CAEPR Working Paper
#2015-001

Backtesting Expected Shortfall: Accounting for Tail Risk

ZAICHAO DU^t

Southwestern University of Finance and Economics, China

JUAN CARLOS ESCANCIANO:

Indiana University, United States of America

January 11, 2015

This paper can be downloaded without charge from the Social Science Research Network electronic library at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2548544

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

©2015 by Zaichao Du and Juan Carlos Escanciano. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Backtesting Expected Shortfall: Accounting for Tail Risk*

ZAICHAO DU[†]

Southwestern University of Finance and Economics, China

JUAN CARLOS ESCANCIANO[‡]

Indiana University, United States of America

January 11, 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 ones 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 a conditional backtest for ES based on cumulative violations, which is the natural analogue of the commonly used conditional backtest for VaR . We establish the asymptotic properties of the test, and investigate its 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 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.

*We would like to thank the computational resources provided by Yan Dong.

[†]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.

1. INTRODUCTION

The quantification of market risk for derivative pricing, for portfolio choice and for managing risk 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 with 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 Yoshida (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. Escanciano and Olmo (2010) and 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

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

for the *mds* property (see Escanciano and Lobato (2009a) for a survey of these methods). In particular, we suggest a Box-Pierce (BP) test (cf. Box and Pierce (1970)). Our BP 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). Indeed, this unconditional backtest for *ES* is asymptotically equivalent to the backtest proposed by Kerkhof and Melenberg (2004). See also Berkowitz (2001), Wong (2008, 2010) and Acerbi and Szekely (2014) for other unconditional backtests for *ES*. In this article, our main focus is on conditional backtests, which, to the best of our knowledge, are not yet available in the literature.²

This article shows 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 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 conditional backtest, and derive its 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 during the 2008-financial crisis: the S&P500, the Deutsche German Stock Index (DAX) and the Hang Seng Index. 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

²Conditional 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.

³Our 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 “elicibility”. We contribute to this discussion by pointing out that, like any other functional of a conditional distribution, ES can be evaluated by comparing nonparametric with parametric fits.

conclude. An Appendix contains mathematical proofs of our results and extensions of the basic setting.

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, $\Omega_{t-1} = \{X_{t-1}, X_{t-2}, \dots; Y_{t-1}, Y_{t-2}, \dots\}$. Let $G_t(\cdot, \Omega_{t-1})$ denote the conditional distribution of Y_t given Ω_{t-1} , i.e. $G_t(\cdot, \Omega_{t-1}) = \Pr(Y_t \leq \cdot | \Omega_{t-1})$. For simplicity, we drop “almost surely” in all equalities involving random variables. Assume $G_t(\cdot, \Omega_{t-1})$ 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

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

That is, the $VaR_t(\alpha)$ is negative of the α -th percentile of the distribution G_t ,

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

Define the α violation 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 revenue at time t is less 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.

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

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 expected loss given that the loss is larger than $VaR_t(\alpha)$, that is,

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

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

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

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 VaR s 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 , $H_t(\alpha)$ has mean $1/\alpha \int_0^\alpha u du = \alpha/2$. Moreover, 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 $u_t = G_t(Y_t, \Omega_{t-1})$. Using that $h_t(u) = 1(Y_t \leq -VaR_t(u)) = 1(u_t \leq u)$, we write

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

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. Working with violations avoids approximations, as the integral can be computed

exactly (cf. 4). 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 $\alpha - u_t = G_t(G_t^{-1}(\alpha, \Omega_{t-1}), \Omega_{t-1}) - G_t(Y_t, \Omega_{t-1})$.⁴

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_t is unknown. In practice, researchers and risk managers specify a parametric conditional distribution $G_t(\cdot, \Omega_{t-1}, \theta_0)$, where θ_0 is some unknown parameter in $\Theta \subset \mathbb{R}^p$, and proceed to estimate θ_0 before using $G_t(\cdot, \Omega_{t-1}, \theta_0)$ for measuring and evaluating risk models. Popular choices for distributions $G_t(\cdot, \Omega_{t-1}, \theta_0)$ are those derived from location-scale models with Student's t distributions, but other choices can be certainly entertained in our setting. With the parametric model at hand, we can define the “generalized errors”

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

and the associated cumulative violations

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

Very much like for *Var*s, 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 backtest for *ES*. The unconditional backtest is simply a t -test for the hypothesis $E[H_t(\alpha)] = \alpha/2$, and it is the analogue for the *ES* of the unconditional *Var* backtest proposed in Kupiec (1995). The conditional backtest is a Portmanteau BP 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*.

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

⁴In fact, $d(y, x) = |G_t(y, \Omega_{t-1}) - G_t(x, \Omega_{t-1})|$ is a distance function.

(rolling and recursive); see, e.g., Escanciano and Olmo (2010) and references therein for details. That is, the in-sample period $\{Y_{-T+1}, \hat{\Omega}_{-T}, \dots, Y_0, \hat{\Omega}_{-1}\}$ of size T is used to estimate θ_0 , say by $\hat{\theta}_T$, where $\hat{\theta}_T$ 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 $\hat{\theta}_T$ we construct residuals

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

and estimated cumulative violations

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

Then, an out-of-sample period $\{Y_1, \hat{\Omega}_0, \dots, Y_n, \hat{\Omega}_{n-1}\}$ of size n is used to evaluate (back-test) the ES model. A standard t -test for the hypothesis $E[H_t(\alpha)] = \alpha/2$ can be entertained. This is the unconditional backtest for ES . Note that simple calculations show that $E[H_t^2(\alpha)] = \alpha/3$, and hence, $v_{ES}(\alpha) = Var(H_t(\alpha)) = \alpha(1/3 - \alpha/4)$. Then, the proposed test rejects $E[H_t(\alpha)] = \alpha/2$ at 5% nominal level if

$$t_{ES} = \frac{\sqrt{n}(\bar{H}(\alpha) - \alpha/2)}{\sqrt{v_{ES}(\alpha)}}$$

is larger in absolute value than 1.96 (the corresponding critical value of a standard normal), where $\bar{H}(\alpha)$ denotes the sample mean of $\{\hat{H}_t(\alpha)\}_{t=1}^n$, i.e.

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

Kerkhof and Melenberg (2004) proposed a backtest for ES that, under some conditions, is asymptotically equivalent to a one-sided version of the t -test.

To introduce our conditional backtest, consider 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. However, in our present context $\{H_t(\alpha)\}_{t=1}^n$ is unobservable, as θ_0 is unknown and Ω_{t-1} is not completely observed. Then we substitute $\hat{H}_t(\alpha)$ for $H_t(\alpha)$ in γ_{nj} and obtain

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

The asymptotic distribution of $\hat{\rho}_{nj}$ is related to that of ρ_{nj} depending on the asymptotic relative size of the in-sample (estimation) size T and the out-of-sample (evaluation) size n . Assume both $T \rightarrow \infty$ and $n \rightarrow \infty$, such that $n/T \rightarrow \lambda < \infty$. There are two scenarios: $\lambda = 0$ (Case 1) and $\lambda > 0$ (Case 2). Case 1 corresponds to the situation where the estimation period is much larger than the evaluation period. In this case, and under the standard assumption that $\hat{\theta}_T$ is $\sqrt{T}$ -consistent, the asymptotic distribution of $\sqrt{n}\hat{\rho}_{nj}$ is the same as that of $\sqrt{n}\rho_{nj}$, which takes a simple form, as the following Theorem shows. Define the m -dimension vector of autocorrelations as $\hat{\rho}_n^{(m)} = (\hat{\rho}_{n1}, \hat{\rho}_{n2}, \dots, \hat{\rho}_{nm})'$ (here A' denotes the transpose of A). The symbol $\longrightarrow^d$ denotes convergence in distribution.

Theorem 1 *Under Assumptions A0-A4 in the Appendix and $\lambda = 0$,*

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

where I_m denotes the identity matrix of dimension m .

Corollary 1 *Under the assumptions of Theorem 1,*

$$BP_{ES}(m) := n \sum_{j=1}^m \hat{\rho}_{nj}^2 \longrightarrow^d \chi_m^2,$$

with χ_m^2 a chi-square distribution with m degrees of freedom.

Let $\chi_{m,1-\tau}^2$ denote the $(1 - \tau)$ -th quantile of the χ_m^2 random variable. Then, the proposed conditional backtest rejects the validity of ES forecasts at the τ -th nominal level if

$$BP_{ES}(m) > \chi_{m,1-\tau}^2.$$

The standard normal approximation for $\sqrt{n}\hat{\rho}_n^{(m)}$ in Theorem 1 can be inaccurate if T is not very large relative to n , i.e. if $\lambda > 0$. In this case, the limiting distribution of $\sqrt{n}\hat{\rho}_n^{(m)}$ is different from that of $\sqrt{n}\rho_n^{(m)}$ (and hence, different from the standard normal). Obtaining the

limiting distribution for $\sqrt{n}\widehat{\rho}_n^{(m)}$ is technically challenging because the cumulative violations $H_t(\alpha, \theta_0)$ are a non-smooth transformation of θ_0 , and the commonly used Taylor expansion argument does not apply. When $\lambda > 0$, the limiting distribution of the conditional backtest will depend on the estimator $\widehat{\theta}_T$ used, and explicitly on its linear expansion, in a complicated manner. The reader is referred to the Appendix for the asymptotic theory in this case. These technical results provided in the Appendix are of independent interest, as they can be used to obtain the asymptotic distribution theory for correlations of non-smooth transformations of generalized residuals.

4. MONTE CARLO SIMULATIONS

To assess the finite sample performance of our proposed tests, including their power performance, 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) 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*, given by

$$t_{vaR} = \frac{\sqrt{n} (\bar{h}(\alpha) - \alpha)}{\sqrt{v_{VaR}(\alpha)}},$$

where $\bar{h}(\alpha)$ is the sample average of $\{\widehat{h}_t(\alpha) = 1(\widehat{u}_t \leq \alpha)\}_{t=1}^n$, and $v_{VaR}(\alpha) = \text{var}(h_t(\alpha)) = \alpha(1 - \alpha)$. The Box-Pierce-type test for *VaR* is given by

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

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 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

$$\begin{aligned} VaR_t(\alpha) &= -a_0 Y_{t-1} - \sigma_t F_v^{-1}(\alpha), \quad \sigma_t^2 = \omega_0 + \alpha_0 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \quad \text{and} \\ ES_t(\alpha) &= -a_0 Y_{t-1} - \sigma_t m(\alpha), \quad m(\alpha) = E[\varepsilon_t | \varepsilon_t \leq F_v^{-1}(\alpha)], \end{aligned} \quad (5)$$

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 $\hat{u}_t = F_{\hat{v}}(\hat{\varepsilon}_t)$, where $F_{\hat{v}}(\cdot)$ denotes the cumulative distribution function (CDF) of Student's t with $\hat{v}$ degrees of freedom; $\hat{\varepsilon}_t = (Y_t - \hat{a}Y_{t-1})/\hat{\sigma}_t$; and $\hat{\sigma}_t$ is as σ_t with estimated parameters replacing the true parameters. We calculate the test statistics $t_{ES}, t_{VaR}, BP_{ES}(m)$ and $BP_{VaR}(m)$, for $m = 1, 3$ and 5 . We repeat the experiment 1000 times. The reported in-sample size is $T = 2500$, although we also experimented with $T = 1000$ and $T = 5000$ (unreported). We consider out-of-sample sizes of $n = 250$ and $n = 500$, to evaluate power performance and the consistency of tests.

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

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

$$\begin{aligned} Y_t &= 0.05Y_{t-1} + v_t, \quad v_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim t_5 \\ \sigma_t^2 &= 0.05 + 0.1v_{t-1}^2 + 0.85\sigma_{t-1}^2, \end{aligned} \tag{6}$$

A_1 : AR(2)-GARCH(1,1) model: $Y_t = 0.05Y_{t-1} + 0.3Y_{t-2} + v_t$.

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

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: $Y_t = 0.05Y_{t-1} + v_t, v_t = h_t \varepsilon_t, \ln h_t^2 = 0.01 + 0.9 \ln h_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - \sqrt{2/\pi}) - 0.8\varepsilon_{t-1}$.

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

A_6 : AR(1)-GARCH(1,1) model with mixed normal innovations: Y_t is as in (6), with $\varepsilon_t \sim [0.5 \cdot N(-3, 1) + 0.5N(-3, 1)]/\sqrt{10}$.

In these models $\{\varepsilon_t\}$ is generally *iid* t_5 , unless otherwise specified. In A_1 , v_t is defined as in (6). These models are studied in Escanciano and Velasco (2010). In A_1 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 1-3 give the empirical sizes and powers of the tests at 5% nominal level. Generally, we find the size performance to be satisfactory. There are some size distortions for backtesting $ES_t(.025)$ and $VaR_t(.01)$ with $n = 250$, but the distortions become much less severe when $n = 500$. Unreported simulations for other values of T , such as $T = 1000$ and $T = 5000$, show improvements in the empirical size for small coverage levels as T gets larger, but these simulations results also suggest that it is more important to have larger out-of-sample sizes (n) than 250 for accurate size performance with $ES_t(.025)$. For larger coverage levels, such as 0.05 and 0.1, the empirical sizes with $n = 250$ are satisfactory even for smaller values of T such as $T = 1000$.

The results for power suggest complementarity between unconditional and conditional backtests for ES . The unconditional backtest detects well alternatives A_2 , A_5 and A_6 , but has low power for A_1 , A_3 and A_4 . In contrast, the conditional backtest with 3 and 5 number of lags has moderate power against all alternatives $A_1 - A_5$ and high power for A_6 . For A_1 , the conditional backtest requires a lag larger than one to be able to detect the second order autoregressive component, as expected. Almost in all cases, the ES backtests have higher power than the VaR counterparts. Finally, we observe that the power for $BP_{ES}(3)$ and $BP_{ES}(5)$ increases with the 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's distribution (see Hansen (1994)) with time-varying higher order moments.⁵

TABLES 1-3 ABOUT HERE

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

⁵These simulations are available from the authors upon request.

financial crisis, as measured by traditional *VaR* backtests with regulatory coverage levels, but 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.

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 4 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 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 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 4 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

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

respectively, as defined in (5).

We estimate the parameters by CMLE using the in-sample data, and perform backtesting with the out-of-sample data. Table 5 reports the CMLE estimates, including estimates for the terms $F_v^{-1}(\alpha)$ and $m(\alpha)$ for the levels of α considered. 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 4.

TABLE 5 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. 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 6 reports the number of violations $\left(V(\alpha) = \sum_{t=1}^n \hat{h}_t(\alpha)\right)$, cumulative violations $\left(CV(\alpha) = \sum_{t=1}^n \hat{H}_t(\alpha)\right)$ 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 6 shows significant discrepancies between violations and cumulative violations at the coverage level suggested by the Basel committee ($\alpha = 0.01$ for VaR).

TABLE 6 ABOUT HERE

Figure 5 plots the cumulative violations $\{\hat{H}_t(0.1)\}$ of the three indexes in the out-of-sample crisis period. We observe large values of $\hat{H}_t(0.1)$, 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 7 reports the p-values of the unconditional tests t_{ES} , t_{VaR} and the conditional tests $BP_{ES}(5)$, $BP_{VaR}(5)$ for the three indexes, respectively. Our conditional test based on *ES*, $BP_{ES}(5)$, generally rejects the null model, while $BP_{VaR}(5)$ does not. Figures 6 and 7 plot the sample autocorrelations of $\hat{H}_t(0.025)$ and $\hat{h}_t(0.01)$, 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 $\{\hat{h}_t(0.01)\}$ for all three series. The autocorrelations of $\hat{h}_t(0.01)$ 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 more extreme losses in the stock market in the US than in Germany and Hong Kong. This also may explain why the unconditional test t_{VaR} in Table 7 has a small p-value for S&P500, and a big p-value for the other two indexes.

The cumulative violations $\{\hat{H}_t(0.025)\}$, 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 $\{\hat{H}_t(0.025)\}$ 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 7, our test based on $\{\hat{H}_t(0.025)\}$ is able to better detect the problems of one of the most commonly used risk models during the 2008 financial crisis.

TABLE 7 ABOUT HERE
FIGURES 6-7 ABOUT HERE

In summary, based on *VaR* backtesting, one cannot find an 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, it has been only recently that *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.

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 to

several choices of m , such as $m = 1, 3$ and 5 , with 5 featuring the best overall performance in our Monte Carlo simulations. 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. For completeness, we discuss the theoretical properties of the data-driven backtest in the Appendix C and refer to future research the evaluation of its finite sample properties.

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 Θ_0 be an arbitrary neighborhood of $\theta_0 \in \Theta$. Finally, 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 $\varphi : [0, 1] \rightarrow \mathbb{R}$, which are right continuous with left limits (*cadlag*), of bounded variation or non-decreasing. The case of cumulative violations corresponds to the special case

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

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: $u_t(\theta) = G_t(Y_t, \Omega_{t-1}, \theta)$, $u_t \equiv u_t(\theta_0)$, $c_\varphi = E[\varphi(u_t)]$ and $v_\varphi = \text{var}(\varphi(u_t))$. Consider the following assumptions.

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

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

Assumption A2: The estimator $\hat{\theta}_T$ is $\sqrt{T}$ -consistent for θ_0 , where θ_0 is in the interior of Θ .

Assumption A3: The effect of information truncation satisfies

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

Assumption A4: $F_t(\theta, x) \equiv \Pr[u_t(\theta) \leq x | \Omega_{t-1}]$ is continuously differentiable in θ and $x \in [0, 1]$ a.s. Moreover,

$$R_j = \frac{-1}{v_\varphi} E \left\{ (\varphi(u_{t-j}) - c_\varphi) \int_0^1 \frac{\partial F_t(\theta_0, x)}{\partial \theta} d\varphi(x) \right\} < \infty, \quad (7)$$

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

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

$$P(Y_t \leq y | \Omega_{t-1}) = G_t(y, \Omega_{t-1}, \theta_0) \text{ for all } y \leq G_t^{-1}(\alpha, \Omega_{t-1}, \theta_0),$$

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 easily holds for many time series models, including stationary and invertible ARMA processes, GARCH processes etc., see e.g. the discussions in Bai (2003). This assumption is not needed when the process is Markovian. Assumption A4 is required for the asymptotic equicontinuity⁶ of certain empirical processes and the uniform law of large numbers.

When $\lambda > 0$ we need to strengthen A2 with A2'.

Assumption A2': In addition to A2, $\hat{\theta}_T$ satisfies the following asymptotic (Bahadur) expansion,

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

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

Appendix B: Case when $\lambda > 0$

It is known that for CMLE Assumption A2' holds with $l_t = I_{\theta_0}^{-1} s_t$, with $I_{\theta_0} = E[s_t s_t']$, $s_t = \nabla_{\theta_0} \ln g_t(\theta)$, and where $g_t(\theta) \equiv g_t(\cdot, \Omega_{t-1}, \theta)$ is the density function of $G_t(\cdot, \Omega_{t-1}, \theta)$.

Theorem 2 *Under Assumptions A0-A4 and $\lambda > 0$,*

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

⁶For definition of asymptotic (uniform) equicontinuity see Chapter 1.5 in van der Vaart and Wellner (1996).

with the ij -th element of Σ given by

$$\Sigma_{ij} = \delta_{ij} + \lambda R_i' E[l_t l_t'] R_j, \quad (8)$$

where

$$R_j = \frac{-1}{v_{ES}(\alpha)} E \left\{ (H_{t-j}(\alpha) - \alpha/2) \int_0^\alpha \frac{\partial F_t(\theta_0, x)}{\partial \theta} dx \right\}, \quad (9)$$

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

Remark A1: An implication of Theorem 2 is that the Box-Pierce test statistic $BP_{ES}(m)$ no longer has a chi-square limit distribution as the tests based on ρ_{nj} .

Remark A2: For the general location-scale model

$$Y_t = \mu_t + \sigma_t \varepsilon_t, \quad (10)$$

where $\mu_t = \mu(\Omega_{t-1}, \theta_0) = E[Y_t \mid \Omega_{t-1}]$, $\sigma_t^2 = \sigma^2(\Omega_{t-1}, \theta_0) = \text{Var}[Y_t \mid \Omega_{t-1}]$ almost surely (a.s.); ε_t follows a distribution with CDF $G(\cdot)$ and density function $g(\cdot)$, we have equation

$$\frac{\partial F_t(\theta_0, x)}{\partial \theta} = g(G^{-1}(x)) \frac{\dot{\mu}_t + G^{-1}(x) \dot{\sigma}_t}{\sigma_t}, \quad (11)$$

and hence

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

whose feasible counterpart is given by

$$\hat{R}_j = \frac{-1}{v_{ES}(\alpha)} \frac{1}{n-j} \sum_{t=j+1}^n (\hat{H}_{t-j}(\alpha) - \alpha/2) g(\hat{\varepsilon}_t) 1(\hat{\varepsilon}_t \leq G^{-1}(\alpha)) \frac{\dot{\mu}_t + \hat{\varepsilon}_t \dot{\sigma}_t}{\sigma_t}.$$

Appendix C: Data-Driven Tests

Regardless of the value of λ , the Portmanteau test $BP_{ES}(m)$ has a practically important limitation, which is that inference based on it can be sensitive to the selected number of autocorrelations m . To overcome this limitation, we propose here a fully automatic Portmanteau test where m is not fixed but selected automatically from the data. We follow the

suggestion in Inglot and Ledwina (2006a, b) and use a combination of Akaike (1974) and Schwarz (1978) criteria to select m , and consider a data-driven test statistic given by

$$BP_{ES}^* = BP_{ES}(m^*), \quad (12)$$

where

$$m^* = \min\{m : 1 \leq m \leq p; L_n(m) \geq L_n(h), h = 1, 2, \dots, p\},$$

$$L_n(m) = BP_{ES}(m) - \pi_n(m, q),$$

p is an arbitrarily large but fixed upper bound,

$$\pi_n(m, q) = \begin{cases} m \log n, & \text{if } \max_{1 \leq j \leq p} \sqrt{n} |\hat{\rho}_{nj}| \leq \sqrt{q \log n}, \\ 2m, & \text{if } \max_{1 \leq j \leq p} \sqrt{n} |\hat{\rho}_{nj}| > \sqrt{q \log n}. \end{cases}$$

and q is some fixed positive number.

As explained in Inglot and Ledwina (2006a, b), the motivation of this selection rule for m is to combine the advantages of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). On the one hand, tests constructed using the BIC criterion are able to properly control the type I error and are more powerful when the serial correlation is present in the first order autocorrelations. On the other hand, tests based on the AIC cannot properly control the type I error, but they are more powerful when the serial correlation is present in high order autocorrelations. This selection rule for m allows the data to choose the preferable criterion according to the data characteristics. There is a theoretical justification and some extensive simulations in the literature suggesting the choice of $q = 2.4$, see Inglot and Ledwina (2006a, b) and Escanciano and Lobato (2009b), among others.

Our next theorem proves the asymptotic null distribution of the data-driven Portmanteau test under the two cases for λ .

Theorem 3 *Under Assumptions A0-A4 in the Appendix,*

- (i) *if $\lambda = 0$, then $BP_{ES}^* \xrightarrow{d} \chi_1^2$.*
- (ii) *if $\lambda > 0$, then $BP_{ES}^* \xrightarrow{d} \Sigma_{11} \chi_1^2$.*

To get an asymptotically distribution-free test in any case, we define

$$Q_{ES}^* = \frac{BP_{ES}^*}{\hat{\Sigma}_{11}}$$

where

$$\widehat{\Sigma}_{ij} = \delta_{ij} + \lambda \widehat{R}_i' W_n \widehat{R}_j, \quad (13)$$

$W_n = n^{-1} \sum_{t=1}^n \widehat{l}_t \widehat{l}_t'$, $\widehat{l}_t$ is a consistent estimator of l_t , $\widehat{R}_j$ is a consistent estimator of R_j . In all our empirical results the natural CMLE is used. From Theorem 3 and some standard arguments, it follows that $Q_{ES}^* \xrightarrow{d} \chi_1^2$ under both scenarios $\lambda = 0$ and $\lambda > 0$.

Appendix D: Proofs

In the Appendix we prove more general versions of Theorems 1 and 2, for which we need the following lemmas. Define the processes

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

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

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

where

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

with $F_t(\cdot, \cdot)$ defined in Assumption A4.

Lemma A1 is a special case of Theorem 1 in Du (2010), 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 $0 \leq y \leq 1$, $R(x, \cdot) \in \Psi$ for $0 \leq x \leq 1$ 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*

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

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

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

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 $j \geq 0$ by

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

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 $\hat{u}_t$ for u_t in γ_{nj} and obtain

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

THEOREM A1: *Under Assumptions A0-A4,*

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

with the ij -th element of Σ given by

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

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

PROOF OF THEOREM A1: We first consider the case of no information truncation. This

occurs if G_t 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

$$\sqrt{n-j}\widehat{\gamma}_{nj} = \sqrt{n-j}\gamma_{nj} + \sqrt{\lambda}\sqrt{T}(\widehat{\theta}_T - \theta_0)' \int_0^1 \int_0^1 \varphi(x)\varphi(y)E_j(dx, dy) + o_p(1), \quad (14)$$

where

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

with R_j defined in (7). 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,

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

Notice that $\sqrt{n}(\rho_{n1}, \rho_{n2}, \dots, \rho_{nm})' \xrightarrow{d} N(0, I_m)$, 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 A1.

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

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

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

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

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

■

PROOF OF THEOREM 1 AND THEOREM 2: The proofs follow from the proof of Theorem A1. ■

PROOF OF THEOREM 3:

Define

$$m_{BIC} = \min\{m : 1 \leq m \leq p; L_{BIC}(m) \geq L_{BIC}(h), h = 1, 2, \dots, p\},$$

where

$$L_{BIC}(m) = BP_{ES}(m) - m \log n.$$

We need to prove that, under the assumptions for Theorem 2,

$$\lim_{n \rightarrow \infty} P(m^* = m_{BIC}) = 1, \quad (15)$$

and that

$$\lim_{n \rightarrow \infty} P(m_{BIC} = 1) = 1. \quad (16)$$

We start by proving (15). Define the event

$$A_n(q) = \left\{ \max_{1 \leq j \leq p} \sqrt{n} |\hat{\rho}_{nj}| > \sqrt{q \log n} \right\}.$$

From Theorem 1 it follows that under H_0

$$\max_{1 \leq j \leq p} \sqrt{n} |\hat{\rho}_{nj}| = O_P(1).$$

Hence $P(A_n(q)) = o(1)$, which implies (15).

To prove that (16) also holds, notice that

$$P(m_{BIC} = 1) = 1 - \sum_{j=2}^p P(m_{BIC} = j) \geq 1 - \sum_{j=2}^p P(L_{BIC}(j) \geq L_{BIC}(1)).$$

Now, for $1 < j \leq p$,

$$\begin{aligned} P(L_{BIC}(j) \geq L_{BIC}(1)) &\leq P(BP_n(j) \geq (j-1) \log(n)) \\ &\leq P\left(n \hat{\rho}_n^{(j)'} \hat{\rho}_n^{(j)} \geq C(j-1) \log(n)\right) + o(1) \\ &= o(1). \end{aligned}$$

The last equality follows from Theorem 1. Therefore, (16) holds, and Theorem 2 follows from an application of the standard CLT for strictly stationary *mds* of Billingsley (1961).

■

TABLES AND FIGURES

Table 1. Backtesting $ES_t(.1)$ and $VaR_t(.05)$ at 5% significance level, $T = 2500$

	t_{ES}	t_{VaR}	$BP_{ES}(1)$	$BP_{VaR}(1)$	$BP_{ES}(3)$	$BP_{VaR}(3)$	$BP_{ES}(5)$	$BP_{VaR}(5)$
$n = 250$								
H_0	0.074	0.085	0.041	0.044	0.053	0.064	0.066	0.073
A_1	0.108	0.102	0.034	0.040	0.217	0.158	0.195	0.131
A_2	0.901	0.824	0.185	0.145	0.251	0.256	0.305	0.286
A_3	0.071	0.091	0.004	0.011	0.284	0.187	0.259	0.171
A_4	0.136	0.110	0.339	0.263	0.377	0.286	0.331	0.259
A_5	0.240	0.213	0.129	0.113	0.200	0.190	0.238	0.215
A_6	1.000	1.000	0.981	0.981	0.995	0.997	0.998	0.997
$n = 500$								
H_0	0.056	0.069	0.036	0.040	0.049	0.048	0.051	0.062
A_1	0.090	0.086	0.046	0.037	0.349	0.213	0.307	0.203
A_2	0.978	0.960	0.211	0.169	0.320	0.272	0.401	0.315
A_3	0.080	0.089	0.029	0.002	0.450	0.294	0.411	0.270
A_4	0.173	0.162	0.489	0.351	0.542	0.391	0.507	0.357
A_5	0.294	0.285	0.131	0.117	0.215	0.203	0.248	0.237
A_6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. Backtesting $ES_t(.05)$ and $Var_t(.025)$ at 5% significance level, $T = 2500$

	t_{ES}	t_{VaR}	$BP_{ES}(1)$	$BP_{VaR}(1)$	$BP_{ES}(3)$	$BP_{VaR}(3)$	$BP_{ES}(5)$	$BP_{VaR}(5)$
$n = 250$								
H_0	0.086	0.065	0.066	0.056	0.075	0.093	0.084	0.095
A_1	0.105	0.062	0.039	0.039	0.142	0.147	0.129	0.126
A_2	0.811	0.636	0.242	0.158	0.282	0.374	0.325	0.382
A_3	0.087	0.063	0.014	0.016	0.236	0.217	0.209	0.180
A_4	0.115	0.073	0.228	0.186	0.255	0.248	0.252	0.215
A_5	0.245	0.194	0.087	0.079	0.152	0.149	0.200	0.180
A_6	1.000	1.000	1.000	0.995	1.000	1.000	1.000	1.000
$n = 500$								
H_0	0.063	0.079	0.052	0.055	0.066	0.083	0.085	0.080
A_1	0.080	0.091	0.050	0.045	0.187	0.152	0.161	0.141
A_2	0.954	0.911	0.260	0.229	0.324	0.234	0.366	0.250
A_3	0.076	0.080	0.005	0.007	0.328	0.255	0.311	0.237
A_4	0.158	0.126	0.312	0.239	0.381	0.308	0.327	0.273
A_5	0.342	0.309	0.087	0.082	0.182	0.181	0.225	0.212
A_6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3. Backtesting $ES_t(.025)$ and $Var_t(.01)$ at 5% significance level, $T = 2500$

	t_{ES}	t_{VaR}	$BP_{ES}(1)$	$BP_{VaR}(1)$	$BP_{ES}(3)$	$BP_{VaR}(3)$	$BP_{ES}(5)$	$BP_{VaR}(5)$
$n = 250$								
H_0	0.112	0.104	0.076	0.127	0.104	0.154	0.120	0.172
A_1	0.131	0.104	0.050	0.113	0.126	0.164	0.145	0.184
A_2	0.726	0.523	0.353	0.523	0.373	0.523	0.386	0.524
A_3	0.123	0.111	0.034	0.107	0.190	0.208	0.192	0.223
A_4	0.103	0.091	0.176	0.190	0.232	0.237	0.228	0.240
A_5	0.229	0.215	0.072	0.084	0.136	0.143	0.171	0.173
A_6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n = 500$								
H_0	0.063	0.079	0.052	0.055	0.066	0.083	0.085	0.080
A_1	0.083	0.151	0.054	0.049	0.149	0.132	0.171	0.164
A_2	0.897	0.857	0.297	0.306	0.322	0.306	0.353	0.307
A_3	0.081	0.146	0.029	0.025	0.246	0.190	0.248	0.219
A_4	0.106	0.137	0.249	0.184	0.309	0.251	0.294	0.249
A_5	0.364	0.329	0.087	0.081	0.179	0.165	0.227	0.206
A_6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. 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
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 5. 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 6. 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 7. p-values for backtesting ES and VaR

S&P500	$ES(0.025)$	$VaR(0.01)$	$ES(0.1)$	$VaR(0.05)$
t	0.011	0.070	0.004	0.010
$BP(5)$	0.007	0.270	0.009	0.052
DAX	$ES(0.025)$	$VaR(0.01)$	$ES(0.1)$	$VaR(0.05)$
t	0.224	0.968	0.045	0.095
$BP(5)$	0.002	0.998	0.091	0.768
HangSeng	$ES(0.025)$	$VaR(0.01)$	$ES(0.1)$	$VaR(0.05)$
t	0.939	0.989	0.194	0.462
$BP(5)$	0.001	0.998	0.002	0.002

Figure 1. Plots of the daily log-returns of SP500, DAX and Hang Seng (HS),
1997.1.1-2009.6.30.

Log-returns, -VaR and -ES of SP500

Figure 2. Log-returns, VaR and ES of S&P500, 2007.7.1-2009.6.30

Figure 3. Log-returns, VaR and ES of DAX, 2007.7.1-2009.6.30

Log-returns, -VaR and -ES of HangSeng

Figure 4. Log-returns, VaR and ES of Hang Seng, 2007.7.1-2009.6.30

Figure 5. Cumulative hits (violations) of S&P500, DAX and Hang Seng (HS),
2007.7.1-2009.6.30

Figure 6. Sample autocorrelations of cumulative hits of S&P500, DAX and Hang Seng (HS),
2007.7.1-2009.6.30

Figure 7. Sample autocorrelations of hits of S&P500, DAX and Hang Seng (HS),
2007.7.1-2009.6.30

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] Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716-723.
- [4] Artzner, P., Delbaen, F., Eber, J-M., Heath, D. (1997) "Thinking coherently", *Risk*, 10, 68-71.
- [5] Artzner, P., Delbaen, F., Eber, J-M., Heath, D. (1999) "Coherent measures of risk", *Mathematical Finance*, 9, 203-228.
- [6] Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models", *Review of Economics and Statistics*, Aug 2003, 85(3): 531-549.
- [7] Basak, S and A Shapiro (2001): "Value-at-risk based risk management: optimal policies and asset prices", *Review of Financial Studies*, 14, 371-405.
- [8] 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>
- [9] Berkowitz, J. (2001), "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economic Statistics*, 19, 465-474.
- [10] Berkowitz, J., Christoffersen, P.F., Pelletier, D. (2011), "Evaluating Value-at-Risk models with desk-level data", *Management Science*, 57, 12, 2213-2227.
- [11] Billingsley, P. (1961), "The Lindeberg-Levy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788-792.
- [12] Bose, A. (1998), "Bahadur Representation of M Estimates", *Annals of Statistics*, 26(2), 771-777.
- [13] 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.

- [14] Christoffersen, P.F., (1998), "Evaluating interval forecasts", *International Economic Review*, 39, 841-862.
- [15] Christoffersen, P.F. (2009), "Value-at-Risk Models", *Handbook of Financial Time Series*, 753-766.
- [16] Du, Z. (2010), "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] Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation", *International Economic Review*, 35, 705–730.
- [25] 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.
- [26] Inglot, T. and Ledwina, T. (2006a), "Towards Data Driven Selection of a Penalty Function for Data Driven Neyman Tests," *Linear Algebra and its Applications*, 417, 124-133.

- [27] Inglot, T. and Ledwina, T. (2006b), "Data Driven Score Tests of Fit for Semiparametric Homoscedastic Linear Regression Model," Preprint 665, *Institute of Mathematics Polish Academy of Sciences*.
- [28] Jorion, P. (2006). Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill, New York.
- [29] Kerkhof, J. ., Melenberg, B. (2004), "Backtesting for risk-based regulatory capital," *Journal of Banking and Finance*, 28, 1845-1865.
- [30] 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.
- [31] Kupiec, P. (1995), "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives*, 3, 73-84.
- [32] Ljung, G.M and Box, G.E.P. (1978), "A Measure of Lack of Fit in Time Series Models", *Biometrika* 65, 297-303.
- [33] McNeil, A., Frey, R. and Embrechts, P. (2005). Quantitative Risk Management, Princeton University Press, Princeton, NJ.
- [34] 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>.
- [35] Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470-472.
- [36] Schwarz, G. (1978), "Estimating the dimension of a model," *Annals of Statistics*, 6, 461-464.
- [37] Shiryaev, A.N. (1996), *Probability*. Springer, New York.
- [38] van der Vaart, A. W. and Wellner, J.A.(1996), *Weak Convergence and Empirical Processes*. Springer, New York.
- [39] Wong, W.K. (2008), "Backtesting trading risk of commercial banks using expected shortfall," *Journal of Banking and Finance*, 32, 1404-1415.

- [40] Wong, W.K. (2010), "Backtesting value-at-risk based on tail losses," *Journal of Empirical Finance*, 17, 526-538.
- [41] Wu, W.B. (2007), "M-Estimation of Linear Models with Dependent Errors", *Annals of Statistics*, Vol. 35, No. 2, 495-521.
- [42] Yamai, Y., Yoshida, T. (2002), "Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization", *Monetary and Economic Studies*, 20, 87-122.
- [43] Yamai, Y., Yoshida, T. (2005), "Value-at-risk versus expected shortfall: A practical perspective," *Journal of Banking and Finance*, 29, 99-115.