062	0.059	0.067	0.044	0.062
A_1	0.061	0.046	0.309	0.211	0.058	0.048	0.328	0.240
A_2	0.813	0.744	0.297	0.242	0.769	0.549	0.279	0.166
A_3	0.031	0.029	0.218	0.150	0.016	0.009	0.221	0.136
A_4	0.071	0.054	0.281	0.221	0.041	0.032	0.286	0.215
A_5	0.148	0.133	0.248	0.214	0.093	0.071	0.168	0.124
A_6	0.650	0.735	0.255	0.305	0.656	0.735	0.295	0.388
$n = 500$, Size and Power (size-corrected)								
H_0	0.061	0.083	0.053	0.054	0.038	0.046	0.048	0.054
A_1	0.101	0.047	0.494	0.367	0.102	0.065	0.572	0.442
A_2	0.979	0.933	0.345	0.265	0.949	0.837	0.336	0.188
A_3	0.034	0.023	0.331	0.234	0.018	0.011	0.364	0.259
A_4	0.106	0.068	0.408	0.297	0.045	0.044	0.439	0.336
A_5	0.207	0.165	0.251	0.237	0.083	0.064	0.196	0.180
A_6	0.929	0.933	0.320	0.379	0.936	0.933	0.359	0.407

Table 8. Empirical rejection rates for backtesting $ES_t(.05)$ and $VaR_t(.025)$ at 5% significance level, $T = 2500$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
				$n = 250$, Size and Power (size-corrected)				
H_0	0.083	0.047	0.081	0.095	0.078	0.047	0.068	0.093
A_1	0.036	0.047	0.187	0.131	0.036	0.035	0.182	0.136
A_2	0.618	0.505	0.251	0.063	0.598	0.346	0.252	0.021
A_3	0.029	0.041	0.136	0.110	0.023	0.015	0.141	0.100
A_4	0.043	0.053	0.180	0.136	0.028	0.023	0.173	0.121
A_5	0.105	0.100	0.168	0.148	0.072	0.056	0.119	0.087
A_6	0.946	0.515	0.670	0.082	0.953	0.946	0.697	0.042
				$n = 500$, Size and Power (size-corrected)				
H_0	0.069	0.071	0.072	0.063	0.049	0.041	0.061	0.069
A_1	0.056	0.048	0.301	0.208	0.050	0.053	0.276	0.212
A_2	0.931	0.862	0.325	0.286	0.901	0.691	0.291	0.176
A_3	0.031	0.039	0.221	0.149	0.018	0.014	0.214	0.147
A_4	0.065	0.057	0.236	0.193	0.030	0.029	0.238	0.200
A_5	0.195	0.162	0.191	0.189	0.120	0.089	0.158	0.146
A_6	1.000	0.999	0.850	0.604	1.000	0.999	0.875	0.584

Table 9. Empirical rejection rates for backtesting $ES_t(.025)$ and $VaR_t(.01)$ at 5% significance level, $T = 2500$

	U_{ES}	U_{VaR}	$C_{ES}(5)$	$C_{VaR}(5)$	MU_{ES}	MU_{VaR}	$MC_{ES}(5)$	$MC_{VaR}(5)$
$n = 250$, Size and Power (size-corrected)								
H_0	0.103	0.093	0.101	0.082	0.098	0.089	0.095	0.071
A_1	0.029	0.077	0.078	0.097	0.027	0.027	0.067	0.057
A_2	0.400	0.527	0.344	0.050	0.290	0.027	0.202	0.000
A_3	0.023	0.088	0.088	0.088	0.018	0.000	0.079	0.084
A_4	0.034	0.082	0.100	0.082	0.017	0.005	0.102	0.100
A_5	0.055	0.055	0.100	0.055	0.013	0.003	0.052	0.024
A_6	0.401	0.956	0.349	0.062	0.420	0.955	0.353	0.001
$n = 500$, Size and Power (size-corrected)								
H_0	0.075	0.124	0.090	0.095	0.062	0.103	0.082	0.092
A_1	0.030	0.040	0.123	0.083	0.028	0.029	0.107	0.090
A_2	0.732	0.606	0.360	0.051	0.720	0.292	0.249	0.001
A_3	0.033	0.058	0.133	0.092	0.022	0.005	0.124	0.114
A_4	0.036	0.054	0.156	0.120	0.021	0.009	0.141	0.124
A_5	0.117	0.058	0.133	0.101	0.071	0.040	0.100	0.082
A_6	1.000	0.967	0.648	0.002	1.000	0.995	0.657	0.002

Table 10. Descriptive statistics for the log-returns (%) of three indexes

	In-sample (1997-2007)			**Out-of-sample (2007-2009)**		
	S&P500	**DAX**	**HangSeng**	**S&P500**	**DAX**	**HangSeng**
No. Obs.	2639	2658	2596	504	509	503
Mean	0.027	0.039	0.019	-0.098	-0.100	-0.034
Median	0.059	0.106	0.042	0.036	0.016	0.052
St.Dev.	1.131	1.577	1.677	2.218	2.045	2.762
Skewness	-0.089	-0.152	0.146	-0.065	0.351	0.167
Excess Kurtosis	3.165	2.348	11.355	4.035	5.273	3.319
Maximum	5.574	7.553	17.250	10.960	10.800	13.410
10 percentile	-1.324	-2.132	-1.712	-2.478	-2.132	-3.238
5 percentile	-1.815	-2.567	-2.552	-3.506	-3.371	-4.381
1 percentile	-2.881	-4.547	-4.306	-6.310	-6.061	-7.305
Minimum	-7.113	-6.652	-14.73	-9.470	-7.433	-13.580

Table 11. CML Estimates

	S&P500	**DAX**	**HangSeng**
a_0	-0.027	0.004	0.034
ω_0	0.007	0.016	0.010
α_0	0.059	0.088	0.058
β_0	0.937	0.910	0.948
v	9	10	4
$F_v^{-1}(0.05)$	-1.617	-1.621	-1.507
$F_v^{-1}(0.01)$	-2.488	-2.472	-2.649
$m(0.1)$	-1.781	-1.779	-1.767
$m(0.025)$	-2.544	-2.521	-2.824

Table 12. Descriptive analysis of violations: Pre-crisis vs Crisis

	Pre-Crisis (2005-2007)			**Crisis (2007-2009)**		
	S&P500	**DAX**	**HangSeng**	**S&P500**	**DAX**	**HangSeng**
$V(0.05)$	20	20	24	41	35	29
$CV(0.1)$	20.309	22.434	24.714	40.026	34.862	30.612
$n \times 0.05$	25.2	25.45	25.15	25.2	25.45	25.15
$V(0.01)$	5	8	2	11	5	5
$CV(0.025)$	6.110	6.360	4.063	13.702	9.101	6.145
$n \times 0.01$	5.04	5.09	5.03	5.04	5.09	5.03

Table 13. p-values for backtesting ES and VaR

S&P500	$ES(0.025)$	$VaR(0.01)$	$ES(0.1)$	$VaR(0.05)$
U	0.011	0.070	0.004	0.010
$C(5)$	0.007	0.270	0.009	0.052
MU	0.019	0.073	0.006	0.013
$MC(5)$	0.017	0.271	0.010	0.053
DAX	$ES(0.025)$	$VaR(0.01)$	$ES(0.1)$	$VaR(0.05)$
U	0.224	0.968	0.045	0.095
$C(5)$	0.002	0.998	0.091	0.768
MU	0.253	0.968	0.052	0.102
$MC(5)$	0.015	0.998	0.095	0.769
HangSeng	$ES(0.025)$	$VaR(0.01)$	$ES(0.1)$	$VaR(0.05)$
U	0.939	0.989	0.194	0.462
$C(5)$	0.002	0.998	0.002	0.002
MU	0.945	0.990	0.310	0.509
$MC(5)$	0.003	0.998	0.004	0.002

Note: U denotes the basic unconditional backtest; MU the robust unconditional backtest; $C(5)$ the basic conditional backtest with $m = 5$; $MC(5)$ the robust conditional backtest.

REFERENCES

[1] Acerbi, C. and Tasche, D. (2002), "On the coherence of expected shortfall," *Journal of Banking & Finance,* 26, 1487-1503.

[2] Acerbi, C. and Szekely, B. (2014), "Backtesting Expected Shortfall," to appear in *Risk Magazine*.

[3] Artzner, P., Delbaen, F., Eber, J-M., Heath, D. (1997), "Thinking coherently", *Risk*, 10, 68–71.

[4] Artzner, P., Delbaen, F., Eber, J-M., Heath, D. (1999), "Coherent measures of risk", *Mathematical Finance*, 9, 203–228.

[5] Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models", *Review of Economics and Statistics*, 85(3): 531-549.

[6] Basak, S. and Shapiro, A. (2001), "Value-at-risk based risk management: optimal policies and asset prices", *Review of Financial Studies,* 14, 371-405.

[7] Basel Committee on Banking Supervision (2012). Consultative Document: Fundamental Review of the Trading Book, Basel, Switzerland. Available at: http://www.bis.org/publ/bcbs219.pdf

[8] Berkowitz, J. (2001), "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, 465-474.

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

[10] Bose, A. (1998), "Bahadur Representation of M Estimates", *Annals of Statistics*, 26(2), 771-777.

[11] Box, G. and Pierce, D. (1970), "Distribution of Residual Autocorrelations in Autorregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association*, 65, 1509-1527.

[12] Christoffersen, P.F. (1998), "Evaluating interval forecasts", *International Economic Review,* 39, 841-862.

[13] Christoffersen, P.F. (2009), "Value-at-Risk Models", *Handbook of Financial Time Series*, 753-766.

[14] Christoffersen, P.F. and Goncalves, P. (2005), "Estimation Risk in Financial Risk Management," *Journal of Risk*, 7(3):1-28.

[15] Colletaz, G., Hurlin, C. and Perignon, C. (2013), "The Risk Map: A New Tool for Validating Risk Models", *Journal of Banking and Finance* 37(10), 3843-3854.

[16] Du, Z. (2015), "Nonparametric Bootstrap Tests for the Independence of Generalized Errors", working paper.

[17] Emmer, S., Kratz, M. and Tasche, D. (2014), "What is the Best Risk Measure in Practice? A Comparison of Standard Measures", ESSEC Business School: Paris, France, 2014; preprint.

[18] Escanciano, J.C. (2007), "Weak Convergence of Non-Stationary Multivariate Marked Processes with Applications to Martingale Testing", *Journal of Multivariate Analysis*, 98, 1321-1336.

[19] Escanciano, J. C. and Lobato, I. N. (2009a). Testing the martingale hypothesis, in T. C. Mills and K. Patterson (eds), Palgrave Handbook of Econometrics, Vol. 2, Palgrave Macmillan, Basingstoke, pp. 972—1003.

[20] Escanciano, J.C. and Lobato, I.N. (2009b), "An Automatic Portmanteau Test for Serial Correlation", *Journal of Econometrics*, 151, 2, 140-149.

[21] Escanciano, J.C. and Olmo, J. (2010), "Backtesting Parametric Value-at-Risk with Estimation Risk", *Journal of Business and Economic Statistics*, 28, 36-51.

[22] Escanciano, J.C. and Pei, P. (2012), "Pitfalls in Backtesting Historical Simulation Models", *Journal of Banking and Finance*, 36, 2233-2244.

[23] Escanciano, J.C. and Velasco, C. (2010), "Specification Tests of Parametric Dynamic Conditional Quantiles", *Journal of Econometrics*, 159, 209-221.

[24] Francq, C. and Zakoian, J.M. (2015), "Risk-parameter estimation in volatility models", *Journal of Econometrics* 184, 158–173.

[25] Gourieroux, C. and Zakoian, J.M. (2013), "Estimation-Adjusted VaR", *Econometric Theory*, 29(04), 735-770.

[26] Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation", *International Economic Review*, 35, 705–730.

[27] Hong, Y. & Lee, T.H. (2003), "Diagnostic Checking for Adequacy of Linear and Nonlinear Time Series Models", *Econometric Theory*, 19, 2003, 1065–1121.

[28] Hong, Y. and Li, H. (2005), "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates", *The Review of Financial Studies*, 18, 1, 37-84.

[29] Hurlin C. and Tokpavi, S. (2006), "Backtesting Value-at-Risk Accuracy: A Simple New Test", *Journal of Risk*, 9, 19-37.

[30] Jorion, P. (2006). Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill, New York.

[31] Kerkhof, J . and Melenberg, B. (2004), “Backtesting for risk-based regulatory capital,” *Journal of Banking and Finance,* 28, 1845-1865.

[32] Kourouma, L., Dupre, D., Sanfilippo, G and Taramasco, O. (2011), “Extreme value at risk and expected shortfall during financial crisis,” http://halshs.archives-ouvertes.fr/docs/00/65/84/95/PDF/CR_2011-03_E2.pdf.

[33] Kupiec, P. (1995), "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives,* 3, 73-84.

[34] McNeil, A. J. and Frey, R. (2000), "Estimation of Tail-Related RiskMeasures for Heteroscedastic Financial Time Series: An Extreme Value approach," *Journal of Empirical Finance*, 7, 271-300.

[35] O'Brien, J. and Szerszen, P. J. (2014), “An Evaluation of Bank VaR Measures for Market Risk During and Before the Financial Crisis,” http://www.federalreserve.gov/pubs/feds/2014/201421/201421abs.html.

[36] Perignon, C. and Smith, D. R. (2008), "A New Approach to Comparing VaR Estimation Methods," *Journal of Derivatives*, 16, 54-66.

[37] Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470-472.

[38] Shiryaev, A.N. (1996), *Probability*. Springer, New York.

[39] van der Vaart, A. W. and Wellner, J.A. (1996), *Weak Convergence and Empirical Processes*. Springer, New York.

[40] Wong, W.K. (2008), “Backtesting trading risk of commercial banks using expected shortfall,” *Journal of Banking and Finance*, 32, 1404-1415.

[41] Wong, W.K. (2010), “Backtesting value-at-risk based on tail losses,” *Journal of Empirical Finance*, 17, 526-538.

[42] Wu, W.B. (2007), "M-Estimation of Linear Models with Dependent Errors", *Annals of Statistics*, Vol. 35, No. 2, 495–521.

[43] Yamai, Y. and Yoshiba, T. (2002), “Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization”, *Monetary and Economic Studies*, 20, 87-122.

[44] Yamai, Y. and Yoshiba, T. (2005), “Value-at-risk versus expected shortfall: A practical perspective,” *Journal of Banking and Finance*, 29, 99-115.