

Testing for a serial correlation in VaR failures through the exponential autoregressive conditional duration model

Marta Małecka¹

ABSTRACT

Although regulatory standards, currently developed by the Basel Committee on Banking Supervision, anticipate a shift from VaR to ES, the evaluation of risk models currently remains based on the VaR measure. Motivated by the Basel regulations, we address the issue of VaR backtesting and contribute to the debate by exploring statistical properties of the exponential autoregressive conditional duration (EACD) VaR test. We show that, under the null, the tested parameter lies at the boundary of the parameter space, which can profoundly affect the accuracy of this test. To compensate for this deficiency, a mixture of chi-square distributions is applied. The resulting accuracy improvement allows for the omission of the Monte Carlo simulations used to implement the EACD VaR test in earlier studies, which dramatically improves the computational efficiency of the procedure. We demonstrate that the EACD approach to testing VaR has the potential to enhance statistical inference in most problematic cases – for small samples and for those close to the null.

Key words: VaR backtesting, exponential autoregressive conditional duration, boundary of the parameter space, test size, test power.

1. Introduction

Value-at-Risk (VaR) and Expected Shortfall (ES) are two measures of market risk that dominate contemporary banking regulation. Since its original inception in business (JP Morgan, 1994) and incorporation to regulatory standards (Basel Committee on Banking Supervision, 1996), VaR has become an industry standard in market risk management. Its constantly widening range of applications include new types of risk and new markets. Despite its widespread use, however, it has several flaws. It does not take account of losses beyond a designated threshold as well as lacks subadditivity, which means that diversification does not, necessarily, imply reduction of risk. Therefore, ES, which remedies this problems, seems to be emerging as a new standard. In the light of the major reform of global supervisory standards, pursued

¹ Department of Statistical Methods, University of Łódź, Poland. E-mail: marta.malecka@uni.lodz.pl.
ORCID: <http://orcid.org/0000-0003-4465-9811>.

by the Basel Committee since 2012 (Basel Committee on Banking Supervision, 2012-2017), ES is recommended for reporting exposures to market risk. Nevertheless, ES fails to satisfy a different mathematical principle – elicibility. Although this criterion has been shown to be erroneously deemed essential to backtesting (Gneiting, 2011), the question about ES-based statistical tests remains open (Acerbi and Szekely, 2014, Chen, 2014, Fissler and Ziegel, 2015, Fissler et al., 2016). No consensus on relevant procedures has yet been reached, either in academic studies or in business practice. Therefore, evaluation of risk models still relies on VaR. In an attempt to include most extreme losses, the regulator has recommended testing VaR on two low coverage levels – 1% and 2.5%. These Basel regulations motivate academics to review, develop and enhance statistical methods of backtesting VaR.

VaR backtesting procedures commonly refer to two criteria: the postulate of unconditional coverage, which treats the overall fraction of VaR violations, and the postulate of conditional coverage, which addresses their serial dependence. Perhaps of greater practical importance is detecting serial correlation of VaR failures, for their clustering may result in a series of catastrophic losses occurring one by one. This, in turn, seriously increases the risk of bankruptcy of a financial institution. The Markov test, which embeds the iid Bernoulli hypothesis within a binary first-order Markov chain and utilizes the likelihood ratio framework, has become the industry standard for testing the conditional coverage property (Christoffersen, 1998). This standard test, however, has been shown to exhibit unsatisfactory power (Lopez, 1999, Christoffersen and Pelletier, 2004, Berkowitz et al., 2011, Pajhede, 2017), which boosted the debate on other possibilities of testing VaR conditional coverage. Among other directions, like spectral tests (Berkowitz et al., 2011, Gordy and McNeil, 2018) or multi-level tests (Berkowitz, 2001, Hurlin and Tokpavi, 2007, Colletaz et al., 2013, Leccadito, Boffelli and Urga, 2014, Wied, Wei and Ziggel, 2016, Kratz et al. 2018), the duration-based approach attracted much attention in the scientific community (Christoffersen and Pelletier, 2004, Candelon et al. 2011, Pelletier and Wei, 2016). In the duration-based framework the sequence of VaR violations is transformed into the duration series. The idea behind this approach follows from the observation that the time that has passed since a VaR violation (hit) should not contain any information about further duration of the no-hit sequence. This implies the memory-free property of the duration series. Within discrete distributions, this property characterizes the geometric distribution (Berkowitz et al., 2011), while the only memory-free continuous distribution is the exponential distribution. To test VaR by means of the exponential distribution it has been proposed to nest the memory free null in the exponential autoregressive conditional model (EACD model, Engle and Russel, 1998). The EACD VaR test has been shown to compare favourably, in terms of its power, to other duration-based tests like the Weibull of the gamma test, especially for small sample sizes (Christoffersen and

Pelletier, 2004, Małecka, 2018). This test, however, suffers from significant size distortions, which means that the asymptotic distribution does not guarantee the correct test level. To make up for this deficiency it has been proposed to use the Monte Carlo method to simulate the null distribution of the test statistic (Christoffersen and Pelletier, 2004). The Monte Carlo approach, however, while ensuring the correct test level, impedes practical implementation of the procedure.

Our work addresses applicability of the exponential autoregressive conditional model to testing for serial correlation in VaR failure series. The goals of the paper are twofold: firstly, we seek to handle the problem of EACD test size distortions without resorting to the use of Monte Carlo simulations and secondly we investigate its power in relation to the standard VaR backtesting procedure. To avoid p-value computation through simulations, we study the asymptotic properties of the test statistic. Exploiting the fact that, in the VaR testing framework, the null value of the parameter vector lies exactly at the boundary of the parameter space, we show that the test statistic does not converge to the standard likelihood ratio (LR) limiting distribution. Using results on asymptotic LR properties under non-regular conditions (Self and Liang, 1987), we suggest p-value computation from the mixture of two chi-square distributions. We experimentally demonstrate the size improvement obtained by the proposed approach.

Given improved accuracy of the EACD VaR test, we investigate its power properties. To mimic a typical VaR failure correlation scheme, we adopt a GARCH model. The comparative evaluation of the EACD test power is conducted in relation to the Markov procedure, which has, so far, won widest recognition in the industry. We indicate cases where the EACD approach allows for power gains, which gives guidance as to practical application of the examined procedures.

Our study is based on earlier works by Christoffersen and Pelletier (2004) and Małecka (2018). The results of Christoffersen and Pelletier are improved by using asymptotic LR properties under non-regular conditions and implementing the EACD VaR test with the limiting mixture distribution. Since this replaces Monte Carlo simulations, our approach improves computational effectiveness of the procedure and facilitates its practical implementation. The results of Christoffersen and Pelletier are also improved by replacing the historical simulation model in the power study with the GARCH-model-based experiment. In this way we obtain the realistic setting, which mimics the volatility clustering of real financial data. In this experiment the serial correlation of VaR failures, as in reality, results from the volatility clustering of the portfolio returns. The volatility clustering is measured by the correlation coefficient of the squared returns, which, in the model we use, can be calculated analytically. Therefore, we are able to study the power of the test as a function of a controlled parameter of the return distribution, which is not attainable with the historical simulation experiment.

We extend the study by Małecka (2018) with respect to the contemporary international regulations in banking supervision. In addition to the typical 5% VaR, we include evaluation of test properties for two lower VaR coverage levels, indicated in the Basel rules. We discuss test accuracy in the context of the coverage level. We also extend the earlier study by depicting powers of the test as a function of volatility clustering. The shapes of the functions, compared to the power function of the standard Markov test, indicate cases where the EACD approach allows for more effective detection of incorrect risk models.

The paper proceeds as follows. Section 2 introduces the notation and presents the duration-based approach to VaR backtesting in relation to the standard Markov procedure. It shows the applicability of the EACD model to testing VaR and discusses the asymptotic distribution of the test statistic. Section 3 provides the study of test properties. Firstly, it details the design of the Monte Carlo experiment, showing a way to control volatility clustering. Secondly, it addresses test accuracy and presents improvements obtained by the use of the asymptotic mixture distribution. Finally, it gives comparative evaluation of test power in relation to the Markov test. The final section summarizes and concludes.

2. Testing VaR Conditional Coverage: EACD vs. Markov-Chain Approach

Let $\{R_t\}$ be the asset or portfolio return process, for which VaR at time t , at the level of tolerance p , is defined as the p -quantile of the relevant return distribution:

$$P(R_t < VaR_t(p)) = p, \quad t = 1, \dots, T. \quad (1)$$

Then, the VaR evaluation framework is based on the stochastic process of VaR failures:

$$I_t = \begin{cases} 1, & R_t < VaR_t(p) \\ 0, & R_t \geq VaR_t(p) \end{cases}, \quad (2)$$

whose realization is referred to as a hit sequence.

The standard Christoffersen's (1998) Markov test of VaR failure independence uses the framework of the binary Markov chain with the transition matrix:

$$\begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}, \quad (3)$$

where π_{ij} denotes the probability of a single-step transition from state i to state j , $i, j \in \{0, 1\}$. The null hypothesis of equal transition probabilities $H_0 : \pi_{01} = \pi_{11}$ implies

the iid Bernoulli process with probability of VaR violence $\pi_1 = \pi_{01} = \pi_{11}$. To verify the above parameter restriction it has been proposed to use the likelihood ratio statistics:

$$LR_{ind} = -2\log \frac{\hat{\pi}_1^{t_1}(1-\hat{\pi}_1)^{t_0}}{\hat{\pi}_{01}^{t_{01}}(1-\hat{\pi}_{01})^{t_{00}}\hat{\pi}_{11}^{t_{11}}(1-\hat{\pi}_{11})^{t_{10}}} \sim_{as} \chi^2_{(1)}, \tag{4}$$

where $\hat{\pi}_1 = \frac{t_1}{t_0 + t_1}$, t_0 is the number of non-exceptions, t_1 – the number of exceptions, $\hat{\pi}_{01} = \frac{t_{01}}{t_0}$, $\hat{\pi}_{11} = \frac{t_{11}}{t_1}$ and t_{ij} – the number of transitions form state i to state j .

The construction of the Markov test implies that it only allows for detecting cases where the hit sequence follows a simple first-order Markov chain. A duration-based approach was proposed as means to capture more general forms of dependence. The duration-based tests use the transformation of the underlying $\{I_t\}$ process into the duration series $\{V_i\}$ defined as:

$$V_i = t_i - t_{i-1}, \tag{5}$$

where t_i denotes the time of the i -th VaR violation. The independence of the $\{I_t\}$ process implies that the time that has passed since a VaR violation (hit) should not contain any information about the further duration of the no-hit sequence. This memory-free property of the duration series motivates the use of the exponential distribution. In the exponential autoregressive conditional test the memory free null is tested against the alternative of the exponential process with a conditional mean. Exploiting the fact that the serially dependent hit sequence is likely to produce an excessive number of relatively short no-hit durations and relatively long no-hit durations, the test checks the autoregression coefficient of the conditional mean of the duration. The EACD approach utilizes the regression of the form:

$$E_{i-1}(V_i) = a + bV_{i-1} \tag{6}$$

(Engle and Russel, 1998). It assumes the exponential distribution, which gives the following conditional pdf function of the duration V_i :

$$f_{EACD}(v_i) = \frac{1}{a + bV_{i-1}} e^{-\frac{v_i}{a + bV_{i-1}}}. \tag{7}$$

Under the null hypothesis $H_0 : b = 0$ the conditional distribution becomes the exponential distribution with a constant mean.

By using the regression of the durations on their past values this test incorporates the information about the ordering of VaR failures. This offers potential power gains over other duration based procedures like the Weibull test or the gamma test, that simply nest the exponential distribution in wider distribution families and verify relevant restrictions.

The EACD-based VaR test verifies the parameter restriction through the likelihood ratio statistic, which requires computation of the loglikelihood function for the unrestricted and restricted case. Taking account of possible presence of censored durations at the beginning and at the end of the series, the loglikelihood takes the form:

$$\begin{aligned} \log L(V, \theta) = & C_1 \log S(V_1) + (1 - C_1) \log f(V_1) + \sum_{i=2}^{N-1} \log f(V_i) + \\ & + C_N \log S(V_N) + (1 - C_N) \log f(V_N), \end{aligned} \quad (8)$$

where C_i is 1 if the duration V_i is censored and 0 otherwise, S is the survival function of the variable V_i , N is the number of VaR failures and θ is the vector of parameters (Christoffersen and Pelletier, 2004).

Assuming parameter values in the interior of the parameter space, the likelihood ratio statistic for one parameter restriction has the chi-square distribution with one degree of freedom χ_1^2 . However, if the tested parameter value lies at or near the boundary of the parameter space, the asymptotic convergence to the chi-square distribution ceases to hold true. This is the case with the EACD VaR test since the null hypothesis imposes the zero value of the autoregression coefficient, and, at the same time, the coefficient satisfies the nonnegativity condition. This means that the vector of EACD model parameters $\theta = [a, b]$ belongs to the space $\Theta = (0, +\infty) \times [0, +\infty)$, which, under the null, reduces to $\Theta_0 = (0, +\infty) \times \{0\}$. In such a case statistical inference based on the asymptotic χ_1^2 may be inaccurate. To overcome the problem of potential size distortions, the EACD VaR test has been originally implemented with the use of the Monte Carlo simulated p-values. Instead, using asymptotic results on the likelihood ratio distribution under non-standard conditions (Self and Liang, 1987), we propose to compute the p-values from the 50:50 mixture of chi-square distributions, with zero and one degrees of freedom:

$$LR_{EACD} \sim as 0.5\chi_0^2 + 0.5\chi_1^2. \quad (9)$$

Using the fact that the chi-square distribution with zero degrees of freedom reduces to the distribution with all its mass cumulated at zero, we get that the value of the test with 50% probability takes the value of 0 and with 50% probability is drawn from the chi-square distribution with one degree of freedom χ_1^2 .

3. Monte Carlo Study of Test Properties

The tests described in Section 2 verify the conditional coverage property of VaR failures referring to the Markov chain framework or, after the transformation of the hit sequence into durations, to the exponential autoregressive conditional duration model. Since the two tests exploit different approaches and make use of different variables, they are likely to differ in power properties. Moreover, as they rely on asymptotic distributions, their finite sample properties are unknown. In the present section, using a finite sample setting, we evaluate and compare the statistical properties of the two tests through the Monte Carlo study. The comparative analysis includes their size and power. We discuss practical implications of the power properties, presenting conclusions as to when to prefer which of the two tests and indicating cases when the two approaches may complement each other.

The finite-sample statistical properties of the tests are evaluated for sample sizes chosen to be realistic for applications in finance: $T = 250, 500, \dots, 1500$. Such samples roughly correspond to daily data covering periods from one year to six years. The size and the power of the tests are approximated by rejection frequencies under the null and under the alternative, respectively. The size study includes significance levels 0.01, 0.05 and 0.1. For powers of the tests, only rejection rates at 0.05 significance level are reported. The size and the power estimates are computed over 10000 Monte Carlo trials.

The size study examines test rejection probabilities when the risk model is correct. We refer to a test as accurate if, under the correct model, the rejection probability corresponds to the assumed level of significance (nominal test size). Therefore, the size study requires generating $\{I_t\}$ series under the correct model, i.e. under the assumptions of the true failure probability and independence of VaR violations. To this end we use the Bernoulli distribution with the probability of success p , equal to the assumed level of VaR tolerance.

The size estimates obtained from the Bernoulli experiment (Tables 1-3) show the accuracy improvement of the EACD test gained by replacing the χ_1^2 distribution by the mixture of distributions $0.5\chi_0^2 + 0.5\chi_1^2$. In the case of the χ_1^2 the procedure is very conservative with the true test level leaning towards zero. This size distortion indicates that practical application of this test should not be based on the asymptotic χ_1^2 distribution. Employment of the mixture $0.5\chi_0^2 + 0.5\chi_1^2$ has the effect that the true test level approaches the nominal size. The test still tends to underreject the null, however the discrepancies between the simulated and the nominal size markedly decrease and the simulated rejection frequencies seem to converge to the desired level with lengthening the sample. The improvement in the accuracy of the test is demonstrated through the fit of the asymptotic and the empirical distribution function, based on a 1500 observation sample (Figure 1).

Table 1. Size estimates for Markov and EACD 1% VaR tests*

Test	Significance level $\alpha = 0.01$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0111	0.0122	0.0116	0.0124	0.0128	0.0128
$LR_{EACD}^{Chi\ square}$	0.0000	0.0000	0.0009	0.0010	0.0018	0.0019
$LR_{EACD}^{Mixture}$	0.0000	0.0008	0.0024	0.0034	0.0051	0.0047
Test	Significance level $\alpha = 0.05$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0234	0.0248	0.0293	0.0269	0.0209	0.0210
$LR_{EACD}^{Chi\ square}$	0.0002	0.0016	0.0085	0.0110	0.0131	0.0146
$LR_{EACD}^{Mixture}$	0.0013	0.0077	0.0203	0.0248	0.0298	0.0358
Test	Significance level $\alpha = 0.1$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0294	0.0418	0.0474	0.0480	0.0425	0.0453
$LR_{EACD}^{Chi\ square}$	0.0013	0.0077	0.0203	0.0248	0.0298	0.0358
$LR_{EACD}^{Mixture}$	0.0080	0.0280	0.0487	0.0616	0.0683	0.0796

* $LR_{EACD}^{Chi\ square}$ denotes the cases when the LR_{EACD} test size was estimated under the χ_1^2 distribution, while $LR_{EACD}^{Mixture}$ – the cases when the size was estimated under the mixture distribution $0.5\chi_0^2 + 0.5\chi_1^2$.

Source: Own work.

Table 2. Size estimates for Markov and EACD 2.5% VaR tests*

Test	Significance level $\alpha = 0.01$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0265	0.0293	0.0310	0.0284	0.0274	0.0274
$LR_{EACD}^{Chi\ square}$	0.0002	0.0007	0.0011	0.0009	0.0019	0.0022
$LR_{EACD}^{Mixture}$	0.0008	0.0016	0.0024	0.0030	0.0042	0.0047
Test	Significance level $\alpha = 0.05$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0393	0.0447	0.0443	0.0448	0.0424	0.0430
$LR_{EACD}^{Chi\ square}$	0.0032	0.0052	0.0079	0.0097	0.0104	0.0119
$LR_{EACD}^{Mixture}$	0.0077	0.0126	0.0189	0.0234	0.0241	0.0253

Table 2. Size estimates for Markov and EACD 2.5% VaR tests* (cont.)

Test	Significance level $\alpha = 0.1$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0526	0.0635	0.0640	0.0685	0.0784	0.0951
$LR_{EACD}^{Chi\ square}$	0.0077	0.0126	0.0189	0.0234	0.0241	0.0253
$LR_{EACD}^{Mixture}$	0.0203	0.0329	0.0429	0.0526	0.0557	0.0552

* $LR_{EACD}^{Chi\ square}$ denotes the cases when the LR_{EACD} test size was estimated under the χ_1^2 distribution, while $LR_{EACD}^{Mixture}$ – the cases when the size was estimated under the mixture distribution $0.5\chi_0^2 + 0.5\chi_1^2$.

Source: Own work.

Table 3. Size estimates for Markov and EACD 5% VaR tests*

Test	Significance level $\alpha = 0.01$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0194	0.0293	0.0281	0.0340	0.0368	0.0400
$LR_{EACD}^{Chi\ square}$	0.0006	0.0008	0.0013	0.0014	0.0026	0.0017
$LR_{EACD}^{Mixture}$	0.0016	0.0036	0.0038	0.0046	0.0055	0.0061
Test	Significance level $\alpha = 0.05$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.0707	0.094	0.1199	0.1276	0.1233	0.1147
$LR_{EACD}^{Chi\ square}$	0.0054	0.0068	0.0105	0.0108	0.0139	0.0142
$LR_{EACD}^{Mixture}$	0.0115	0.0196	0.0258	0.0279	0.0301	0.0308
Test	Significance level $\alpha = 0.1$					
	Series length					
	250	500	750	1000	1250	1500
LR_{Ind}	0.1012	0.1797	0.2033	0.1792	0.1656	0.1654
$LR_{EACD}^{Chi\ square}$	0.0114	0.0197	0.0209	0.0265	0.0298	0.0289
$LR_{EACD}^{Mixture}$	0.0321	0.0483	0.058	0.062	0.0646	0.0675

* $LR_{EACD}^{Chi\ square}$ denotes the cases when the LR_{EACD} test size was estimated under the χ_1^2 distribution, while $LR_{EACD}^{Mixture}$ – the cases when the size was estimated under the mixture distribution $0.5\chi_0^2 + 0.5\chi_1^2$.

Source: Own work.

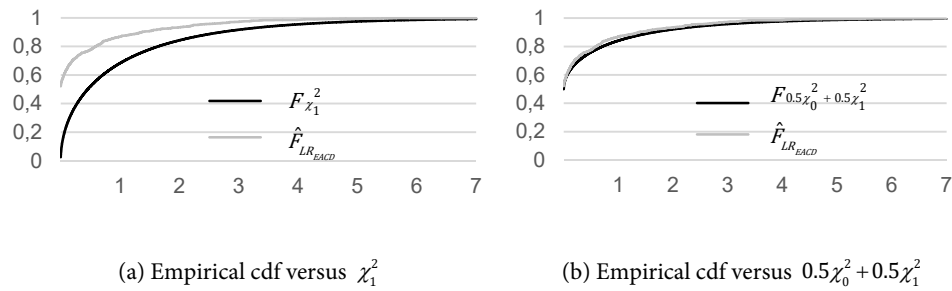


Figure 1. Empirical cdf of the LR_{EACD} test statistic, based on 1,500 observations, versus χ_1^2 and $0.5\chi_0^2 + 0.5\chi_1^2$ asymptotic distributions.

Source: Own work.

Comparative evaluation of the size results in relation to the relevant outcomes for the benchmark Markov procedure shows that the EACD $LR_{EACD}^{Mixture}$ rejection frequencies tend to be closer to the nominal test size. The EACD test seems also more reliable as the relation of the estimated size to the chosen significance level is remarkably stable across significance levels and VaR coverage levels, especially for large samples. For the Markov LR_{Ind} this relation changes rapidly with both: chosen significance and VaR coverage. Contrary to the systematically undervalued but apparently convergent $LR_{EACD}^{Mixture}$ rejection frequencies, the LR_{Ind} test changes from being undersized to being oversized. Its rejection frequencies are much overvalued for 5% VaR, while for lower coverage levels they shift from being overvalued in tails to undervalued closer to the central area of the distribution. The differences between the estimated and the nominal size of LR_{Ind} range from minor – for 1% VaR and 0.01 significance – to large – for 5% VaR and 0.01 significance. In the last case the estimated size overvalues the nominal significance four times. Therefore, in the light of the results for all considered significance and coverage levels, the EACD approach to testing VaR offers accuracy improvement in comparison to the standard Markov-chain-based VaR test.

The size improvement attainable with the proposed method confirms that the asymptotic mixture distribution works well for the EACD test. However, this method does not solve the problem of the inaccurate size for small samples. Our results show that this problem cannot be handled by any of the considered approaches. In particular, none of the asymptotic approximations, including both standard likelihood ratio distribution and the non-standard mixture distribution, is relevant for daily data covering a one-year period. Therefore, if the sample size is limited, it seems recommendable to resort to the Monte-Carlo-based methods.

The comparison of the size results across 1%, 2.5% and 5% coverage levels shows that the EACD $LR_{EACD}^{Mixture}$ test performs best for 5% VaR. In this case, the observed size

distortions are lowest and convergence to the desired levels is fastest. This is particularly visible for popular significance levels 0.01 and 0.05. The recommended VaR coverage, for EACD-based VaR backtesting, is thus not in line with contemporary trends in banking supervision. The lower coverage levels, suggested by the Basel rules, lead to $LR_{EACD}^{Mixture}$ accuracy loss.

The power study investigates test performance under the assumption of an incorrect risk model. The test is regarded as more effective if its rejection frequencies, under the incorrect model, are higher. Since the considered procedures are aimed at checking the conditional coverage property, we study their ability to reject clustered VaR violations. Therefore, the simulation experiment in the power study is designed to reflect the volatility clustering of return data. The volatility clustering, in turn, implies the undesired serial correlation of VaR violations.

The $\{I_t\}$ series under the incorrect model is computed as the hit sequence from the GARCH(1,1) return process and the constant VaR level. The VaR level is set to the value of the unconditional p -quantile of the returns. This produces VaR failures with the tendency to correlate in time and, at the same time, guarantees the correct overall VaR failure rate. Through employment of the GARCH process we obtain the realistic setting, which mimics the volatility clustering of real financial data. The volatility clustering is measured by the correlation coefficient of the squared returns, which, under the specification we use, can be calculated analytically. This enables us to study the power of the test as a function of a controlled parameter of the return distribution. In order to calculate analytically the correlation coefficient of the squared returns, we use the GARCH model of the form:

$$\begin{aligned} R_t &= \sqrt{h_t} Z_t, \quad Z_t : N(0,1), \\ h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \end{aligned} \tag{10}$$

Under specification (10) the correlation of the squared returns ρ is given by

$$\rho = \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2}. \tag{11}$$

This is, however, subject to the restriction

$$(\alpha + \beta)^2 + 2\alpha^2 < 1 \tag{12}$$

If condition (12) does not hold, the correlations of the GARCH model are time-varying. In such a case, they have been shown to behave approximately as:

$$\rho = \alpha + \frac{\beta}{3}. \tag{13}$$

We choose the GARCH parameters on realistic levels $\omega = 0.01$, $\beta = 0.85$ and the correlation coefficient ρ to vary from 0.05 to 0.5. The α parameter is set to levels that

guarantee the desired value of ρ . Under the above parameter values the restriction (12) holds for correlations not higher than 0.4, thus a range of values from 0.05 to 0.4 is used.

Due to observed size distortions, in the power exercise we adopt the Monte Carlo test technique, which provides exact tests by replacing theoretical null distributions of test statistics by their sample analogues [6]. Through ensuring a correct test level we obtain comparability of the power results. Since estimating EACD model parameters requires at least three durations, which corresponds to at least two VaR violations, there are cases when the test is not feasible. Rejecting these cases constitutes a non-random sample selection rule. Therefore, we present effective power rates, which correspond to multiplying raw power by the rate of valid test runs. Referring to the results of the size study, in the power exercise we rely on 5% VaR. We report rejection frequencies for 0.05 significance level.

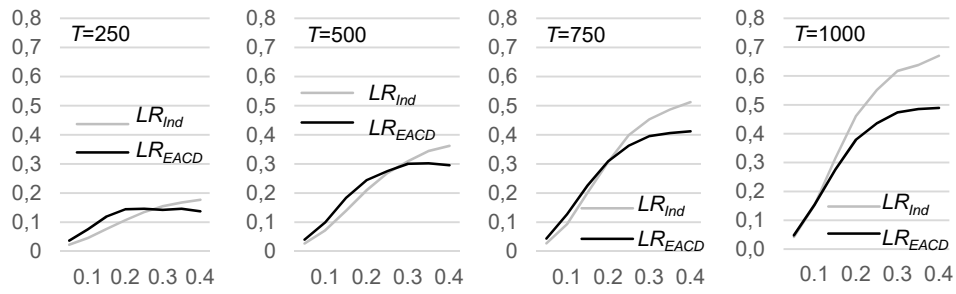
The power of the EACD LR_{EACD} test is evaluated in relation to rejection frequencies of the Markov LR_{Ind} test (Table 4). The results show superiority of the EACD procedure at short distances from the null. In the case of volatility clustering corresponding to 0.05 and 0.1 correlation of the squared returns, the EACD procedure exhibits higher power than the benchmark for all series lengths. Subsequent experiments show that this comparative advantage tends to vanish for a stronger correlation. However, it is observed relatively long for small samples.

Table 4. Power estimates for 5% VaR Markov and EACD tests on 0.05 significance level

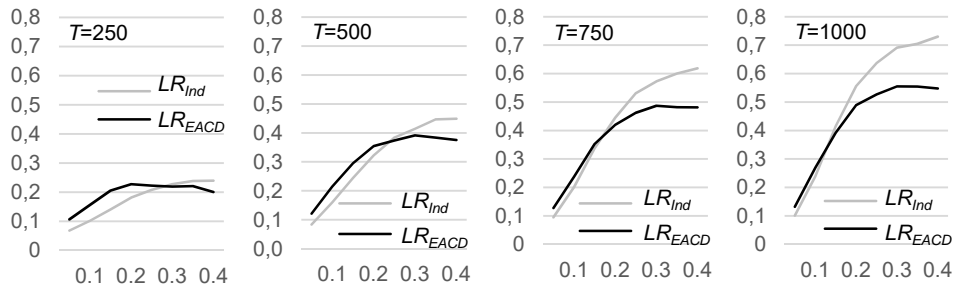
Test	Volatility clustering*	Series Length					
		250	500	750	1000	1250	1500
LR_{Ind}	0.05	0.068**	0.083**	0.095**	0.101**	0.102**	0.104**
	0.10	0.102**	0.160**	0.201**	0.239**	0.281**	0.316**
	0.15	0.141**	0.244**	0.337**	0.415	0.487	0.546
	0.20	0.181**	0.322**	0.445	0.557	0.633	0.708
	0.25	0.208**	0.384	0.531	0.637	0.722	0.803
	0.30	0.228	0.414	0.573	0.691	0.775	0.836
	0.35	0.239	0.447	0.600	0.705	0.787	0.859
	0.40	0.240	0.449	0.619	0.730	0.807	0.866
LR_{EACD}	0.05	0.107**	0.121**	0.128**	0.132**	0.137**	0.155**
	0.10	0.156**	0.215**	0.238**	0.269**	0.287**	0.321**
	0.15	0.206**	0.295**	0.353*	0.392	0.439	0.477
	0.20	0.228**	0.355**	0.420*	0.490	0.523	0.583
	0.25	0.223**	0.374	0.463	0.527	0.582	0.635
	0.30	0.220	0.392	0.487	0.555	0.618	0.657
	0.35	0.221	0.384	0.482	0.555	0.615	0.668
	0.40	0.201	0.375	0.482	0.548	0.614	0.662

*The volatility clustering in the simulated process is measured by the correlation coefficient of the squared returns ρ .

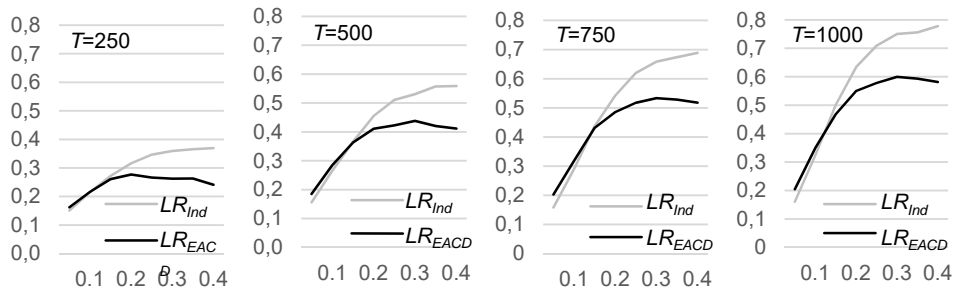
**Cases when the estimated power of LR_{EACD} exceeds that of LR_{Ind} are marked with double asterix.



(a) Significance level $\alpha=0.01$



(b) Significance level $\alpha=0.05$



(c) Significance level $\alpha=0.1$

Figure 2. Power function estimates against volatility clustering for VaR tests, $T=250, \dots, 1000$.

Source: Own work.

The relative performance of the tests is depicted by the power functions plotted against the strength of volatility clustering (Figure 2). The figures extend the power study, illustrating estimated powers for three significance levels: 0.01, 0.05 and 0.1. The sample sizes range from 250 to 1000, as for longer series the test performance is

relatively stable and follows the trends observed for 1000 sample. All plots confirm that the EACD power tends to grow faster than the Markov test power close to the null. Thus, this test is likely to outperform the benchmark procedure in detecting low-scale correlations. Its advantage in power against low-scale correlations is especially large for small samples. This suggests that the EACD approach has the potential to improve testing efficiency in cases when statistical inference is particularly troublesome – for small samples and close to the null.

The power functions illustrate also the downswing in the EACD test performance for largest correlations. This suggests that after exceeding some critical value of the correlation of the squared returns, the test power starts to deteriorate. Large correlation in squared returns is likely to produce VaR violations occurring one by one in turbulent periods, followed by long calm periods without any violation. This translates into a typical setup of a duration sequence with series of very short durations from turbulent time, interrupted by one long duration, corresponding to the calm period. The single outstanding duration replaces series of long durations. In such a setting the autoregressive model of durations tends to be insignificant. Thus, the excessive correlation of the squared returns works against the power of the test. This supports the practical conclusion that the EACD approach to testing VaR is particularly recommendable for detecting low-scale distortions from the null.

Combining the size and the power results, the EACD procedure seems complementary to the standard Markov test, as their relative performance depends on the distance from the null. At the null, the EACD test outperforms the benchmark procedure. This means that it is less likely to overreject the correct risk model. Of practical importance is the fact that its performance at the null is remarkably stable across significance levels and VaR coverage levels. Thus, its accuracy is only slightly influenced by the user's parameter choices. Close to the null, the EACD power grows quickly, which makes the test more sensitive to low-scale correlations. In practice it means that it is more likely to detect incorrect risk models when clustering of VaR failures is relatively small. On the other hand, the Markov test performs better at detecting a large correlation of VaR violations. Thus, in the light of their statistical properties, it is advisable that the procedures be employed simultaneously in testing conditional coverage property. Another practical guideline from our results is that the contrary decisions of the two tests may occur due to low-scale correlation rather than the type one error. Therefore, such an outcome of the backtesting procedure signals that the risk model should be recognized as incorrect.

4. Conclusion

The paper tackled the issue of evaluating risk models with respect to the contemporary changes in international banking regulation. In accordance with the Basel recommendations, we inquired into ways of assessing risk models based on the VaR measure. In this context we studied applicability of the EACD model. We considered the EACD test as means of testing conditional coverage property of VaR violations. We addressed the construction, asymptotic distribution as well as the finite sample size and power properties of the test.

With reference to the accuracy of backtesting, we sought to handle the problem of EACD test size distortions without resorting to the use of Monte Carlo simulations. Based on the observation that the conditional coverage property implies the parameter restriction that lies at the boundary of the parameter space, we suggested p-value computation from the mixture of chi-square distributions. In this way we obtained the procedure which is both accurate and computationally effective as it replaces the originally proposed Monte Carlo method. Since its construction is based on the duration series instead of the hit sequence, it also has the potential to exhibit power against more general forms of dependence than the standard VaR test, which operates within the framework of the first order Markov chain.

Via simulations we showed improvement in the test accuracy owned to replacing the asymptotic likelihood ratio distribution with a mixture of chi-square distributions. We confirmed the convergence of the true test level to the nominal size of the test. With the use of the GARCH model we designed the experiment, which enabled us to study the power of the tests against various levels of volatility clustering in return data. The estimated power functions showed that the EACD test outperforms the benchmark Markov procedure at the null and its power grows faster close to the null. Thus, this procedure may be useful to detect low-scale correlations and in this sense it may complement the standard Markov test. This comparative advantage of the EACD test turned out to be particularly large for shortest examined series lengths. Therefore, our results suggested that the EACD approach to VaR testing may aid statistical inference in most troublesome cases – for small samples and close to the null.

References

- ACERBI, C., SZEKELY, B., (2014). Backtesting Expected Shortfall, *Risk*, November.
- BASEL COMMITTEE ON BANKING SUPERVISION, (1996). Amendment to the capital accord to incorporate market risks: Technical document, available online: <http://www.bis.org/publ/bcbs24.pdf> (accessed June 4, 2018).

- BASEL COMMITTEE ON BANKING SUPERVISION, (2012). Fundamental Review of the Trading Book: Technical document, available online: <http://www.bis.org/publ/bcbs219.pdf> (accessed June 4, 2018).
- BASEL COMMITTEE ON BANKING SUPERVISION, (2013). Fundamental Review of the Trading Book: A Revised Market Risk Framework: Technical document, available online: <http://www.bis.org/publ/bcbs265.pdf> (accessed June 4, 2018).
- BASEL COMMITTEE ON BANKING SUPERVISION, (2014). Fundamental Review of the Trading Book: Outstanding Issues: Technical document, available online: <http://www.bis.org/bcbs/publ/d305.pdf> (accessed June 4, 2018).
- BASEL COMMITTEE ON BANKING SUPERVISION, (2015). Fundamental review of the trading book - interim impact analysis: Technical document, available online: <http://www.bis.org/bcbs/publ/d346.pdf> (accessed June 4, 2018).
- BASEL COMMITTEE ON BANKING SUPERVISION, (2016). Minimum capital requirements for market risk: Technical document, available online: <http://www.bis.org/bcbs/publ/d352.pdf> (accessed June 4, 2018).
- BASEL COMMITTEE ON BANKING SUPERVISION, (2017). High-level summary of Basel III Reforms: Technical document, available online: https://www.bis.org/bcbs/publ/d424_hlsummary.pdf (accessed June 4, 2018).
- BERKOWITZ, J., (2001). Testing Density Forecasts with Applications to Risk Management, *J Bus Econ Stat*, Vol. 19(4), pp. 465–474, doi: <https://dx.doi.org/10.1198/07350010152596718>.
- BERKOWITZ, J., CHRISTOFFERSEN, P., PELLETIER, D., (2011). Evaluating Value-at-Risk Models with Desk-Level Data, *Manage Sci*, Vol. 12(57), pp. 2213–2227, doi: <https://dx.doi.org/10.1287/mnsc.1080.0964>.
- CANDELON, B., COLLETAZ, G., HURLIN, C., TOKPAVI, S., (2011). Backtesting Value-at-Risk: a GMM duration-based test, *J Financ Economet*, Vol. 9(2), pp. 314–343, doi: <https://doi.org/10.1093/jfinec/nbq025>.
- CHEN, J. M., (2014). Measuring market risk under the Basel accords: VaR, stressed VaR, and expected shortfall. *Aestimatio, The IEB International Journal of Finance*, Vol. 8, pp.184–201, doi: <https://doi.org/10.2139/ssrn.2252463>.
- CHRISTOFFERSEN, P., (1998). Evaluating Interval Forecasts, *Int Econ Rev*, Vol. 39(4), pp. 841–862, doi: <https://doi.org/10.2307/2527341>.

- CHRISTOFFERSEN, P., PELLETIER, D., (2004). Backtesting Value-at-Risk: A Duration-Based Approach, *J Financ Economet*, Vol. 2(1), pp. 84–108, doi: <https://doi.org/10.1093/jjfinec/nbh004>.
- COLLETAZ, G., HURLIN, C., PERIGNON, C., (2013). The Risk Map: a New Tool for Risk Management, *J Bank Financ*, Vol. 37(10), pp. 3843–3854, doi: <https://doi.org/10.1016/j.jbankfin.2013.06.006>.
- DUFOUR, J. M., (2006). Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics, *J Econometrics*, Vol. 133(2), pp. 443–477, doi: <https://doi.org/10.1016/j.jeconom.2005.06.007>.
- ENGLE, R. F., RUSSEL, J. R., (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, *Econometrica*, Vol. 66(5), pp. 1127–62, doi: <https://doi.org/10.2307/2999632>.
- FISLER, T., ZIEGEL, J. F., GNEITING, T., (2016). Expected shortfall is jointly elicitable with value at risk – Implications for backtesting, *Risk*, Vol. 29, pp. 58–61.
- FISLER, T., ZIEGEL, J. F., (2016). Higher order elicibility and Osband’s principle, *Ann Stat*, Vol. 44(4), pp. 1680–707, doi: <https://doi.org/10.1214/16-AOS1439>.
- GNEITING, T., (2011). Making and evaluating point forecasts, *J Am Stat Assoc*, Vol. 106(494), pp. 746–762, doi: <https://doi.org/10.1198/jasa.2011.r10138>.
- GORDY, M. B., MCNEIL, A. J., (2018). Spectral Backtests of Forecast Distributions with Application to Risk Management, in: *Finance and Economics Discussion Series 2018-021*, Board of Governors of the Federal Reserve System, Washington.
- HURLIN, CH., TOKPAVI, S., (2007). Backtesting value-at-risk accuracy: a simple new test, *J Risk*, Vol. 9(2), pp. 19–37, doi: <https://doi.org/10.21314/JOR.2007.148>.
- KRATZ, M., LOK, Y. H., MCNEIL, A. J., (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall, *J Bank Financ*, Vol. 88, pp. 393–407, doi: <https://doi.org/10.1016/j.jbankfin.2018.01.002>.
- LECCADITO, A., BOFFELLI, S., URGA, G., (2014). Evaluating the Accuracy of Value-at-Risk Forecasts: New Multilevel Tests, *Int J Forecasting*, Vol. 30(2), pp. 206–216, 014.doi: <https://doi.org/10.1016/j.ijforecast.2013.07>.
- LOPEZ, J., (1999). Methods for Evaluating Value-at-Risk Estimates, *FRBSF Economic Review*, Vol. 2, pp. 3–17.
- MAŁECKA, M., (2018). Exponential Autoregressive Conditional Duration Approach to Testing VaR, in: *ICoMS 2018: Proceedings of the 2018 International Conference*

- on *Mathematics and Statistics*, ACM, New York, pp. 6–10, doi: <https://doi.org/10.1145/3274250.3274254>.
- PAJHEDE, T., (2017). Backtesting Value-at-Risk: A Generalized Markov Test, *J Forecast*, Vol. 36(5), pp. 597–613, doi: <https://doi.org/10.1002/for.2456>.
- PELLETIER, D., WEI, W., (2016). The geometric-VaR backtesting method, *J Financ Economet*, Vol. 14(4), pp. 725–745, doi: <https://doi.org/10.1093/jjfinec/nbv015>.
- SELF, S. F., LIANG, K. Y., (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions, *J Am Stat Assoc*, Vol. 82(398), pp. 605–610, doi: <https://doi.org/10.2307/2289471>.
- WIED, D., WEI, G. N. F., ZIGGEL, D., (2016). Evaluating Value-at-Risk forecasts: a new set of multivariate backtests, *J Bank Financ*, Vol. 72, pp. 121–132, doi: <https://doi.org/10.1016/j.jbankfin.2016.07.014>.