Department of Economics
FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Thor Pajhede Nielsen

Tales From the Unit Interval: Backtesting, Forecasting and Modeling

Academic advisor: Anders Rahbek

Date of submission: 31 January, 2017

1 Acknowledgments

First and foremost I wish to thank my thesis supervisors, Anders Rahbek and Heino Bohn. Since I started my PhD studies they have been a constant source of support and encouragement in my research. I have had the pleasure of many interesting and illuminating discussions over the years, for that I am very grateful.

At the Department of Economics, University of Copenhagen, I would like to thank the faculty and my fellow PhD students for a friendly environment and insightful discussions over the years. A special thank also goes to Rasmus Søndergaard, Nick Fabrin, Anne Lundgaard, Philipp Kless, Morten Tabor and Emil Nejstgaard for making life as a PhD student a great experience and for their many comments on my work.

In addition to my colleagues at the Department of Economics, I would like to thank, Peter Christoffersen and Oliver Linton for valuable conversations related to my research.

I would also like to thank Dennis Kristensen and the staff of University College London, for helping me arrange my research stay at University College London (UCL) in the Spring of 2016. I enjoyed being a visitor at UCL, the faculty and doctoral students were kind and welcoming during my stay.

Last but not least, I would like to thank Lena Hasselberg, my family, and my friends for their support during my studies.

2 Summary

This thesis comprises three self-contained chapters, which are all related to either credit or market risk.

Chapter one discusses a new likelihood ratio test for evaluating Value-at-Risk (VaR) forecasts. We provide closed form expressions for the tests as well as asymptotic theory. Not only do the generalized tests have power against k'th order dependence by definition, but also included simulations indicate improved power performance over existing tests. The chapter is forthcoming in the Journal of Forecasting.

Chapter two discusses how to best forecast VaR of a portfolio. In particular one faces a number of choices in how to construct a model; univariate or multivariate models, interday or intraday based data and distributional alternatives. We consider a portfolio of 44 major US stocks from the S&P 500 index and compare forecasts using both recently developed backtests and the model confidence set approach. We also consider the square-root-of-time scaling rule for a 10 day period as suggested in the Basel accords.

Chapter three discusses an observation driven, conditionally beta distributed model. The model includes both explanatory variables and autoregressive dependence in the mean and precision parameters using the mean-precision parametrization of the beta distribution suggested by Ferrari et al. (2004). Our model is a generalization of the βARMA model proposed in Rocha et al. (2009). We also highlight some errors in their derivations of the score and information, which has implications for the asymptotic theory. Included simulations suggest that standard asymptotics for estimators and test statistics apply. In an empirical application to Moody's monthly US 12-month issuer default rates in the period 1972-2015, we revisit the results of Agosto et al. (2016) in examining the conditional independence hypothesis of Lando et al. (2010).

3 Summary in Danish

Denne afhandling indeholder tre selvtændige kapitler som alle er relaterede til enten kredit eller markeds risiko.

Kapitel et diskuttere nye likelihood ratio tests til evaluering af Value-at-Risk (VaR) forecasts. Vi giver lukkede form udtryk for testene samt asymptotisk teori. De nye tests har styrke mod k'te ordens afhængighed per definition, mens inkluderede simulationer antyder forbedret styrke i forhold til eksisterende tests. Kapitlet bliver udgivet i Journal of Forecasting.

Kapitel to diskutere hvordan man bedst kan forecaste VaR for en stor portefølje af aktier. Specifikt skal man træffe valg angående hvordan en model for risiko skal opbygges: Univariat eller multivariat, interday eller intraday baseret data og antagelser vedr. fordeling. I en empirisk applikation benytter vi en portefølge af 44 amerikanske aktier fra S&P 500 indekset hvor vi sammenligner en række metoders forecasting ved brug af både nyligt udviklede backtests samt model confidence set metoden. Vi diskutere også square-root-of-time skalerings reglen for en 10 dages periode som foreslås i Basel reglerne.

Kapitel tre diskutere en observations drevet, betinget beta fordelt model. Modellen inkludere både forklarende variable og autoregressiv afhængighed i middelværdien samt præcisions parametrene. Vores model er en generalisering af βARMA modellen foreslået i Rocha et al. (2009). Vi viser nogle fejl i deres udledninger af scoren og informationen som har implikationer for den asymptotiske teori. Inkluderede simulationer antyder at standard asymptotik gælder. I en empirisk applikation på Moody's amerikanske 12 måneders default rate i perioden 1972-2015 replikere vi resultaterne fra Agosto et al. (2016) hvor vi undersøger hypotesen om betinget uafhængighed fremsat i Lando et al. (2010).

Table of Contents

Backtesting Value-at-Risk: A Generalized Markov Test[*,†]

Thor Pajhede[‡]

January 31, 2017

Abstract

Testing the validity of Value-at-Risk (VaR) forecasts, or backtesting, is an integral part of modern market risk management and regulation. This is often done by applying independence and coverage tests developed in Christoffersen (1998) to so-called hit-sequences derived from VaR forecasts and realized losses. However, as pointed out in the literature, see Christoffersen and Pelletier (2004), these aforementioned tests suffer from low rejection frequencies, or (empirical) power when applied to hit-sequences derived from simulations matching empirical stylized characteristics of return data. One key observation of the studies is that higher order dependence in the hit-sequences may cause the observed lower power performance. We propose to generalize the backtest framework for Value-at-Risk forecasts, by extending the original first order dependence of Christoffersen (1998) to allow for a higher, or k'th, order dependence. We provide closed form expressions for the tests as well as asymptotic theory. Not only do the generalized tests have power against k'th order dependence by definition, but also included simulations indicate improved power performance when replicating the aforementioned studies. Further, included simulations show much improved size properties of one of the suggested tests.

Keywords: Value-at-Risk, Backtesting, Markov Chain, Duration, quantile, likelihood ratio, maximum likelihood.

JEL codes: C12, C15, C52, C32.

---

[*]**This chapter has been accepted for publication in Journal of Forecasting.**

[†]The author acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. We thank Peter Christoffersen, Anders Rahbek, Heino Bohn Nielsen, Lena Hasselberg, Oliver Linton and an anonymous referee as well as the participants of the 2016 SOFIE conference for helpful comments. The Matlab code used to implement the Backtests in this paper as well as several others is free to use, and will be made available from econ.ku.dk/pajhede/miscellaneous.html.

[‡]Department of Economics, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K, Denmark. CREATES, University of Aarhus. E-mail: thor.nielsen@econ.ku.dk

1 Introduction

Since its introduction in the 90s Value-at-Risk (VaR), as measured by the p'th quantile of a forecasted distribution of losses, has become widely used when reporting aggregate market risk. This again has prompted a rich literature on validation of VaR forecasts, so-called backtesting, as is often applied empirically by regulatory authorities, academics and financial institutions. See Campbell (2007) for a review of the backtesting procedures and an economic motivation for the backtesting criteria.

The leading reference on backtesting is Christoffersen (1998), wherein the evaluation of accurate VaR forecasts was first formalized. Specifically it was shown that the occurrences of losses beyond a specified VaR level, termed *violations* or *hits*, should occur independently and with a constant probability matching the p'th quantile. Based on this, the widely applied *conditional coverage* and *independence* tests were proposed. However, as documented in Christoffersen and Pelletier (2004) and Berkowitz et al. (2011) the tests have low empirical power in simulation studies matching empirical stylized facts of returns data.

To address this we propose to derive tests in a more general setting than the original framework of Christoffersen (1998). Specifically, we propose tests within a general backtest framework extending the underlying Markovian model of Christoffersen (1998) to allow for higher, or k'th, order dependence. Within the quite general k'th order dependence model, we consider two structures, or specifications: one which we label as the generalized Markov specification, and the other as the generalized Markov duration specification. Preceding the details given in Section 2.2, the generalized Markov specification can be viewed as similar to the extension of autoregressive models from order one to order k when testing for white noise, while the Markov duration specification mimic duration modeling approaches to backtesting of Christoffersen and Pelletier (2004), Haas (2006) and Pelletier and Wei (2015).

We provide asymptotic theory and closed form expressions for the implied tests for conditional coverage and independence within these generalized specifications. Moreover, simulations illustrate that the new generalized tests solve some of the leading issues with regards to low empirical power.

Note in this respect, that by definition the proposed tests will have power against higher order dependence, and in particular so when compared to the tests derived in the Markovian framework. That the tests seem to perform well in empirically stylized simulations is additional reason to prefer these.

The rest of the paper is organized as follows. Section 2 sets out the backtesting criteria i.e. Unconditional Coverage, Independence, and Conditional Coverage. Subsection 2.1 reviews the popular classic Markov backtests due to Christoffersen (1998) and Kupiec (1995). Subsection 2.2 introduces our new framework. We consider two specifications from this framework, the generalized Markov and the Markov duration specifications. From them we derive tests of unconditional coverage, independence and conditional coverage. Section 3 examines the power and size properties of the various tests using a simulation framework. Section 4 concludes.

2 Hit-sequence Based Backtesting

Let R_t denote the realization of a return of an asset or a portfolio of assets at time t. The ex ante VaR for time t and coverage rate p, denoted as $\text{VaR}_{t|t-1}(p)$, conditional on all information, $\mathcal{F}_{t-1}$, available at time $t-1$ (for example past returns and macroeconomic indicators) is defined as the p'th conditional quantile of the distribution of R_t:

$$P(R_t < \text{VaR}_{t|t-1}(p)|\mathcal{F}_{t-1}) = p, \qquad t = 1, ..., T.$$

Typically the coverage rate used is 1% or 5%. Several parametric (for example GARCH models) and non-parametric (for example Historical Simulation) methods are used to forecast $\text{VaR}_{t|t-1}(p)$, see McNeil et al. (2005).

Backtesting is the procedure of comparing realized losses to the forecasted VaR. To implement backtesting of a VaR forecast, we follow Christoffersen (1998) in defining the hit-sequence, $\{I_t\}_{t=1}^{T}$, as follows:

Definition 1. The hit-sequence, $\{I_t\}_{t=1}^{T}$, for a sequence of VaR forecast, $\left\{\text{VaR}_{t|t-1}(p)\right\}_{t=1}^{T}$, is defined as,

$$I_t := 1\left(R_t < \text{VaR}_{t|t-1}(p)\right), \qquad t = 1, ...T \tag{2.1}$$

Where $1(\cdot)$ is the indicator function. Thus, the hit-sequence is by construction a binary time series indicating whether a loss at time t greater than the VaR, termed a violation or a hit, was realized.

A VaR forecast is valid, in the sense of actually having forecasted the desired quantile, only if the associated hit-sequence satisfies the following criteria due to Christoffersen (1998):

- **The unconditional coverage criteria:** The unconditional probability of a violation must be exactly equal to the coverage rate p:

$$H_{UC}: \ P(I_t = 1) = p$$

- **The independence criteria**: The conditional probability of a violation must be constant:

$$H_{Ind}: \ P(I_t = 1|\mathcal{F}_{t-1}) = P(I_t = 1)$$

Combining these criteria we obtain the conditional coverage criteria:

- **The conditional coverage criteria**: The probability of a violation must be constant and equal to the coverage rate:

$$H_{CC}: \ P(I_t = 1|\mathcal{F}_{t-1}) = P(I_t = 1) = p$$

It follows, see Christoffersen (1998), that the hit-sequence of a valid VaR forecast, is in fact a sequence of i.i.d. Bernoulli distributed variables:

$$I_t \underset{i.i.d.}{\sim} Bernoulli(p), \qquad t = 1, ..., T. \tag{2.2}$$

The classic Markov framework of Christoffersen (1998) models the hit-sequence of (2.2) as a first order Markov chain. As detailed in the following subsection 2.1 this allows testing of both the unconditional coverage and independence criteria using likelihood-ratio tests. Furthermore, these tests have closed form expressions, standard asymptotics and are easy to implement. However, as previously mentioned in the introduction, the tests have also been found to suffer from low power when dependence is not Markovian of order one.

In subsection 2.2, we extend the classic Markov framework to allow for higher, or k'th order, dependence. We detail how our approach preserves all of the aforementioned advantages of the classic Markov testing, but also have power against more general forms of dependence.

2.1 Classic Markov Testing

The first backtest by Kupiec (1995), models the hit-sequence as an i.i.d. Bernoulli sequence with an unknown probability parameter $\pi_1 \in]0,1[$, that is:

$$I_t \underset{i.i.d.}{\sim} Bernoulli(\pi_1), \qquad t = 1, ..., T \tag{2.3}$$

The likelihood for the Bernoulli sequence (2.3) is given by $\mathcal{L}_T(\pi_1) = \pi_1^{T_1}(1-\pi_1)^{T_0}$ where $T_1 = \sum_{t=1}^{T} I_t$, $T_0 = T - T_1$ and the maximum likelihood (ML) estimate of π_1 is given by $\hat{\pi}_1 = T_1/T$.

From this a likelihood-ratio test of the restriction $H_{UC}: \ \pi_1 = p$, corresponding to the criteria of unconditional coverage, can be constructed in the usual way. It follows that the likelihood-ratio statistic, under the hypothesis stated in the parenthesis, for unconditional coverage satisfies, as $T \to \infty$,

$$Q_{UC}(\pi_1 = p) = -2log\left(\frac{p^{T_1}(1-p)^{T_0}}{\hat{\pi}_1^{T_1}(1-\hat{\pi}_1)^{T_0}}\right) \xrightarrow{d} \chi^2(1). \tag{2.4}$$

This test is often termed the proportion of failures (PF) test. Because the model from which the test was derived, see equation (2.3), does not allow for any dependence structure in the hit-sequence it is clear that the test is unsuited to detect dependence in the hit-sequence.

The need to also test the independence criteria led Christoffersen (1998) to develop the Markov tests of independence and conditional coverage. To do so it was proposed to model the conditional distribution of I_t given I_{t-1}, $I_t|I_{t-1}$ as a first order Markov chain. We write this first order Markov chain as

$$I_t|I_{t-1} \underset{i.i.d}{\sim} Bernoulli(p_t(\theta)),$$

with transition probability,

$$p_t(\theta) = I_{t-1}\pi_{11} + (1 - I_{t-1})\pi_{01}, \quad \theta = (\pi_{11}, \pi_{01})' \in]0,1[^2.$$

Here π_{ij} is the probability of observing i on day $t-1$ being followed by observing j on day t for $i, j = 0, 1$. Equivalently this may be expressed in terms of the transition probability matrix given by

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}. \tag{2.5}$$

In terms of Π, independence is implied by the restriction $H_{Ind}: \ \pi_{01} = \pi_{11}$ while the combined hypothesis of conditional coverage can be tested by the additional restriction $H_{CC}: \ \pi_{01} = \pi_{11} = p$.

The likelihood for the unrestricted Markov chain $\{I_t\}_{t=1}^T$, with the first observation (I_0) fixed, is given by

$$\mathcal{L}_T(\pi_{01}, \pi_{11}) = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}.$$

Here T_{ij} indicates the number of observations of the hit-sequence where a j follows an i. Noting that $T_1 = T_{11} + T_{10}$ and $T_0 = T_{01} + T_{00}$, the ML estimates are $\hat{\pi}_{01} = T_{01}/T_0$, $\hat{\pi}_{11} = T_{11}/T_1$. It follows that the likelihood-ratio test statistic of independence, as $T \to \infty$, satisfies,

$$Q_{Ind}(\pi_{01} = \pi_{11}) = -2log\left(\frac{(1-\hat{\pi}_1)^{T_{00}}\hat{\pi}_1^{T_{01}}(1-\hat{\pi}_1)^{T_{10}}\hat{\pi}_1^{T_{11}}}{(1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}}\right) \xrightarrow{d} \chi^2(1).$$

Likewise, the likelihood-ratio test statistic for conditional coverage (the so-called joint test), satisfies,

$$Q_{CC}(\pi_{01} = \pi_{11} = p) = -2log\left(\frac{(1-p)^{T_{00}}p^{T_{01}}(1-p)^{T_{10}}p^{T_{11}}}{(1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}}\right) \xrightarrow{d} \chi^2(2).$$

The tests of Christoffersen (1998) have standard asymptotics, closed form expressions and remain popular in the applied literature. However, because they only model the hit-sequence as a first order Markov chain the ability to detect higher order dependence may be limited. Furthermore, simulation studies have shown them to have a low power in realistic settings. In the following subsection 2.2 we extend the model to allow for higher order dependence, in order to remedy the shortcomings of the classical framework but still derive tests that are easy to implement and interpret.

2.2 Generalized Markov framework

As detailed in the first part of the section, we now extend the classic Markov framework to a k'th order Markov chain. Specifically, let the Markov chain be given by,

$$I_t|\mathcal{F}_{t-1,k} \underset{i.i.d}{\sim} Bernoulli(p_t(\theta)), \quad \mathcal{F}_{t-1,k} = I_{t-1}, ..., I_{t-k}, \qquad t = 1, ..., T. \tag{2.6}$$

The transition probabilities of (2.6) are given by,

$$p_t(\theta) = P(I_t = 1|\mathcal{F}_{t-1,k}), \qquad t = 1, ..., T, \tag{2.7}$$

With θ a 2^k vector of the individual parameters, corresponding to the possible permutations of $I_{t-1}, I_{t-2}, ..., I_{t-k}$.

Equivalently, one could specify a k-tuple $\tilde{I}_t = (I_t, ..., I_{t-k+1})'$ which would then follow a Markov chain governed by a $2^k \times 2^k$ transition matrix P. Since the rows of P must sum to 1 and each state is only accessible from 2 other states, this implies that each row has two non-zero elements[1], which restricts it to the 2^k parameters also found in θ.

The likelihood for this model conditioned on k observations prior to $t = 1$ fixed, is given by,

$$\mathcal{L}_T(\theta) = \prod_{t=1}^{T} p_t(\theta)^{I_t}(1 - p_t(\theta))^{1-I_t},$$

and the log-likelihood by, $L_T(\theta) = \sum_{t=1}^{T} log(p_t(\theta))I_t + log(1 - p_t(\theta))(1 - I_t)$.

The principal motivation was to allow for dependence of order $k > 1$. However since the number of parameters increase at the geometric rate of 2^k, estimating the model becomes infeasible for larger values of k. In order to have a feasible number of parameters we therefore impose parametric structures on the model of equation (2.7). Examples of such structures or restrictions, inspired by the tests of Christoffersen (1998) Christoffersen and Pelletier (2004) and Haas (2006), are presented in the following subsections 2.2.1 and 2.2.2. The criteria of independence and conditional coverage impose further restrictions, which are used to create likelihood-ratio tests. Specifically if the restriction $p_t(\theta) = p$ holds for all t, then the Markov chain of equation (2.6) reduces to the i.i.d. Bernoulli sequence of equation (2.2).

There is no clear choice of k. A too low value might not adequately allow for the modeling of higher order dependence. While a too high k conditions on too many observations making the effective sample size small. For $k = 1$ the tests suggested in the following subsections reduce to the tests of Christoffersen (1998) described in section 2. A natural choice of k is to use 5, 10 or 20, corresponding to testing for a change in the probability of a hit in the week, two weeks or 1 month following a hit.

As pointed out by an anonymous referee the proposed framework could also be used in a dynamic quantile type regression estimated by ordinary least squares, see Engle and Manganelli (2004), by use of appropriate indicator functions. In fact, the proofs of asymptotic distribution for the DQ test remain valid in such a case.

[1] Intuitively, if $k = 1$, one can recall that the two permutations of I_{t-1} (either 1 or 0) meant that the classical tests of Christoffersen (1998) are based on a Markov chain with 2 parameters in θ which are gathered into a 2×2 transition matrix.

However, in our framework we will make use of LR test statistics which are known to have good power properties and for these models yield somewhat simpler test statistics which may be beneficial to some practitioners.

2.2.1 The Generalized Markov Test

In terms of the unrestricted model in (2.6), we first consider the restriction that the probability of a hit at time t, $p_t(\theta)$, is a function of only whether or not a hit has occurred in $I_{t-1}, I_{t-2}, ..., I_{t-k}$. This reduces the parameters of the model to two, or equivalently,

$$p_t(\theta) = J_{t-1}p_E + (1 - J_{t-1})p_S, \quad J_{t-1} := 1\left(\sum_{i=1}^{k} I_{t-1} > 0\right). \tag{2.8}$$

The bivariate parameter vector $\theta = (p_E,\ p_S)'$ belongs to the parameter space $\Theta =]0, 1[^2$. Intuitively, this corresponds to an excited (p_E) and a steady (p_S) probability. Because the restricted model retains the interpretation of two categories similar to the Markov tests of Christoffersen (1998), we will refer to it as the the generalized Markov specification.

The likelihood is then given by,

$$\mathcal{L}_T(\theta) = (1 - p_S)^{T_{00}} p_S^{T_{01}} (1 - p_E)^{T_{10}} p_E^{T_{11}},$$

where T_{ij} are the counts; $T_{11} := \sum_{t=1}^{T} I_t J_{t-1}$, $T_{01} := \sum_{t=1}^{T} I_t(1 - J_{t-1})$, $T_{10} := \sum_{t=1}^{T}(1 - I_t)J_{t-1}$, $T_{00} := \sum_{t=1}^{T}(1 - I_t)(1 - J_{t-1})$. That is, T_{11} (T_{10}) is the number of hits (no hits) observed where one or more hits were observed in the preceding k observations. $T_{01}(T_{00})$ is the number of hits (no hits) observed where there was not observed a hit in the prior k observations.

This leads to the ML estimates (see Appendix A),

$$\hat{p}_S = \frac{T_{01}}{T_{01} + T_{00}} \qquad \text{and} \qquad \hat{p}_E = \frac{T_{11}}{T_{11} + T_{10}}.$$

It follows that we require $T_{11} + T_{10} > 0$ corresponding to at least 1 hit in the hit-sequence which must occur before T, to calculate the test. This is the same requirement as for the tests of Christoffersen (1998).

To test the hypothesis of independence, we consider the restriction $H_{Ind}:\ p_E = p_S := \phi$, that is, whether there is a constant probability of a hit. The restricted parameter space, Θ_H, is in this case given by,

$$\Theta_H = \{\theta |\ \theta = H\phi,\ \phi \in]0, 1[\},$$

where $H = (1, 1)'$, with ML estimate of ϕ given by (see Appendix A)

$$\hat{\phi} = \frac{T_{01} + T_{11}}{T_{01} + T_{11} + T_{00} + T_{10}} = \frac{T_1}{T}. \tag{2.9}$$

Defining the unrestricted estimator, the estimator restricted under H_{Ind} and the estimator restricted under

H_{CC} as

$$\hat{\theta} := \underset{\theta \in \Theta}{argmax}\ \mathcal{L}_T(\theta) = (\hat{p}_S, \hat{p}_E)', \quad \tilde{\theta} := \underset{\theta \in \Theta_H}{argmax}\ \mathcal{L}_T(\theta) = H\hat{\phi} \quad \text{and} \quad \theta_0 = Hp.$$

As was the case for the classic Markov tests of the previous section, the likelihood ratio test statistic of independence conveniently factorizes, see Appendix A, into tests for conditional coverage and unconditional coverage as follows (with the hypothesis of each test in parenthesis)

$$\begin{aligned} Q_{G-Ind}(\theta = H\phi) &= -2log\left(\frac{\mathcal{L}_T(\tilde{\theta})}{\mathcal{L}_T(\hat{\theta})}\right) = Q_{G-CC}(\theta = Hp) - Q_{G-UC}(H\phi = Hp) \\ &= \left(-2\left[L_T(\theta_0) - L_T(\hat{\theta})\right]\right) - \left(-2\left[L_T(\theta_0) - L_T(\tilde{\theta})\right]\right) \end{aligned}$$

Note the simple relation, $Q_{G-CC}(\theta = Hp) = Q_{G-Ind}(\theta = H\phi) + Q_{G-UC}(H\phi = Hp)$. This provides a simple way of analyzing a rejection of CC. If a rejection of conditional coverage is found one can examine if it was due to dependence, an incorrect coverage or both, using the $Q_{G-Ind}(\theta = H\phi)$ and $Q_{G-UC}(H\phi = Hp)$ tests.

The test statistic of independence has the following expression

$$\begin{aligned} Q_{G-Ind}(\theta = H\phi) &= -2log\left(\frac{\mathcal{L}_T(\tilde{\theta})}{\mathcal{L}_T(\hat{\theta})}\right) \\ &= -2\{log(1-\hat{\phi})(T_{00}+T_{10}) + log(\hat{\phi})(T_{01}+T_{11}) \\ &- log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} - log(1-\hat{p}_E)T_{10} - log(\hat{p}_E)T_{11}\} \end{aligned} \tag{2.10}$$

The test statistic of conditional coverage has the following expression

$$\begin{aligned} Q_{G-CC}(\theta = Hp) &= -2log\left(\frac{\mathcal{L}_T(\theta_0)}{\mathcal{L}_T(\hat{\theta})}\right) \\ &= -2\{log(1-p)(T_{00}+T_{10}) + log(p)(T_{01}+T_{11}) \\ &- log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} - log(1-\hat{p}_E)T_{10} - log(\hat{p}_E)T_{11}\} \end{aligned} \tag{2.11}$$

The test statistic for unconditional coverage, $Q_{G-UC}(H\phi = Hp)$ is by definition simply the proportion of failures test of section 2, where the first k observations are dropped from the sample.

The distribution of the generalized Markov tests of for independence, conditional coverage and unconditional coverage are asymptotically $\chi^2(1)$, $\chi^2(2)$ and $\chi^2(1)$. That is, We have the following results:

Theorem 1. *For $T \to \infty$, and under the null-hypothesis that $\{I_t\}$ is an i.i.d. Bernoulli sequence with probability parameter p,*

$$Q_{G-Ind}(\theta = H\phi) \xrightarrow{d} \chi^2(1),$$

$$Q_{G-CC}(\theta = Hp) \xrightarrow{d} \chi^2(2),$$

$$Q_{G-UC}(H\phi = Hp) \xrightarrow{d} \chi^2(1).$$

For a proof see Appendix A.

2.2.2 The Markov Duration Test

In terms of the unrestricted model in 2.6, we now consider the restriction that the probability of a hit at time t, $p_t(\theta)$, is a function of the number of observations since the last hit (the *duration*) in the preceding k lags, after which the probability is a constant. This reduces the parameters of the model to $k+1$, or equivalently,

$$p_t(\theta) = J(1)_{t-1}p_{E1} + ... + J(k)_{t-1}p_{Ek} + (1 - \sum_{i=1}^{k} J(i)_{t-1})p_S, \tag{2.12}$$

where

$$J(1)_{t-1} := 1\left(I_{t-1} = 1\right)\ , ...,\ J(k)_{t-1} := 1\left(I_{t-1} = 0, ..., I_{t-k} = 1\right).$$

Specifically this implies $p_{E1} = P(I_t = 1|I_{t-1} = 1)$, $p_{Ek} = P(I_t = 1|I_{t-1} = 0, ..., I_{t-k} = 1)$ and $p_S = P(I_t = 1|I_{t-1} = 0, ..., I_{t-k} = 0)$. Because the restricted model is similar to the underlying models of the duration based backtests of Christoffersen and Pelletier (2004), Haas (2006) and Pelletier and Wei (2015) we will refer to this as the Markov duration specification.

The parameter vector $\theta = (p_{E1}, ..., p_{Ek},\ p_S)'$ belongs to the parameter space $\Theta =]0,1[^{k+1}$. The Markov duration specification is less restrictive than that of the generalized Markov specification and contains it as the special case $p_{E1} = ... = p_{Ek}$. Despite being less restrictive, the specification ensures that the number of parameters in (2.6) only grows linearly with k.

The likelihood is given by,

$$\mathcal{L}_T(\theta) = (1-p_S)^{T_{00}}p_S^{T_{01}}\prod_{i=1}^{k}(1-p_{Ei})^{T_{10}(i)}p_{Ei}^{T_{11}(i)},$$

where $T_{10}(i) = \sum_{t=i+1}^{T}(1-I_t)J(i)_{t-1}$ is the number of zeros observed after having observed a hit in I_{t-i}, but not in any I_{t-j} where $i > j$. $T_{11}(i)$ is the number of ones observed after having observed a hit I_{t-i} lags previously, but not in any I_{t-j} where $i > j$.

This leads to the ML estimates (see Appendix B),

$$\hat{p}_S = \frac{T_{01}}{T_{01}+T_{00}} \qquad \text{and} \qquad \hat{p}_{Ei} = \frac{T_{11}(i)}{T_{11}(i)+T_{10}(i)}, \quad i = 1, ..., k,$$

It follows that we require $T_{11}(i) + T_{10}(i) > 0$, corresponding to at least 1 hit not in the last k observations, to calculate the test. T_{01} and T_{00} are identical to those defined in the previous section.

To test the hypothesis of independence, consider the restriction $H_{Ind}:\ p_{E1} = ... = p_{Ek} = p_S := \phi$, that is, whether there is a constant probability of a hit. The restricted parameter space is given by

$$\Theta_H = \left\{\theta|\ \theta = H\phi,\ \phi \in]0,1[\right\},$$

Where $H = (1, ..., 1)'$ is a $k \times 1$ vector and with ML estimate $\hat{\phi}$ unchanged.

Defining the unrestricted estimator, the estimator restricted under H_{Ind} and the estimator restricted under H_{CC} as

$$\hat{\theta} := \underset{\theta\in\Theta}{argmax}\ \mathcal{L}_T(\theta) = (\hat{p}_S, \hat{p}_{E1}, ..., \hat{p}_{Ek})', \quad \tilde{\theta} := \underset{\theta\in\Theta_H}{argmax}\ \mathcal{L}_T(\theta) = H\hat{\phi} \quad \text{and} \quad \theta_0 = Hp.$$

The likelihood ratio test statistic factorizes as into tests of conditional coverage and unconditional coverage as in the previous subsection (with the hypothesis of each test in parenthesis)

$$\begin{aligned} Q_{D-Ind}(\theta = H\phi) &= -2log\left(\frac{\mathcal{L}_T(\tilde{\theta})}{\mathcal{L}_T(\hat{\theta})}\right) = Q_{D-CC}(\theta = Hp) - Q_{D-UC}(H\phi = Hp) \\ &= \left(-2\left[L_T(\theta_0) - L_T(\hat{\theta})\right]\right) - \left(-2\left[L_T(\theta_0) - L_T(\tilde{\theta})\right]\right), \end{aligned}$$

We will refer to these tests as the Markov-Duration tests of independence, unconditional coverage and conditional coverage. We again have the relation between the tests that $Q_{D-CC}(\theta = Hp) = Q_{D-Ind}(\theta = H\phi) + Q_{D-UC}(H\phi = Hp)$.

Intuitively, $Q_{D-Ind}(\theta = H\phi)$ tests whether the hazard function can be reduced to a constant and $Q_{D-CC}(\theta = Hp)$ tests if that constant is exactly p. They can be viewed as duration tests, with the hazard rate being entirely free of restrictions except a truncation to a constant beyond the k'th lag.

The test statistic of independence has the following expression

$$\begin{aligned} Q_{D-Ind}(\theta = H\phi) &= -2log\left(\frac{\mathcal{L}_T(\tilde{\theta})}{\mathcal{L}_T(\hat{\theta})}\right) \\ &= -2\Big(log(1-\hat{\phi})(T_{00}+T_{10}) \times log(\hat{\phi})(T_{01}+T_{11}) - log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} \\ &- \sum_{i=1}^{k} log(1-\hat{p}_{Ei})T_{10}(i) - \sum_{i=1}^{k} log(\hat{p}_{Ei})T_{11}(i)\Big). \end{aligned} \tag{2.13}$$

The test statistic of conditional coverage has the following expression

$$\begin{aligned} Q_{D-CC}(\theta = Hp) &= -2log\left(\frac{\mathcal{L}_T(\theta_0)}{\mathcal{L}_T(\hat{\theta})}\right) \\ &= -2(log(1-p)(T_{00}+T_{10}) \times log(p)(T_{01}+T_{11}) - log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} \\ &- \sum_{i=1}^{k} log(1-\hat{p}_{Ei})T_{10}(i) - \sum_{i=1}^{k} log(\hat{p}_{Ei})T_{01}(i)). \end{aligned} \tag{2.14}$$

Lastly, $Q_{D-UC}(H\phi = Hp)$ is simply the proportion of failures test of section 2, where the first k observations are dropped from the calculations (equivalent to the $Q_{G-UC}(H\phi = Hp)$ test statistic).

The distribution of the Markov duration tests of for independence, conditional coverage and unconditional coverage are asymptotically χ^2 distributed. That is, we have the following results:

Theorem 2. *For $T \to \infty$, and under the null-hypothesis that $\{I_t\}$ is an i.i.d. Bernoulli sequence with probability parameter p,*

$$Q_{D-Ind}(\theta = H\phi) \xrightarrow{d} \chi^2(k),$$

$$Q_{D-CC}(\theta = Hp) \xrightarrow{d} \chi^2(k+1),$$

$$Q_{D-UC}(H\phi = Hp) \xrightarrow{d} \chi^2(1).$$

For a proof see Appendix B.

In section 3.1 we demonstrate in a simulation study that using the asymptotic distributions of Theorems 1 and 2 to calculate p-values can cause a distortion of the size. Instead the Monte Carlo Method of Dufour (2006) can be used to simulate the exact distribution under the null hypothesis and obtain valid p-values. It is the tests using the Monte Carlo Method of Dufour (2006) which should be used in practice and it is what is used in our empirical power simulations found in sections 3.2 and 3.3.

3 Simulation Study of Size and Power

In this section we conduct a simulation study to investigate the empirical size and power properties of the generalized Markov and duration tests of conditional coverage developed in section 2.2. Further, we evaluate the *empirical rejection frequency* (ERF) of the tests using a simulation setup not contained in the general model of equation (2.6), generating the returns using a GARCH model and forecasting the VaR using *historical simulation* (HS). This later simulation is commonly included in papers which develop VaR backtests and we refer to it as scenario power.

We use $k = 1$, 5, and 10 lags for each of the conditional coverage tests, see equations (2.11) and (2.14) from section 2.2, where we note that for $k = 1$, the generalized Markov and generalized duration tests both reduce to the original joint test of Christoffersen (1998). We use sample sizes $T = 500$, $1,000$, $1,500$, $2,500$, $5,000$ and $N = 100,000$ replications for each sample size. For the size simulations we use $p = 1\%$, 5% and 10%, where the latter is included to illustrate the improved size properties for larger values of p. For the power simulations we use only $p = 1\%$ and 5% reflecting empirically relevant cases. We use a significance level of 5% for all simulations. In the empirical power and scenario power simulations in subsections 3.2 and 3.3 we use the Monte Carlo testing technique of Dufour (2006) (see Appendix C) to obtain tests with a size of 5%. Tests which could not be calculated are treated as not rejecting a well specified forecast since they indicate neither dependence in, nor an excess number of, violations, the primary concern of backtesting. Typically the feasibility ratio of backtests is very close to 1 when considering 500 or more observations even when using a low coverage rate, Table 6 of Berkowitz et al. (2011) presents feasibility ratios for a variety of backtests.

To facilitate comparison with newer backtests we also include the discrete duration backtest for conditional coverage of Haas (2006), the GMM j-tests of Candelon et al. (2011) with either 3 or 5 moment conditions, the Ljung-Box test with 5 or 10 lags and lastly the DQ test of Engle and Manganelli (2004).

3.1 Empirical Size

It is a well established fact of the backtest literature that the use of asymptotic distributions critical values can create significant size distortions in existing tests, see Christoffersen and Pelletier (2004). To examine the size distortion of the tests developed in this paper, and to examine when the asymptotic critical values can be used, we simulate the hit-sequence, $\{I_t\}_{t=1}^{T}$, under the null hypothesis of Conditional Coverage as an independent Bernoulli sequence. Recalling equation (2.2), we simulate the hit-sequence, $\{I_t\}_{t=1}^{T}$, using the data generating process (DGP):

$$I_t \underset{i.i.d}{\sim} Bernoulli(p), \qquad t = 1, ..., T$$

ERFs of the generalized Markov and generalized duration tests of conditional coverage, when using the asymptotic distributions critical value are presented in Table 1. From the table it is clear that using the critical values of the asymptotic distributions can cause size distortion, especially when testing a low p or when using a small sample. In general most tests, with the exception of the LB and DQ tests, appear to be undersized when using the low $p = 1\%$. Though the generalized Markov test, of equation (2.11), is only slightly undersized for $k = 5$ and $T > 1,000$. When the higher $p = 5\%$ or 10%, is used, the size properties are generally much improved for all tests. Especially so for $k > 5$. The generalized duration test, of equation (2.14), has somewhat varying size properties. For the low $p = 1\%$ it is undersized while for $p = 5\%$ or 10% it is oversized, though not to high degree when $T = 5,000$ observations are used. Out of the tests considered, the generalized Markov test with $k = 10$ lags used appears to have the best size properties.

Because of the size distortion the empirical power and scenario power simulations in subsections 3.2 and 3.3 use the Monte Carlo testing technique of Dufour (2006) (see Appendix C) to obtain tests with a size of 5%.

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	0.99	2.37	2.74	0.49	0.18	10.84	1.71	1.17	11.97	16.13	18.76	18.41
1000	2.68	3.11	3.86	0.81	0.23	8.83	3.85	3.56	10.45	13.85	10.65	12.43
1500	3.33	3.79	4.23	0.99	0.39	7.18	3.67	3.11	10.26	11.50	9.54	12.64
2500	2.80	4.04	4.85	1.17	0.62	6.33	4.21	3.37	9.19	11.05	11.74	9.90
5000	3.39	5.27	5.30	2.12	1.70	5.69	4.83	4.13	7.75	8.83	8.53	9.01

(a) $p = 1\%$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	4.21	5.13	5.25	4.11	4.03	6.33	4.21	3.64	5.48	5.61	5.72	6.20
1000	6.02	5.12	5.01	6.32	7.06	6.14	4.96	4.29	5.05	5.08	4.96	5.08
1500	6.47	5.16	4.82	7.08	8.05	5.30	4.33	3.49	4.94	5.29	5.27	5.29
2500	5.64	4.92	4.80	6.29	6.62	5.03	4.78	4.15	5.08	5.19	5.10	5.71
5000	4.91	5.16	4.94	5.54	5.67	5.27	4.80	4.01	4.83	5.06	5.02	5.08

(b) $p = 5\%$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	4.92	4.81	4.72	6.43	7.62	5.24	4.50	3.57	4.71	4.99	4.89	5.16
1000	5.61	5.19	5.17	5.55	6.85	5.18	4.96	4.05	4.91	4.92	4.99	4.93
1500	5.31	5.19	5.47	5.46	5.58	5.46	5.20	4.05	4.74	4.64	4.54	4.62
2500	5.04	5.24	5.10	5.19	5.44	5.22	5.08	3.92	5.12	5.15	4.88	5.10
5000	4.77	4.91	5.01	4.79	4.89	4.75	5.38	4.19	4.83	4.56	4.63	4.56

(c) $p = 10\%$

Table 1: ERF when simulating under the null hypothesis of Conditional Coverage (the empirical size) and using the asymptotic distributions 95% critical value. The hit-sequences were drawn as i.i.d. *Bernoulli*(p) sequences. The results reported are based on $100,000$ replications for each test and sample size. The test Markov and Duration tests refer to the generalized Markov and generalized Duration tests developed in this paper, the Markov-1 test is also found in Christoffersen (1998) as the joint test. The Haas test is the test of Haas (2006), the GMM test is the test of Candelon et al. (2011), the LB test is the Ljung-Box test and the DQ test is the test of Engle and Manganelli (2004).

3.2 Empirical power

To evaluate the power, the probability of rejecting $\theta \in \Theta_0$ when $\theta \notin \Theta_0$, of the tests of conditional coverage we specify two *DGP's* using the generalized Markov specification and the generalized duration specification. Let the Markov chain generating the hit-sequence, I_t, be given by

$$I_t|\mathcal{F}_{t-1,k} \underset{i.i.d}{\sim} Bernoulli(p_t(\theta)), \quad \mathcal{F}_{t-1,k} = I_{t-1}, ..., I_{t-k}, \qquad t = 1, ..., T. \tag{3.1}$$

with transition probabilities of (3.1) given by one of two DGP's: Equation (2.8) or Equation (2.12), which we will refer to as specification 1 and 2. For specification 1 we set $p_S = 5\%$ and $p_E = 10\%$ with $k = 5$, this corresponds to the hit-sequence of a VaR forecast of coverage rate 5% which is misspecified in such a way that for $k = 5$ days following a hit the actual quantile modeled is the 10% quantile. We repeat these simulations using $k = 10$. For specification 2 we set $p_S = 5\%$, $P_{E1} = P_{E2} = 10\%$, $P_{E3} = P_{E4} = 8\%$ and $P_{E5} = 6\%$ for $k = 5$. , this corresponds to the hit-sequence of a VaR forecast of coverage rate 5% which is misspecified in such a way that for 5 days following a hit the actual quantile modeled is 10%, decreasing to 8% after 3 days, to 6% after 4 days and the returns to 5% after $k = 5$ days. We repeat this simulation for $k = 10$ we set $p_s = 5\%$, $P_{E1} = P_{E2} = P_{E3} = 10\%$, $P_{E4} = P_{E5} = P_{E6} = 9\%$, $P_{E7} = P_{E8} = 8\%$ and $P_{E9} = P_{E10} = 7\%$. The resulting empirical power[2] of the backtests are presented in Table 2 for the four DGP's shown. Several simulations not shown were carried out with similar results. For all simulations we use the Monte Carlo testing technique of Dufour (2006) rather than the critical values implied by the asymptotic distributions in evaluating the tests.

From Table 2 it can be seen that for less than 1,000 observations and especially for the Markov-1 test, the attained power can be quite limited. The DGP with the lowest order of dependence, $k = 5$, also results in tests which has a lower empirical power compared to the DGP with $k = 10$ order dependence. Further, we see that using the tests with the highest empirical power can greatly improve the empirical power compared to the joint test of Christoffersen (1998). Not surprisingly the DQ test of Engle and Manganelli (2004) performs respectably, but worse than the Markov tests using the correct number of lags.

[2]Strictly speaking, the term empirical power is only appropriate for those tests based on models which contain the DGP as a special case, eg. for $k = 5$ the simulations indicate the empirical power of the Markov-5, Duration-5 and Duration-10 tests.

3.3 Scenario Power Using GARCH Returns and Historical Simulation

The scenario power simulation consists of two elements, a model with parameters matching those found in empirical studies for generating non-i.i.d. returns and a forecast method which does not produce a valid forecast. Similar to Christoffersen and Pelletier (2004), Haas (2006), Berkowitz et al. (2011) and Candelon et al. (2011), we thus simulate a series of returns from a GARCH model and estimate VaR using the popular *HS* method[3]. Specifically, let the returns, R_t, be generated by a $GARCH(1,1) - t(d)$ with a skew and a conditional t distribution as:

$$R_t = \sigma_t z_t \sqrt{\frac{d-2}{d}}, \tag{3.2}$$

where the conditional variance is given by,

$$\sigma_t^2 = \omega + \alpha\sigma_{t-1}^2 \left(\sqrt{\frac{d-2}{d}} z_{t-1} - \theta \right)^2 + \beta\sigma_{t-1}^2. \tag{3.3}$$

Here z_t is an i.i.d. draw from a student t-distribution with d degrees of freedom. The parameter values are similar to estimates of this GARCH model on daily S&P500 returns, see Christoffersen and Pelletier (2004). Specifically, we set $d = 8$ degrees of freedom with parametrization of the coefficients as $\alpha = 0.1$, $\theta = 0.5$, $\beta = 0.85$ and $\omega = 3.9683e^{-6}$. The value of ω was set to target an annual standard deviation of 0.20 and the parametrization implies a daily volatility persistence of 0.975. We use a burn-in period of $5,000$ observations for each simulation to remove traces from initialization of the process. For more details see Christoffersen and Pelletier (2004) which presents figures of the generated returns, estimated VaR using HS and hazard functions of the hit-sequence from a similar simulation experiment.

Forecasting $\text{VaR}_{t|t-1}(p)$ is done using HS, see equation (3.4). HS is known to be under-responsive to changes in conditional risk as it assigns an equal probability weight to all past observations, ignoring the temporal ordering. Furthermore, the method responds asymmetrically, increasing risk (as measured by VaR) following large losses but not following large gains. HS generates a hit-sequence which violates conditional coverage, see Pritsker (2006) for a discussion of the problems associated with HS.

The forecast is found by taking the empirical p percentile, of a rolling window of the T_W latest returns. We set T_W to be either 250 or 500, both lengths are used so that we may evaluate the robustness of the results with respect to changes in the data generating process.

$$\text{VaR}_{t|t-1}(p) = percentile\left(\{R_j\}_{j=t-T_W}^{t-1}, p \right), \qquad t = 1, ..., T \tag{3.4}$$

Because the forecast is slow to update to changes in volatility, this will generate clusters of violations. We then use the returns and VaR forecasts to create the hit-sequence as specified in definition 2.1, that is to say $\{I_t\}_{t=1}^{T}$ is

$$I_t := 1(R_t < \text{VaR}_{t|t-1}(p)), \qquad t = 1, ...T$$

[3]Perignon and Smith (2010) find that 73% of banks that disclosed their VaR forecast method used HS.

The resulting ERFs of the backtests are presented in figure 4.1. Note that we use the Monte Carlo testing technique of Dufour (2006) rather than the critical values implied by the asymptotic distributions in evaluating the tests.

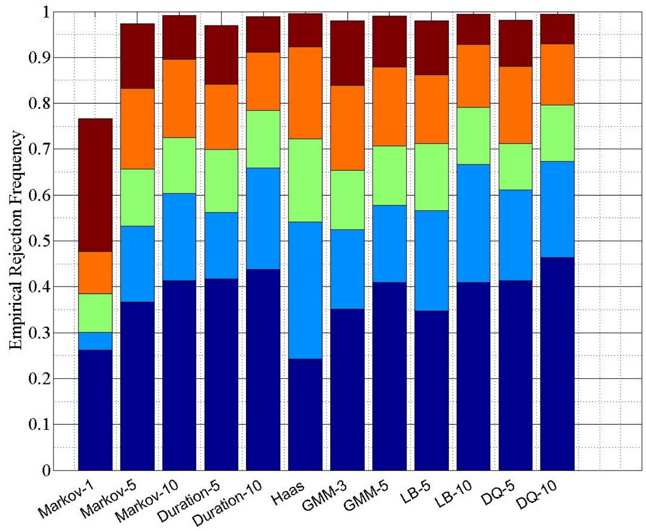
Inspecting figure 4.1, it is clear that the duration and generalized Markov tests improve the ERF compared with the original joint test. For example, when $1,000$ observations are available, using either the generalized Markov test or generalized duration tests with $k = 10$ lags will roughly double the ERF for either coverage rate. For the lower coverage rate, the Markov Duration test appears to perform slightly better than the generalized Markov test. However for the higher coverage rate the duration test can perform much worse, indicating its power is less robust (though still better than the original test). The results seem quite robust to the choice of T_W, although in general slightly better power was found when using $T_W = 500$ for all tests. This last result is as expected, since a longer window would be expected to increase the dependence in the hit-sequence. All the tests displayed largely similar power when using $1,000$ or more observations, with the exception of the test due to Christoffersen (1998). The tests due to Haas (2006) and Candelon et al. (2011) had somewhat varying power when using less than $1,000$ observations but had good power properties when using $1,000$ or more.

4 Concluding Remarks

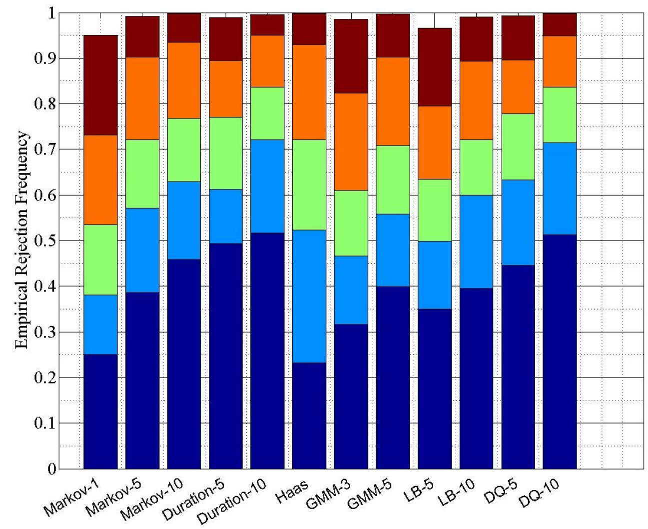
To summarize, we have introduced the generalized Markov framework for deriving Value-at-Risk backtests. Using the generalized Markov framework we suggested two specifications within this framework, the generalized Markov specification and the Markov Duration specification, inspired by the original backtests of Christoffersen (1998) and of the duration based backtests due to Christoffersen and Pelletier (2004), Haas (2006) and Pelletier and Wei (2015).

Based on these specifications we derived likelihood-ratio test statistics for the criteria of independence, unconditional coverage and conditional coverage. We provided closed form expressions for the tests, as well as asymptotic theory. Our tests have the advantage, compared to the original tests of Christoffersen (1998), that they possess power against k'th order dependence. Furthermore, the tests of conditional coverage is equivalent to the sum of the tests for independence and unconditional coverage. This allows one to evaluate rejection of conditional coverage as being caused by either dependence, an incorrect coverage rate or both.

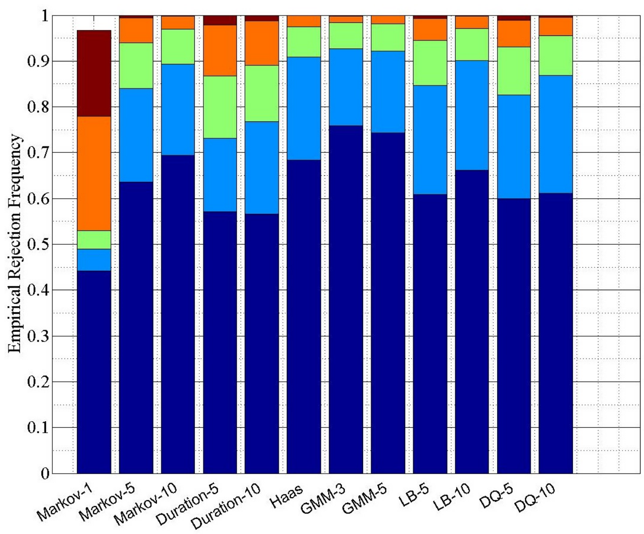
Using a simulation study we found evidence of improved size properties for the generalized Markov test compared to the original Markov test of Christoffersen (1998), though worse size properties for the Markov duration test. Simulations also indicated much improved empirical power while correcting for size distortions for the tests of conditional coverage based on either the generalized Markov or Duration specification compared to the original Markov test of Christoffersen (1998). Comparison to existing backtests showed that the generalized Markov test and the Markov duration test performed excellent in terms of empirical power, further their performance was robust to changes in the DGP used in the power simulations, unlike some of the backtests considered in the study.



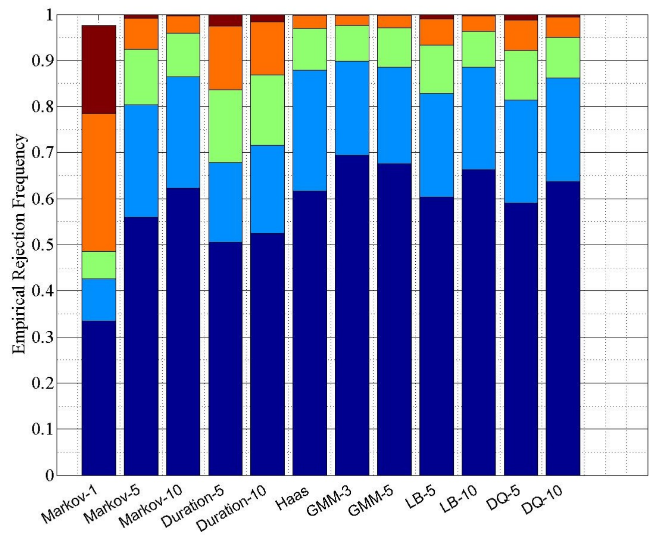
(a) $p = 1\%$ and $T_W = 500$



(b) $p = 1\%$ and $T_W = 250$



(c) $p = 5\%$ and $T_W = 500$



(d) $p = 5\%$ and $T_W = 250$

Figure 4.1: ERFs in percent for conditional coverage tests. The hit-sequences were simulated using a GARCH DGP with $VaR_{t|t-1}(p)$ estimated by historical simulation, using a rolling window of length T_W. The test Markov and Duration tests refer to the generalized Markov and generalized Duration tests developed in this paper, the Markov-1 test is also found in Christoffersen (1998) as the joint test. The Haas test is the test of Haas (2006), the GMM test is the test of Candelon et al. (2011), the LB test is the Ljung-Box test and the DQ test is the test of Engle and Manganelli (2004). The Monte Carlo testing technique of Dufour (2006) was used to ensure a size of 5%. The sample sizes used are 500 (dark blue), 1,000 (blue), 1,500 (green), 2,500 (orange) and 5,000 (red) as indicated by the colors.

References

Berkowitz, J., P. Christoffersen, and D. Pelletier (2011, December). Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science 57*(12), 2213–2227.

Billingsley, P. (1962). *statistical inference for Markov processes.* Chicago Press.

Brown, B. (1971). Martingale central limit theorems,. *The Annals of Mathematical Statistics 42*(1), 59–66.

Campbell, S. D. (2007). A review of backtesting and backtesting procedures. *Journal of Risk 9*(2), 1–18.

Candelon, B., G. Colletaz, C. Hurlin, and S. Tokpavi (2011, Spring). Backtesting Value-at-Risk: A GMM Duration-Based Test. *Journal of Financial Econometrics 9*(2), 314–343.

Christoffersen, P. and D. Pelletier (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics 2*, 84–108.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–62.

Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics 133*(2), 443–477.

Engle, R. F. and S. Manganelli (2004, October). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics 22*, 367–381.

Haas, M. (2006). Improved duration-based backtesting of value-at-risk. *Journal of Risk 8*(2), 17–38.

Jensen, S. r. T. and A. Rahbek (2004, December). Asymptotic Inference For Nonstationary Garch. *Econometric Theory 20*(06), 1203–1226.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95-24, Board of Governors of the Federal Reserve System (U.S.).

Lehmann, E. (1999). *Elements of Large-Sample Theory.* Springer.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools (Princeton Series in Finance).* Princeton University Press.

Pelletier, D. and W. Wei (2015). The geometric-var backtesting method. Forthcoming in journal of financial econometrics.

Perignon, C. and D. R. Smith (2010). The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking & Finance 34*(2), 362–377.

Pritsker, M. (2006, February). The hidden dangers of historical simulation. *Journal of Banking & Finance 30*(2), 561–582.

A Derivation of the Generalized Markov Test Distributions

A.1 Asymptotic Distribution for the Conditional Coverage Test

This proof is based on the usual Taylor expansions and verifies the conditions in Lemma 1 of Jensen and Rahbek (2004) for asymptotic inference, see also Theorem 7.7.3 of Lehmann (1999) or Billingsley (1962).

I. The score of the likelihood evaluated at the true value θ_0 satisfies $\frac{1}{\sqrt{T}}S_T(\theta_0) \xrightarrow{d} N(0,\Sigma)$ as $T \to \infty$

II. The observed information of the likelihood evaluated at the true value θ_0 satisfies $\frac{1}{T}i_T(\theta_0) \xrightarrow{p} \Sigma$ as $T \to \infty$

III. $\sup\limits_{\theta \in N(\theta_0)} \frac{1}{T}\left|\frac{\partial^3 L_T(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \leq C_T \xrightarrow{p} C \leq \infty$ as $T \to \infty$ and where $N(\theta_0)$ is a compact neighborhood around the true value θ_0 and $\theta_i, \theta_j, \theta_k = \{p_E, p_S\}$.

Condition (I): Recalling that $J_{t-1} = 1\left(\sum_{i=1}^{k} I_{t-1} > 0\right)$, the log-likelihood conditional on first k observations fixed is given by

$$L_T(\theta) = \sum_{t=1}^{T} L_t(\theta) = \sum_{t=1}^{T} I_t log\left(J_{t-1}p_E + (1-J_{t-1})p_S\right) + (1-I_t)log\left(1 - (J_{t-1}p_E + (1-J_{t-1})p_S)\right)$$

Next, the score with respect to θ is given by,

$$S_T(\theta) = \sum_{t=1}^{T} s_t(\theta) = \sum_{t=1}^{T} \frac{\partial L_t(\theta)}{\partial\theta} = \sum_{t=1}^{T} \begin{bmatrix} \frac{I_t J_{t-1}}{J_{t-1}p_E+(1-J_{t-1})p_S} - \frac{(1-I_t)J_{t-1}}{1-(J_{t-1}p_E+(1-J_{t-1})p_S)} \\ \frac{I_t(1-J_{t-1})}{J_{t-1}p_E+(1-J_{t-1})p_S} - \frac{(1-I_t)(1-J_{t-1})}{1-(J_{t-1}p_E+(1-J_{t-1})p_S)} \end{bmatrix} = \begin{bmatrix} \frac{T_{11}}{p_E} - \frac{T_{10}}{1-p_E} \\ \frac{T_{01}}{p_S} - \frac{T_{00}}{1-p_S} \end{bmatrix}$$

Here $T_{11} := \sum_{t=1}^{T} I_t J_{t-1}$, $T_{01} := \sum_{t=1}^{T} I_t(1-J_{t-1})$, $T_{10} := \sum_{t=1}^{T}(1-I_t)J_{t-1}$, $T_{00} := \sum_{t=1}^{T}(1-I_t)(1-J_{t-1})$. Recalling the definition of $\hat{\theta}$ we have that

$$\hat{\theta} = \begin{bmatrix} \frac{T_{11}}{T_{10}+T_{11}} \\ \frac{T_{01}}{T_{00}+T_{01}} \end{bmatrix} = \begin{bmatrix} \frac{T_{11}}{T_1} \\ \frac{T_{01}}{T_0} \end{bmatrix},$$

with $T_1 := T_{11} + T_{01}$, $T_0 := T_{10} + T_{00}$.

The distribution of $S_T(\theta_0)$, where θ_0 is the true value of $\theta \in \Theta_H$, can be found as

$$S_T(\theta_0) = \sum_{t=1}^{T} s_t(\theta_0) = \sum_{t=1}^{T} \frac{1}{p(1-p)}(I_t - p) \begin{bmatrix} J_{t-1} \\ 1 - J_{t-1} \end{bmatrix}.$$

Since $s_t(\theta_0)$ is a vector of martingale difference sequences with respect to $\mathcal{F}_{t-1}$, with conditional covariance matrix

$$E(s_t(\theta_0)s_t(\theta_0)') = E\begin{bmatrix} \frac{(I_t-p)^2 J_{t-1}}{p^2(1-p)^2} & 0 \\ 0 & \frac{(I_t-p)^2(1-J_{t-1})}{p^2(1-p)^2} \end{bmatrix} = \begin{bmatrix} \frac{1-(1-p)^k}{p(1-p)} & 0 \\ 0 & \frac{(1-p)^k}{p(1-p)} \end{bmatrix} =: \Sigma$$

and as $s_t(\theta_0)$ is stationary with finite third order moments, it follows from the martingale difference central limit theorem in Brown (1971) that as $T \to \infty$

$$\frac{1}{\sqrt{T}} S_T(\theta_0) \xrightarrow{d} \Sigma^{1/2} U$$

where $U := wlim \left(\Sigma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{p(1-p)} (I_t - p) \begin{bmatrix} J_{t-1} \\ 1 - J_{t-1} \end{bmatrix} \right) = N(0, I_2)$ and where I_2 is the identity matrix.

Condition (II): The observed information is given by

$$i_T(\theta) := -\frac{\partial^2 L_T(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{T_{11}}{p_E^2} + \frac{T_{10}}{(p_E-1)^2} & 0 \\ 0 & \frac{T_{01}}{p_S^2} + \frac{T_{00}}{(p_S-1)^2} \end{bmatrix}.$$

It follows that as $T \to \infty$

$$\begin{aligned} \frac{1}{T} i_T(\theta_0) &\xrightarrow{p} \begin{bmatrix} \frac{p\left(1-(1-p)^k\right)}{p^2} + \frac{(1-p)\left(1-(1-p)^k\right)}{(p-1)^2} & 0 \\ 0 & \frac{p(1-p)^k}{p^2} + \frac{(1-p)(1-p)^k}{(p-1)^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1-(1-p)^k}{p(1-p)} & 0 \\ 0 & \frac{(1-p)^k}{p(1-p)} \end{bmatrix} = \Sigma, \end{aligned}$$

By using the law of large numbers for i.i.d. observations. Observe in particular that as $T \to \infty$, $\frac{1}{T} T_{11} = \frac{1}{T} \sum_{t=1}^{T} I_t J_{t-1} \xrightarrow{p} p\left(1 - (1-p)^k\right)$, $\frac{1}{T} T_{10} = \frac{1}{T} \sum_{t=1}^{T} (1 - I_t) J_{t-1} \xrightarrow{p} (1-p)\left(1 - (1-p)^k\right)$, $\frac{1}{T} T_{00} = \frac{1}{T} \sum_{t=1}^{T} (1 - I_t)(1 - J_{t-1}) \xrightarrow{p} (1-p)(1-p)^k$ and $\frac{1}{T} T_{01} = \frac{1}{T} \sum_{t=1}^{T} I_t (1 - J_{t-1}) \xrightarrow{p} p(1-p)^k$.

Condition (III): Define $0 < p_E^L \le p_E \le p_E^U < 1$ and $0 < p_S^L \le p_S \le p_S^U < 1$ such that $max(p_S^L, p_E^L) \le p \le min(p_S^U, p_E^U)$. We verify, using the above results, that since $\frac{\partial^2 L_T(\theta)}{\partial p_E \partial p_S} = 0$ it follows that

$$\frac{1}{T} \left| \frac{\partial^3 L_T(\theta)}{\partial^3 p_E} \right| = \frac{1}{T} \left| \frac{2T_{11}}{\left(p_E^L\right)^3} + \frac{2T_{10}}{\left(p_E^U - 1\right)^3} \right| \le \frac{1}{T} \left(\frac{2T_{11}}{\left(p_E^L\right)^3} + \frac{2T_{10}}{\left(p_E^U - 1\right)^3} \right) = C_T \xrightarrow{p} c, \ for\ T \to \infty$$

$$\frac{1}{T} \left| \frac{\partial^3 L_T(\theta)}{\partial^3 p_S} \right| = \frac{1}{T} \left| \frac{2T_{01}}{\left(p_S^L\right)^3} + \frac{2T_{10}}{\left(p_S^U - 1\right)^3} \right| \le \frac{1}{T} \left(\frac{2T_{01}}{\left(p_S^L\right)^3} + \frac{2T_{10}}{\left(p_S^U - 1\right)^3} \right) = C_T \xrightarrow{p} c, \ for\ T \to \infty$$

Having verified the conditions we can derive the Q_{G-CC} test statistics asymptotic distribution for $T \to \infty$ as

$$
\begin{aligned}
Q_{G-CC} &= -2\left(L_T(\theta_0) - L_T(\hat{\theta})\right) \\
&= (\hat{\theta}-\theta_0)^{'} i(\theta_0)(\hat{\theta}-\theta_0) + o_p(1) \\
&= \frac{1}{\sqrt{T}} S_T(\theta_0)^{'} \left(\frac{1}{T} i(\theta_0)\right)^{-1} \frac{1}{\sqrt{T}} S_T(\theta_0) + o_p(1) \\
&\xrightarrow{d} U^{'}\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}U = U^{'}U \sim \chi^2(2)
\end{aligned}
$$

A.2 Asymptotic Distribution for the Unconditional Coverage Test

The asymptotic distribution of the Q_{G-UC} test is found in the same fashion using

$$
\frac{\partial L(\theta)}{\partial p} = \frac{\partial L(\theta)}{\partial \theta}\frac{\partial \theta}{\partial p} = S_T(\theta)^{'} H, \quad -\frac{\partial^2 L(\theta)}{\partial p \partial p} = H^{'} i(\theta_0)_T H
$$

where we recall that $H = (1,1)'$and the definition of $\tilde{\theta}$, it then follows that

$$
\tilde{\theta} = \frac{T_1}{T_0 + T_1}.
$$

Then as $T \to \infty$

$$
\begin{aligned}
Q_{G-UC} &= -2\left(L_T(\theta_0) - L_T(\tilde{\theta})\right) \\
&= [\frac{1}{\sqrt{T}} S_T(\theta)^{'} H]^{'} \left(\frac{1}{T} H^{'} i_T(\theta_0) H\right)^{-1} [\frac{1}{\sqrt{T}} S_T(\theta)^{'} H] + o_p(1) \\
&\xrightarrow{d} U^{'}\Sigma^{1/2} H (H^{'}\Sigma H)^{-1} H \Sigma^{1/2} U \\
&\sim \chi^2(1)
\end{aligned}
$$

A.3 Asymptotic Distribution for the Independence Test

Using the projection $I = \Sigma^{1/2}H(H^{'}\Sigma H)^{-1}H^{'}\Sigma^{1/2} + \Sigma^{-1/2}H_{\perp}(H_{\perp}^{'}\Sigma^{-1}H_{\perp})^{-1}H_{\perp}^{'}\Sigma^{-1/2}$, where $H_{\perp}$ designates the orthogonal complement of H, we can now find the asymptotic distribution of Q_{G-Ind} as $T \to \infty$

$$
\begin{aligned}
Q_{G-Ind} &= Q_{G-CC} - Q_{G-UC} \\
&\xrightarrow{d} U^{'}U - U^{'}\Sigma^{1/2}H(H^{'}\Sigma H)^{-1}H^{'}\Sigma^{1/2}U \\
&= U^{'}\left(I - \Sigma^{1/2}H(H^{'}\Sigma H)^{-1}H^{'}\Sigma^{1/2}\right)U \\
&= U^{'}\Sigma^{-1/2}H_{\perp}(H_{\perp}^{'}\Sigma^{-1}H_{\perp})^{-1}H_{\perp}^{'}\Sigma^{-1/2}U = A^{'}A \sim \chi^2(1),
\end{aligned}
$$

where $A := (H_{\perp}^{'}\Sigma^{-1}H_{\perp})^{-1/2}H_{\perp}^{'}\Sigma^{-1/2}U$.

B Derivation of the Markov Duration Test Distributions

B.1 Asymptotic Distribution for the Conditional Coverage Test

We proceed as in the proof for Theorem 1.

Condition (I): Recalling that $J(k)_{t-1} = 1\left(I_{t-1} = 0, ..., I_{t-k} = 1\right)$, the log-likelihood conditional on first k observations fixed is given by

$$\begin{aligned} L_T(\theta) &= \sum_{t=1}^{T} I_t log\left(J(1)_{t-1}p_{E1} + ... + J(k)_{t-1}p_{Ek} + (1 - \sum_{i=1}^{k} J(i)_{t-1})p_S\right) \\ &\quad +(1-I_t)log\left(1 - \left(J(1)_{t-1}p_{E1} + ... + J(k)_{t-1}p_{Ek} + (1 - \sum_{i=1}^{k} J(i)_{t-1})p_S\right)\right) \end{aligned}$$

Next, the score with respect to θ is given by,

$$S_T(\theta) = \sum_{t=1}^{T} s_t(\theta) = \sum_{t=1}^{T} \frac{\partial L_t(\theta)}{\partial \theta} = \sum_{t=1}^{T} \begin{bmatrix} \frac{I_t J(1)_{t-1}}{p_{E1}} - \frac{(1-I_t)J(1)_{t-1}}{1-p_{E1}} \\ ... \\ \frac{I_t J(k)_{t-1}}{p_{Ek}} - \frac{(1-I_t)J(k)_{t-1}}{1-p_{Ek}} \\ \frac{I_t(1-\sum_{i=1}^{k} J(i)_{t-1})}{p_S} - \frac{(1-I_t)(1-\sum_{i=1}^{k} J(i)_{t-1})}{1-p_S)} \end{bmatrix} = \begin{bmatrix} \frac{T_{11}(1)}{p_{E1}} - \frac{T_{10}(1)}{1-p_{E1}} \\ \vdots \\ \frac{T_{11}(k)}{p_{Ek}} - \frac{T_{10}(k)}{1-p_{Ek}} \\ \frac{T_{11}}{p_S} - \frac{T_{00}}{1-p_S} \end{bmatrix}$$

Recalling the definition of $\hat{\theta}$ we have that $\hat{\theta} = \begin{bmatrix} \frac{T_{11}(1)}{T_{11}(1)+T_{10}(1)} & \cdots & \frac{T_{11}(k)}{T_{11}(k)+T_{10}(k)} & \frac{T_{11}}{T_{11}+T_{01}} \end{bmatrix}'$.
Recalling that $J_{t-1} := 1\left(\sum_{i=1}^{k} I_{t-1} > 0\right)$, the distribution of $S_T(\theta_0)$, where θ_0 is the true value of $\theta \in \Theta_H$, can be found as

$$S_T(\theta_0) = \sum_{t=1}^{T} s_t(\theta_0) = \sum_{t=1}^{T} \frac{1}{p(1-p)}(I_t - p) \begin{bmatrix} J(1)_{t-1} \\ \vdots \\ J(k)_{t-1} \\ 1 - J_{t-1} \end{bmatrix}.$$

Since $s_t(\theta_0)$ is a vector of martingale difference sequences with respect to $\mathcal{F}_{t-1}$, with conditional covariance matrix

$$E(s_t(\theta_0)s_t(\theta_0)') = \begin{bmatrix} \frac{1}{(1-p)} & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & (1-p)^{k-2} & 0 \\ 0 & 0 & 0 & \frac{(1-p)^{k-1}}{p} \end{bmatrix} =: \Sigma$$

and as $s_t(\theta_0)$ is stationary with finite third order moments, it follows from the martingale difference central limit theorem in Brown (1971) that as $T \to \infty$

$$\frac{1}{\sqrt{T}} S_T(\theta_0) \xrightarrow{d} \Sigma^{1/2} U$$

where $U := wlim \left(\Sigma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{p(1-p)} (I_t - p) \begin{bmatrix} J(1)_{t-1} & \ldots & J(k)_{t-1} & 1 - J_{t-1} \end{bmatrix}' \right) = N(0, I_{k+1})$ and where I_{k+1} is the identity matrix.

Condition (II): The observed information is given by

$$i_T(\theta_0) := -\frac{\partial^2 L_T(\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{T_{11}(1)}{p_{E1}^2} + \frac{T_{10}(1)}{(1-p_{E1})^2} & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \frac{T_{11}(k)}{p_{Ek}^2} + \frac{T_{10}(k)}{(1-p_{Ek})^2} & 0 \\ 0 & 0 & 0 & \frac{T_{01}}{p_S^2} + \frac{T_{00}}{(1-p_S)^2} \end{bmatrix}$$

It follows that as $T \to \infty$ $\frac{1}{T} i_T(\theta_0) \xrightarrow{p} \Sigma$ by the law of large numbers for i.i.d. observations, and that as $T \to \infty$, $\frac{1}{T} T_{11}(k) \to (1-p)^{k-1} p^2$ and $\frac{1}{T} T_{01}(k) \to (1-p)^k p$.

Condition (III): Define $0 < p_{Ek}^L \leq p_{Ek} \leq p_{Ek}^U < 1$ and $0 < p_S^L \leq p_S \leq p_S^U < 1$ such that $max(p_S^L, p_E^L) \leq p \leq min(p_S^U, p_E^U)$. Since $\frac{\partial^2 L_T(\theta)}{\partial p_E \partial p_S} = 0$ it follows that we only need the following

$$\frac{1}{T}\left|\frac{\partial^3 L_T(\theta)}{\partial^3 p_{Ek}}\right| = \frac{1}{T} \left|\frac{2T_{11}(k)}{\left(p_{Ek}^L\right)^3} + \frac{2T_{10}(k)}{\left(p_{Ek}^U - 1\right)^3}\right| \leq \frac{1}{T}\left(\frac{2T_{11}(k)}{\left(p_{Ek}^L\right)^3} + \frac{2T_{10}(k)}{\left(p_{Ek}^U - 1\right)^3}\right) = C_T \xrightarrow{p} c, \ for\ T \to \infty$$

$$\frac{1}{T}\left|\frac{\partial^3 L_T(\theta)}{\partial^3 p_S}\right| = \frac{1}{T}\left|\frac{2T_{0,1}}{\left(p_S^L\right)^3} + \frac{2T_{1,0}}{\left(p_S^U - 1\right)^3}\right| \leq \frac{1}{T}\left(\frac{2T_{0,1}}{\left(p_S^L\right)^3} + \frac{2T_{1,0}}{\left(p_S^U - 1\right)^3}\right) = C_T \xrightarrow{p} c, \ for\ T \to \infty$$

Having verified the conditions we can derive the Q_{D-CC} test statistics asymptotic distribution for $T \to \infty$ as

$$Q_{D-CC} \xrightarrow{d} \chi^2(k+1)$$

B.2 Asymptotic Distribution for the Unconditional Coverage Test

The asymptotic distribution of the Q_{D-UC} test is found in the same fashion using

$$\frac{\partial L(\theta)}{\partial p} = \frac{\partial L(\theta)}{\partial \theta}\frac{\partial \theta}{\partial p} = S_T(\theta)^{'} H, \quad -\frac{\partial^2 L(\theta)}{\partial p \partial p} = H^{'} i(\theta_0)_T H$$

where we recall that $H = (1, ..., 1)'$and the definition of $\tilde{\theta}$, it then follows that $\tilde{\theta} = \frac{T_1}{T_0+T_1}$.

Then as $T \to \infty$

$$Q_{D-UC} \quad \overset{d}{\to} \quad \chi^2(1)$$

B.3 Asymptotic Distribution for the Independence Test

Using the projection $I = \Sigma^{1/2} H(H^{'}\Sigma H)^{-1} H^{'}\Sigma^{1/2} + \Sigma^{-1/2} H_{\perp}(H_{\perp}^{'}\Sigma^{-1} H_{\perp})^{-1} H_{\perp}^{'}\Sigma^{-1/2}$ we can now find the asymptotic distribution of Q_{D-Ind} as $T \to \infty$

$$Q_{D-Ind} \quad \overset{d}{\to} \quad \chi^2(1),$$

C The Monte Carlo Testing Technique Dufour (2006)

In this section we outline the Monte Carlo testing technique of Dufour (2006) used in the empirical power simulations of section C. The technique used is given by the following algorithm:

I. Generate M i.i.d. hit-sequences of length T, $\{I_t\}_{t=1}^{T}$, under the null of conditional coverage, H_{CC}, by drawing from a Bernoulli sequence, as:

$$I_t \underset{i.i.d}{\sim} Bernoulli(p), \qquad t = 1, ..., T$$

II. Calculate the test statistic, S_i, for each of the generated hit-sequence, $i = 1, ..., M$ and denote by S_0 the original test value. Throughout this paper we use $M = 99, 999$.

III. Draw U_i for $i = 0, ..., M$ from the uniform $U(0, 1)$ distribution. Calculate the p-values as $\hat{p}_M(S_0) = \frac{M\hat{G}_M(S_0)+1}{M+1}$ where $\hat{G}_M(S_0) = 1 - \frac{1}{M}\sum_{i=1}^{M} 1\,(S_i \leq S_0) + \frac{1}{M}\sum_{i=1}^{M} 1\,(S_i = S_0)\,1\,(U_i \geq U_0)$.

D Power Tables

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	28.63	46.71	34.77	39.00	32.21	30.953	27.46	27.183	34.89	27.87	49.88	44.42
1000	44.70	77.17	59.64	60.69	51.38	61.536	49.35	51.255	59.76	48.36	74.22	67.78
1500	53.51	91.78	77.19	77.77	68.93	80.998	69.59	73.567	72.91	63.24	86.48	81.68
2500	81.02	99.06	94.04	96.26	91.52	96.12	89.13	92.189	92.61	85.82	97.55	95.48
5000	98.1	100	99.87	99.98	99.91	99.93	99.64	99.8	99.83	99.32	99.99	99.95

(a) Specification 1: $k = 5$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	52.13	60.23	71.79	51.03	56.06	54.59	55.65	51.90	22.47	31.07	61.57	67.67
1000	76.42	87.44	95.4	75.29	79.59	88.02	86.31	85.25	37.44	50.41	85.36	89.01
1500	87.38	96.71	99.37	89.19	93.32	97.45	97.05	96.99	45.41	64.44	94.45	96.74
2500	98.67	99.84	99.99	99.08	99.66	99.9	99.84	99.88	67.49	87.28	99.62	99.75
5000	100	100	100	100	100	100	100	100	92.369	99.35	100	100

(b) Specification 1: $k = 10$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	24.48	36.67	26.85	31.65	25.67	25.42	21.30	21.13	30.59	24.52	42.70	37.62
1000	38.33	64.62	47.46	49.09	40.02	52.42	37.94	40.53	52.08	41.69	64.62	57.77
1500	45.60	82.92	64.70	66.07	56.14	72.74	56.73	61.73	65.26	55.11	78.54	71.98
2500	73.71	96.52	86.039	90.10	82.24	91.86	78.42	83.41	87.10	78.85	93.96	90.36
5000	96.02	99.92	99.00	99.78	99.35	99.78	97.86	98.98	99.16	97.83	99.82	99.61

(c) Specification 2: $k = 5$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	37.58	47.79	52.64	40.65	39.41	40.87	38.72	36.50	26.21	28.31	51.93	53.78
1000	59.63	77.08	82.32	62.31	59.26	75.24	68.67	68.87	43.52	45.40	75.91	76.36
1500	71.77	91.38	94.28	79.03	77.91	91.29	87.32	88.09	54.27	58.60	88.12	88.57
2500	93.11	99.08	99.64	96.49	96.26	99.27	98.19	98.52	77.16	81.48	98.06	98.07
5000	99.83	100	100	99.97	100	100	100	100	96.57	98.42	100	100

(d) Specification 2: $k = 10$

Table 2: Empirical power in percent for conditional coverage tests. The hit-sequences were simulated using a k'th order Markov chain specified in equation (3.1) of section 3. The tests refer to the generalized Markov and generalized Duration tests developed in this paper, the Markov-1 test is also found in Christoffersen (1998) as the joint test. The Haas test is the test of Haas (2006), the GMM test (with 3 or 5 moments) is the test of Candelon et al. (2011), the LB test is the Ljung-Box test (with 5 or 10 lags) and the DQ test is the test of Engle and Manganelli (2004) (with 5 or 10 lags). The Monte Carlo testing technique of Dufour (2006) was used to ensure a size of 5%

E Scenario Power Tables

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	26.13	36.72	41.28	41.76	43.75	24.29	35.10	40.92	34.78	40.98	41.34	46.33
1000	30.11	53.19	60.33	56.22	65.90	54.09	52.41	57.79	56.62	66.73	61.16	67.32
1500	38.54	65.71	72.49	69.92	78.50	72.27	65.38	70.66	71.25	79.06	71.23	79.68
2500	47.66	83.28	89.55	84.14	91.09	92.26	83.85	87.92	86.17	92.76	88.08	92.95
5000	76.65	97.34	99.14	96.97	98.87	99.59	98.03	99.05	98.01	99.45	98.10	99.41

(a) $p = 1\%$ and $T_W = 500$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	25.05	38.65	45.87	49.40	51.64	23.20	31.58	39.90	35.04	39.47	44.57	51.29
1000	38.12	57.12	62.95	61.25	72.17	52.34	46.58	55.80	49.93	59.91	63.29	71.47
1500	53.45	72.06	76.77	77.06	83.63	72.16	61.02	70.89	63.52	72.13	77.86	83.64
2500	73.20	90.20	93.50	89.44	94.97	92.99	82.37	90.18	79.51	89.36	89.58	94.93
5000	95.06	99.20	99.80	98.93	99.55	99.81	98.57	99.63	96.63	99.05	99.24	99.78

(b) $p = 1\%$ and $T_W = 250$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	44.21	63.56	69.46	57.12	56.65	68.34	75.86	74.27	60.86	66.12	59.91	61.10
1000	48.95	84.06	89.27	73.22	76.72	90.83	92.69	92.16	84.70	90.12	82.60	86.91
1500	52.94	93.93	96.96	86.75	89.03	97.46	98.33	98.09	94.50	97.04	93.02	95.57
2500	78.00	99.37	99.77	97.83	98.75	99.90	99.86	99.89	99.26	99.75	98.86	99.59
5000	96.68	100	100	99.99	100	100	100.00	100.00	100	100	100	100

(c) $p = 5\%$ and $T_W = 500$

T	Markov -1	Markov -5	Markov -10	Duration -5	Duration -10	Haas	GMM -3	GMM -5	LB -5	LB -10	DQ -5	DQ -10
500	33.43	55.98	62.24	50.54	52.42	61.68	69.44	67.58	60.31	66.34	59.08	63.78
1000	42.68	80.44	86.49	67.90	71.62	87.91	89.84	88.58	82.83	88.61	81.38	86.19
1500	48.56	92.49	95.89	83.62	86.84	96.93	97.61	97.15	93.30	96.26	92.13	94.96
2500	78.43	99.20	99.72	97.45	98.37	99.85	99.82	99.77	99.07	99.64	98.72	99.48
5000	97.61	100	100	99.99	99.99	100	100.00	100	100	100	100	100

(d) $p = 5\%$ and $T_W = 250$

Table 3: ERFs in percent for conditional coverage tests. The hit-sequences were simulated using a GARCH DGP with $VaR_{t|t-1}(p)$ estimated by historical simulation, using a rolling window of length T_W. The test Markov and Duration tests refer to the generalized Markov and generalized Duration tests developed in this paper, the Markov-1 test is also found in Christoffersen (1998) as the joint test. The Haas test is the test of Haas (2006), the GMM test (with 3 or 5 moments) is the test of Candelon et al. (2011), the LB test is the Ljung-Box test (with 5 or 10 lags) and the DQ test is the test of Engle and Manganelli (2004) (with 5 or 10 lags). The Monte Carlo testing technique of Dufour (2006) was used to ensure a size of 5%

Forecasting Portfolio Value-at-Risk: Choices in Portfolio Aggregation, Data-frequency and Distribution*

Thor Pajhede[†]And Erik Kole[‡]

January 31, 2017

Abstract

When forecasting the Value-at-Risk (VaR) of a portfolio, one faces a number of choices in how to construct a model; univariate or multivariate models, interday or intraday based data and distributional alternatives. We examine the performance of 40 models from the combinations of these choices for forecasting the VaR at different coverage levels of a realistically sized stock portfolio. Our data is a portfolio of 44 major US stocks from the S&P 500 index for which we compare the forecasting ability of the models - using both recently developed backtests and the model confidence set approach. We generally find mixed results neither clearly favoring univariate nor multivariate models. Likewise there is no clear advantage to either using intraday or interday based models. No apparent advantage is found in using realized kernels compared to using realized variance sampled at 10 minutes frequency. However, when combining univariate models with filtered historical simulation methods, we consistently find good forecasting ability. Models with fixed parameters such as the EWMA model and multivariate variations, both intraday and interday based, generally also have strong performance. We also consider the square-root-of-time scaling rule for a 10 day period as suggested in the Basel Accords and find that it consistently performs poorly, generally leading to VaR forecasts for which we reject the independence but not the unconditional coverage criteria.

Keywords: Value-at-Risk, Backtest, Risk Management, Portfolio.

JEL codes: C12, C15, C52, C32.

*The author acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. We thank Oliver Linton, Lena Hasselberg, Asger Lunde, Peter Reinhard Hansen, Rasmus Søndergaard, Rogier Quaedvlieg, Anders Rahbek and Heino Bohn Nielsen for helpful comments. All calculations were done in Matlab2016B and Ox.

[†]Department of Economics, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K, Denmark. CREATES, University of Aarhus. E-mail: thor.nielsen@econ.ku.dk

[‡]Erasmus University Rotterdam - Erasmus School of Economics - Econometric Institute; ERIM; Tinbergen Institute

1 Introduction

Financial econometrics has developed a range of tools for modeling and forecasting market risk. Most prominent among these tools is the (G)ARCH class of models first suggested in Engle (1982), which has since been expanded on by many authors to allow for effects such as different conditional distributions, leverage effects or a multivariate setting. More recently, the availability of intraday data, specifically using realized variance and realized covariance, see Andersen et al. (2003), and models using these measures have been developed and currently compose a very active area of research.

In this paper we examine the importance of choosing between univariate and multivariate models (portfolio aggregation), intraday or interday data (data frequency) and conditional distributions when forecasting VaR for a portfolio of several stocks at different coverage levels and different forecasting horizons. After considering all of these options one must also consider what conditional distribution best fits the data, typical distributions used are the Gaussian, t or empirical distributions.

How to best forecast risk has important implications for both practitioners and researchers. Using interday data based models may be the most common approach as models using daily data has been extensively studied over the past 30 years. However, intraday data may provide a clearer signal of market volatility, resulting in more accurate forecasts, see Andersen and Benzoni (2008), but is possibly affected by market microstructure effects and may be costly to obtain and store. Similarly, modeling large portfolios by a multivariate model may lead to improvements as the dependencies of the portfolio are taken into account. However, as the number of assets in a portfolio increases the estimated number of parameters typically increases faster, the so-called curse of dimensionality. This could reduce the forecasting ability of the model. It follows that the easiest models to use when forecasting a portfolios VaR would be univariate interday based models, but it may well be the case that adopting models which are intraday based, multivariate or both is required for an accurate forecast.

Previously, several papers have argued that the VaR of a portfolio is best forecasted using univariate rather than multivariate models, see eg. Persand and Brooks (2003) and Berkowitz and O'Brien (2001), while McAleer and da Veiga (2008) found rather mixed evidence. However, as noted in Santos et al. (2013), the results of these studies are based on small portfolios of around 4 assets. Conversely, using several larger portfolios, they find evidence in favor of multivariate models. However, the backtests used are known to have low power and to be size distorted, see Christoffersen (1998) and Pajhede (2017). Kole et al. (2015) find that for longer forecasting periods multivariate models with daily frequency performs the best. However, none of the papers so far mentioned consider any intraday based models.

Using univariate models, several papers find superior forecasts for stock indexes, currencies and single stocks when using intraday data, see Koopman et al. (2005), Clements et al. (2006) and Shephard and Sheppard (2010). This again suggests that the choice of data frequency may be important when forecasting VaR.

Boudt et al. (2014) suggest a model for forecasting large covariance matrices using high frequency data. In an empirical application they published the first paper that compares multivariate interday and intraday type models for

VaR prediction, they find that their model provides superior forecasts compared to interday type models. However, the study is limited by using a single dataset, the size of the portfolio and not comparing to more sophisticated multivariate models as in Santos et al. (2013) or Kole et al. (2015) and restricting themselves to Gaussian models, which has been shown to perform worse in both univariate and multivariate models, see Luc Bauwens (2005) and Braione and Scholtes (2014). A central issue is the topic of dealing with asynchronicity in data, with the existing methods imposing a large loss of data for larger observations, see the discussion in section 4.

This paper contributes to the VaR literature for large portfolios by providing a comprehensive comparison of the forecasts of VaR for a large portfolio from across the categories of univariate, multivariate, interday and intraday models. We include several models from all categories and allow for Gaussian, Student-t or empirical distributions for univariate models and Gaussian or empirical distributions for multivariate models. We examine multiple forecast horizons and VaR coverage levels.

We compare the competing forecasts in two ways; Hit-sequence based backtests known to have good power properties, see Pajhede (2017), and the model confidence set (MCS) approach of Hansen et al. (2011).

The rest of the paper is organized as follows. Section 2 formalizes VaR and its calculation from a conditional distribution. Section 3 introduces interday based models, specifically univariate and multivariate GARCH type models as well as Conditional Autoregressive Value at Risk (CAViaR) models. Section 4 gives a brief discussion of realized variance and then discuss several volatility models using high frequency data, both univariate and multivariate. Section 5 details the tools used for comparison, hit-sequence based backtests and the MCS approach. 6 contains the empirical application. Section 7 concludes.

2 Value-at-Risk

Let $y_{p,t} = w'_{t-1} r_t$ denote the realization of a return of a portfolio of N assets at day t, where w'_{t-1} and r_t are the $N \times 1$ vectors of portfolio weights and asset returns respectively. The ex-ante VaR for time t and coverage rate α, $\mathrm{VaR}_{t|t-1}(\alpha)$, conditional on all information, $\mathcal{F}_{t-1}$, available at time $t-1$ is the α'th conditional quantile of the distribution of r_t:

$$P(y_{p,t} < \mathrm{VaR}_{t|t-1}(\alpha)|\mathcal{F}_{t-1}) = \alpha. \tag{2.1}$$

Throughout the rest of the paper we will consider the 1% and 5% coverage rates, as is common in the literature. When there is no ambiguity, we simplify the notation by writing $\mathrm{VaR}_{t|t-1}$.

We calculate the VaR of Equation (2.1) from a conditional distribution as

$$\mathrm{VaR}_{t|t-1}(\alpha) = \mu_{p,t} + \sigma_{p,t} F^{-1}(\alpha), \tag{2.2}$$

where $\mu_{p,t} = E\left(y_{p,t}|\mathcal{F}_{t-1}\right)$ and $\sigma_{p,t} = \sqrt{var\left(y_{p,t}|\mathcal{F}_{t-1}\right)}$ are the means and standard deviations of the portfolio return,

$y_{p,t}$, while $F^{-1}(\alpha)$ is the α quantile of the distribution of the standardized returns, $e_{p,t} = (y_{p,t} - \mu_{p,t})/\sigma_{p,t}$.

We set the expected return to a constant, μ_p, which we estimate by the average for the historical portfolio returns. Since we are forecasting one day ahead, the expected return of the portfolio should be negligible. The standard deviation, $\sigma_{p,t}$, is estimated directly when using a univariate model and as $\sigma_{p,t}^2 = w_t H_t w_t'$ when using a multivariate model for the conditional covariance matrix, $H_t = E\left((y_t - \mu_p)^2 | \mathcal{F}_{t-1}\right)$.

For the univariate models we calculate $F^{-1}(\alpha)$ from the following distributions; Gaussian, t and using filtered historical simulation (FHS), see Barone-Adesi and Giannopoulos (2001), based on the the empirical distribution of $e_{p,t}$. For the multivariate models, we use the multivariate Gaussian and FHS.

The only exception to this setup is the CAViaR model of Engle and Manganelli (2004), which directly models the VaR quantile.

To reduce the number of models we only consider specifications using a single lag for all models, eg. the GARCH(P,Q) model is implemented as the GARCH(1,1) model, this is also known to work well in practice, see Lunde and Hansen (2005). For the FHS forecasts we assume a Gaussian distribution for the model to estimate the parameters and standardize the returns.

3 Interday Based Models

In this section we describe the interday based univariate and multivariate models and estimation procedures which we use to generate VaR forecasts.

3.1 Univariate Models

Our selection of GARCH type models includes a variety of models representing most of the major strains of the literature. We include the GARCH model of Bollerslev (1986), the GJR-GARCH model of Glosten et al. (1993) which accounts for leverage effects, the APARCH model of Ding et al. (1993) which also includes leverage effects as well as the threshold ARCH class of Rabemananjara and Zakoian (1993), the RiskMetrics model and lastly the recent tv-t-GAS of Creal et al. (2013) which may produce forecasts more robust to outliers. The only non-GARCH type of models included is the CAViaR model of Engle and Manganelli (2004) of which we include 3 variations. Most models were estimated using code modified from the Oxford MFE toolbox of Kevin Shephard[1] or from the website of Simone Manganelli[2].

All the univariate GARCH type models can be written as

[1] https://www.kevinsheppard.com/MFE_Toolbox

[2] http://www.simonemanganelli.org/Simone/Research.html

$$\begin{cases} y_{p,t} = \mu_p + \epsilon_t \\ \epsilon_t = \sigma_{p,t} e_t \\ e_t \sim F \end{cases},$$

where F is either $N(0,1)$ or t_v. The GARCH specification is given by

$$\sigma^2_{p,t} = \omega + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{p,t-1},$$

The GJR-GARCH specification is given by

$$\sigma^2_{p,t} = \omega + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{p,t-1} + \delta 1\left(\epsilon_{t-1} < 1\right) y^2_{p,t-1},$$

where $1(\cdot)$ is the indicator function. The APARCH specification is given by

$$\sigma^2_{p,t} = \omega + \alpha\left(|e_{p,t-1}| + \delta e_{p,t-1}\right)^{\lambda} + \beta\sigma^2_{p,t-1}.$$

Lastly, the exponentially weighted moving average model (EWMA1994) from risk-metrics model is given by

$$\sigma^2_{p,t} = 0.94\sigma^2_{p,t-1} + 0.06\epsilon^2_{t-1}.$$

Rather than specifying the entire distribution of returns, the CAViaR model is a model for only the desired quantile. We use two specifications suggested in Engle and Manganelli (2004); the symmetric absolute value specification

$$\text{VaR}_{t|t-1} = \alpha + \beta_1 \text{VaR}_{t-1|t-2} + \beta_2|\epsilon_{t-1}|$$

and the asymmetric slope specification

$$\text{VaR}_{t|t-1} = \alpha + \beta_1 \text{VaR}_{t-1|t-2} + \beta_2(\epsilon_{t-1})1\left(\epsilon_{t-1} > 0\right) + \beta_3(\epsilon_{t-1})1\left(\epsilon_{t-1} < 0\right).$$

Estimation is achieved by minimizing the tick loss function as in quantile regression[3].

3.2 Multivariate Models

To model the conditional covariance matrix, H_t, we consider a number of multivariate GARCH type models. The primary challenge when using multivariate models is the curse of dimensionality, that is, the number of parameters growing rapidly for larger systems. For example, the VEC model of Bollerslev et al. (1988) parameters of order $O(N^4)$

[3] In the empirical application we use code modified from the website of Simone Manganelli.

and estimation is therefore not feasible for larger portfolios. In selecting models the first requirement is that they are feasible for a portfolio size of 44 assets.

The first model considered is the variance targeting scalar BEKK model of Engle and Kroner (1995), given by

$$H_t = (1 - a - b)\,\Gamma + A r_{t-1} r_{t-1}' + B H_{t-1}. \tag{3.1}$$

The use of variance targeting reduces the number of parameters in the scalar form to just the two a and b parameters, which are easily estimated by maximum likelihood using either a Gaussian or a multivariate t distribution. The unconditional variance, Γ, is estimated as the sample covariance $\hat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t'$, see Pedersen and Rahbek (2014).

Our second multivariate model is the constant conditional correlations (CCC) model of Bollerslev (1990), in this model the standard deviations, contained in D_t, are allowed to vary over time while the correlations, R, are kept constant. That is, the covariance matrix, H_t, is modeled as

$$H_t = D_t R D_t, \tag{3.2}$$

In estimating D_t we use the GJR-GARCH which has previously shown good performance for similar applications, see Santos et al. (2013), we do not consider alternative univariate models to keep the number of considered models manageable. This model is attractive because of its simplicity and ease of estimation even for larger portfolios. We estimate the CCC model using the two step procedures of first estimating D_t using univariate models, and then estimating R from the standardized residuals.

The variance targeting dynamic conditional correlations (DCC) model of Engle (2002) uses a similar decomposition of H_t but, here the conditional correlation matrix, R_t, is now time varying. R_t is equivalent to the covariance matrix of the standardized residuals, $R_t = V_{t-1}(e_t)$, which is modeled as

$$R_t = diag\left(Q_t\right)^{-1/2} Q_t diag\left(Q_t\right)^{-1/2} \tag{3.3}$$

where

$$Q_t = (1 - \alpha - \beta)\,\overline{R} + \alpha e_{t-1} e_{t-1}' + \beta Q_{t-1}.$$

with $\overline{R} = \frac{1}{T}\sum_{t=1}^{T} e_t e_t'$.

The models covered so far have been general models, designed to model multiple time-series in general, however the RiskMetrics group has also developed multivariate models, specifically intended to model the risk of portfolios. The RiskMetrics 1994 (RM94) model is given as

$$H_t = (1 - \lambda)\, r_{t-1} r_{t-1}' + \lambda H_{t-1}, \tag{3.4}$$

with $\lambda = 0.94$ for daily data.

4 Intraday Based Models

Merton (1980) noted that in a geometric Brownian motion, using higher frequencies of data will allow one to estimate volatility at arbitrary precision. In a real world setting this idea is subject to the discreteness of trading and other market microstructure effect generating noise in the estimates when using frequencies below 5-20 minutes, see Hansen and Lunde (2006).

The characteristics of this noise are dependent on the sampling frequency, but may be correlated with the price, time-dependent and has changed over time. To benefit from sampling at higher frequencies while being robust to noise, several methods for measuring realized variance (RV) have been estimated, we discuss these in subsection 4.1. We use two different estimates in our empirical application, realized volatility at a low frequency and kernel at higher frequencies.

When dealing with multivariate high frequency data one has to deal with the issue of synchronicity, while the daily models could safely be dependent on each stock being traded every day, it is highly unlikely that all the stocks of a portfolio are traded at exactly the same time throughout a day or the same number of times.

Apart from dealing with noise induced by market microstructure, one also has to be mindful of data cleaning which is known to be one of the most critical aspects of using intraday data, see Barndorff-Nielsen et al. (2009).

4.1 Measuring Variation

4.1.1 Univariate

To measure the daily variation in a stocks price at a higher frequency, we now divide day t into n subintervals

$$t_0 < t_1 < ... < t_{n-1} < t_n = t,$$

for which we observe the intradaily returns

$$r_{j,n} = \log(p_{t_j}) - log(p_{t_{j-1}}), \quad j = 1, ..., n$$

with p_{t_j} the price at observation j on day t.

Since Andersen and Bollerslev (1998) used RV, also commonly referred to as realized volatility, to evaluate the forecasting ability of volatility models it has been apparent that RV, at least in principle, offers a measure of volatility which is much less noisy than squared daily returns. We define the RV of day t as

$$RV_t := \sum_{i=1}^{n} r_{j,n}^2$$

However, as noted as early as Merton (1980) RV is susceptible to noise, see Roll (1984) and Hansen and Lunde (2006). Hence sampling at higher frequencies can make RV even noisier than simply using squared daily returns. Hansen and Lunde (2006) suggest not to sample more frequently than the $5 - 10$ minute frequency to reduce bias, though this comes at a price of not using much of the data efficiently. In this paper we use a 10 minute frequency for estimating RV.

However, this loss of data has motivated robust measures of realized volatility such as pre-averaging, subsampling and realized kernels, see Barndorff-Nielsen et al. (2008). Realized kernel estimators are for day t is defined as

$$RK_t = \sum_{h=-H}^{H} k\left(\frac{h}{H}\right)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{h} r_{j,n} r_{j-|h|,n},$$

where $k(\cdot)$ is a kernel function. We used a Parzen kernel with an optimally selected bandwidth (H), see Barndorff-Nielsen et al. (2008).

4.1.2 Multivariate

When attempting to measure the volatility of several assets at high frequency one must deal with the issue of asynchronicity. That is, stocks are generally traded at different times, making a return of the portfolio of stocks difficult to measure. One solution to the issue of asynchronicity has been to consider refresh sampling the prices, see Barndorff-Nielsen et al. (2008), which samples prices using the newest trades but only once all assets have changed price at least once since the last sampling. That is, for $t \in [0,1]$ with 0 being the start of trading and 1 being the end, and the portfolio consisting of N assets the first refresh-time is $\tau_1^N = max(t_1^{(1)}, ..., t_1^{(N)})$ with $t_1^{(i)}$ indicating the first time asset i is traded. Subsequent refresh-times occur as $\tau_{j+1}^N = max(t_m^{(i)} | t_m^{(i)} > \tau_j^N, m = 1, ..., N_i \wedge i = 1, ..., N)$ where N_i is the number of trades for asset i. Clearly this means that the available refreshed prices will be a decreasing as the number of assets included increases, since the frequency of observed prices are determined by the least frequently traded asset[4]. A secondary issue is the Epps effect, the tendency for microstructure noise to cause a bias in realized covariances (RC) towards zero, see eg. Sheppard (2006).

We use the refresh sampled prices to calculate the (univariate) realized variance and realized kernel estimates at the portfolio level for use in eg. GARCH-X models, see section 4. This is a very simple approach to using high frequency data for portfolio risk forecasting. We are not aware of any other applications of this particular method.

Similar to how we use the realized variance for the univariate estimates we will for the multivariate estimates use RC . RC is estimated by the outer product of returns,

[4]Specifically, for the full portfolio of 44 stocks in section 6, we have $1,557$ daily prices on average.

$$\hat{\Sigma}_t^{RC} = \sum_{j=1}^{n} r_{j,n} r'_{j,n}.$$

similar to how realized variance is calculated. RC should be sampled as frequently as possible when there is no errors in the prices. However, to minimize the Epps effect while being mindful of the loss in observations we use prices sampled at most at 20 minute intervals[5].

In future work we plan to also employ the CholCov method of Boudt et al. (2014) to estimate the covariance matrix, Σ_t, of the portfolio at time t by a Cholesky decomposition of the covariance matrix

$$\hat{\Sigma}_t^{Cholcov} = H_t G_t H_t'.$$

This ensures a positive semidefinite estimate while a sequential estimation of the elements of H_t and G_t using an expanding number of asset returns, based on refreshed prices, uses as many observations as possible. The method is attractive because it ensures both a positive semidefinite matrix and is efficient in its usage of observations available.

4.2 Univariate Models

The GARCH-X model is an extension of the GARCH model of section 3.1 with an added exogenous variable in the form of a realized volatility measure (RM). Specifically the model is given by

$$\sigma_{p,t}^2 = \omega + \alpha y_{p,t-1}^2 + \beta \sigma_{p,t-1}^2 + \gamma RM_{t-1}.$$

This specification is identical to the expression for the variance in the HEAVY models of Noureldin et al. (2011) when setting $\alpha = 0$, but unlike the HEAVY models does not contain an expression for the realized measure which can be useful for multiple step forecasts. For completeness we also implement the HEAVY model, that is, the GARCH-X model with restriction $\alpha = 0$.

The Heterogeneous Autoregressive model (HAR) of Corsi (2009) is an additive cascade model of realized volatility from different frequencies (daily, weekly and monthly). This simple model can generate a number of features commonly seen in return data such as long memory and fat tails. The model is given by

$$\sigma_{p,t}^2 = \alpha + \beta RM_{t-1}^d + \gamma RM_{t-1}^w + \kappa RM_{t-1}^m,$$

where RM_{t-1}^d, RM_{t-1}^w and RM_{t-1}^m are the daily, weekly and monthly realized measures. The weekly and monthly realized volatility are estimated by using daily data for the past 5 and 30 days respectively but otherwise in a similar manner to daily realized volatility.

[5] Sheppard (2006) finds bias when sampling more frequently than 18 minutes in data observed between 1993 and 2002, as our dataset is more recent and the number of trades have increased rapidly over time we are confident that using a 20 minute sampling frequency avoids most negative microstructure effects.

4.3 Multivariate Models

Our first intraday multivariate model is a simple analog to the RM94 model of equation (3.4), but using either $\hat{\Sigma}_t^{RC}$ rather than the outer product of returns to update the covariance matrix. The model, which we will refer to as the RM94-RM model, is given as

$$H_t = (1 - \lambda)\,\hat{\Sigma}_{t-1}^{RC} + \lambda H_{t-1}$$

We let $\lambda = 0.25$, weighting scheme has been changed compared to the RM94 model to reflect the increased precision of using a realized measure. The weighting is somewhat arbitrary and is chosen largely to be intuitive. We are not aware of other papers using this setup to forecast the covariance matrix of a portfolio.

The HEAVY model of Noureldin et al. (2011), is similar to the variance targeting scalar BEKK model of equation (3.1), but with the difference that the estimated integrated covariance matrix is used rather than the outer product of returns so that the model is given by

$$H_t = (1 - A - B)\,\Gamma + A\hat{\Sigma}_{t-1}^{RC} + BH_{t-1}. \tag{4.1}$$

Bauwens et al. (2012) propose the realized consistent dynamic conditional correlation model (Re-cDCC), a DCC-like structure for realized covariance matrices. Specifically H_t is factorized as

$$H_t = D_t R_t D_t \tag{4.2}$$

and R_t as

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2},$$

but where the correlation driving process, Q_t, is defined as

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha P_{t-1}^{*} + \beta Q_{t-1},$$

where

$$P_{t-1}^{*} = diag(Q_t)^{1/2} D_t^{-1} \hat{\Sigma}_{t-1t} D_t^{-1} diag(Q_t)^{1/2}.$$

To reduce the number of estimated parameters we use correlation targeting by replacing $\overline{Q}$ with the mean of P_{t-1}^{*}. We use ARFIMA(1,d,0) models for the volatilities in D_t, estimated on the individual series. We estimate using a Gaussian likelihood.

The HEAVY and Re-cDCC models are also included in the VaR 1 day forecasting exercise found in Boudt et al.

(2014), where both models are estimated using a Wishart composite likelihood, see Shephard et al. (2008), and use the CholCov, see section 4.1, estimate for the realized measure.

5 Evaluating VaR Forecasts

In this section we describe two tools for comparing VaR forecasts, hit-sequence based backtests and the model confidence set approach.

Hit-sequence based backtesting is useful for evaluating whether a forecast successfully forecasted the VaR, however the early backtests commonly used in the literature are known to have low power, see Christoffersen (1998). We therefore use the backtests developed in Pajhede (2017) which in simulations have been found to have excellent power properties and are easy to interpret.

The model confidence set approach of Hansen et al. (2011) yields a set of models which contain the best model, in terms of minimizing some loss function when forecasting, with a specified degree of certainty. This approach is useful in comparing multiple competing models of different types.

5.1 Hit-Sequence Based Backtesting

Backtesting is the procedure of comparing realized losses to the forecasted VaR. To implement backtesting of a VaR forecast, we use the hit-sequence

$$I_t := 1\left(R_t < \text{VaR}_{t|t-1}(p)\right), \qquad t = 1, ...T \tag{5.1}$$

Where $1(\cdot)$ is the indicator function. Thus, the hit-sequence is by construction a binary time series indicating whether a loss at time t greater than the VaR, termed a violation or a hit, was realized.

A VaR forecast is valid, in the sense of actually having forecasted the desired quantile, only if the associated hit-sequence satisfies the following criteria due to Christoffersen (1998):

- The unconditional coverage criteria: The unconditional probability of a violation must be exactly equal to the coverage rate p:

$$H_{UC}: \; P(I_t = 1) = p$$

- The independence criteria: The conditional probability of a violation must be constant:

$$H_{Ind}: \; P(I_t = 1|\mathcal{F}_{t-1}) = P(I_t = 1)$$

Combining these criteria we obtain the conditional coverage criteria:

- The conditional coverage criteria: The probability of a violation must be constant and equal to the coverage rate:

$$H_{CC}:\ P(I_t = 1|\mathcal{F}_{t-1}) = P(I_t = 1) = p$$

It follows, see Christoffersen (1998), that the hit-sequence of a valid VaR forecast, is in fact a sequence of i.i.d. Bernoulli distributed variables:

$$I_t \underset{i.i.d.}{\sim} Bernoulli(p), \qquad t = 1, ..., T. \tag{5.2}$$

Pajhede (2017) generalizes the backtests of Kupiec (1995) and Christoffersen (1998) to test for unconditional coverage, independence and conditional coverage in a k'th order Markov chain model. The so called generalized Markov tests suggested in Pajhede (2017) have shown good power properties in simulation studies and are easy to apply. We implement the backtests using the Backtest toolbox[6].

$$\hat{p}_S = \frac{T_{01}}{T_{01} + T_{00}}, \qquad \hat{p}_E = \frac{T_{11}}{T_{11} + T_{10}} \qquad \text{and} \qquad \hat{\phi} = \frac{T_{01} + T_{11}}{T_{01} + T_{11} + T_{00} + T_{10}}$$

Where $J_{t-1} := 1\left(\sum_{i=1}^{k} I_{t-1} > 0\right)$ and the counts; $T_{11} := \sum_{t=1}^{T} I_t J_{t-1}$, $T_{01} := \sum_{t=1}^{T} I_t(1 - J_{t-1})$, $T_{10} := \sum_{t=1}^{T}(1 - I_t)J_{t-1}$, $T_{00} := \sum_{t=1}^{T}(1 - I_t)(1 - J_{t-1})$. We note that $\hat{p}_E$ can only be calculated if there is at least 1 hit in the hit-sequence.

- The test statistic of independence has the following expression

$$\begin{aligned} Q_{G-Ind} = -2\{&log(1-\hat{\phi})(T_{00}+T_{10}) + log(\hat{\phi})(T_{01}+T_{11}) \\ &- log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} - log(1-\hat{p}_E)T_{10} - log(\hat{p}_E)T_{11}\} \end{aligned} \tag{5.3}$$

- The test statistic of unconditional coverage has the following expression

$$Q_{G-UC} = -2\left\{log\left(p^{T_1}(1-p)^{T_0}\right) - log\left(\hat{\phi}^{T_1}(1-\hat{\phi})^{T_0}\right)\right\} \xrightarrow{d} \chi^2(1). \tag{5.4}$$

- The test statistic of conditional coverage has the following expression

$$\begin{aligned} Q_{G-CC} &= -2\{log(1-p)(T_{00}+T_{10}) + log(p)(T_{01}+T_{11}) \\ &\quad - log(1-\hat{p}_S)T_{00} - log(\hat{p}_S)T_{01} - log(1-\hat{p}_E)T_{10} - log(\hat{p}_E)T_{11}\} \\ &= Q_{UC} + Q_{G-Ind} \end{aligned} \tag{5.5}$$

The last expression highlights that a rejection of conditional coverage can easily be examined as to whether it was due to a rejection of UC, IND or both. We note that simulations in Pajhede (2017) suggest that $k = 10$ may be an appropriate choice of lags and therefore we use this for all tests.

We have that for $T \to \infty$, and under the null-hypothesis that $\{I_t\}$ is an i.i.d. Bernoulli sequence with probability parameter p, that

$$Q_{G-Ind}\chi^2(1), \quad Q_{G-UC} \xrightarrow{d} \chi^2(1) \quad \text{and} \quad Q_{G-CC} \xrightarrow{d} \chi^2(2).$$

[6] See Pajhede (2015) for details

However, as for all hit-sequence based backtests size distortion can be severe even for quite large samples, see Pajhede (2017), and so the Monte Carlo testing technique of Dufour (2006) is used to correct the p-values.

5.2 Model Confidence Set

The MCS approach of Hansen et al. (2011) can be used to select a set of models which - with arbitrary precision - contain the model with the best forecast. Specifically, the "best" model is defined as the one minimizing the mean loss using the tick loss function $L_t = \left(\text{VaR}_{t|t-1}(p) - Y_t\right)(I_t - \alpha)$ with sampling of a block bootstrap. The bootstrap is used to estimate a distribution of the mean loses and eliminate the models using t-statistics.

We select a 90% confidence set, that is, the best model in terms of minimizing the loss will with at least 90% certainty belong to the model confidence set. All calculations were done using the Oxford MFE toolbox of Kevin Shephard[7] with a block length of 2 and $10,000$ replications.

6 Empirical Application

In this section we perform an empirical application by using the models detailed in the previous sections to forecast the 1 day VaR, these forecasts are then compared by the methods covered in section 5. The test results are presented in an aggregate manner, showing which types of model have the strongest performance rather than singling out one model as superior. For the full results, see appendix A, B and C.

In the last subsection 6.4 we also consider the results of multi-period forecasting, specifically the 10 day VaR.

6.1 Data

Data was obtained from the TAQ database and consist of trading data for 44 of the largest stocks in the S&P500[8] as measured by market capitalization on the date of 01/01/2012. The data spans the period 03/01/2011 to 31/12/2015 with a total of 1258 observations. In the empirical application we use a rolling window of 500 observations for estimation, updated daily, with 1 day ahead forecasts for a total of 758 forecasts to be used for evaluating the accuracy of the forecasts. The results of Pajhede (2017) suggest that this is sufficient for good power properties of the VaR backtests. The data was generously supplied by Asger Lunde and has been cleaned using the step-by-step cleaning procedure described in Barndorff-Nielsen et al. (2009). RV and RK where estimated using code modified from the Oxford MFE toolbox of Kevin Shephard[9]

[7] https://www.kevinsheppard.com/MFE_Toolbox

[8] The tickers are: AAPL, ABT, AIG, AMZN, AXP, BAC, BMY, C, CAT, COP, CSCO, CVS, CVX, DIS, GE, GOOG, GS, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MO, MRK, MSFT, ORCL, OXY, PEP, PFE, PG, PM, QCOM, SLB, T, UNH, UPS, USB, UTX, V, VZ, WFC.

[9] https://www.kevinsheppard.com/MFE_Toolbox

6.2 Backtest Results

The backtest results are presented in tables 1, 2 and 3. To extract the main results of the data we organize the results by their rejection at a 5% significance level, using the conditional coverage backtest presented in section 5.1. We also sort the results by different categories such as univariate/multivariate, conditional distribution, realized measure used and model specification. In the following, we will refer to forecasts for which the conditional coverage test rejects the null of conditional coverage at a 5% significance level as a rejected forecast.

For the 1% VaR we see from table 1 that multivariate and univariate models have similar performance, with 4 out of 12 and 7 out of 28 being rejected respectively. Regarding the univariate models, we see that all 7 rejections are from the interday based models and none from using intraday, which also entails no apparent difference in using either RV or RK for forecasting the 1% VaR. For the multivariate models, based on intraday data, there is 1 rejection out of 5 forecasts and 3 rejections out of 7 using interday data. In terms of the conditionally specified distribution, of the 9 forecasts generated from a univariate model using FHS only 1 forecast is rejected. However, 4 out of 5 forecasts using multivariate models and the FHS method are rejected, suggesting FHS may be more appropriate in a univariate setting. 1 of the 2 CAViaR model based forecasts is rejected. The forecasts based on the simple EWMA1994, RM94 and RM94-RM models are not rejected.

For the 5% VaR we see from table 2 that there is a slight edge to the forecasts generated using multivariate models with 5 rejections out of 12 compared to the 16 rejections out of 28 from the univariate models. The intraday and interday based models now suggest a worse performance using intraday based models compared to interday based models with 10 rejections out of 16 forecasts and 6 rejections out of 12 forecasts respectively - With equal performance of forecasts based on models using RV and RK. There appears to be similar performance of intraday and interday based forecasts when using a multivariate model, with 3 out of 7 rejections for interday based forecasts and 2 out of 5 for intraday based forecasts. In terms of conditional distribution, we see that no forecast out of the 9 generated using a univariate model and FHS is rejected. However, for the multivariate models we again see quite poor performance using FHS in conjunction with a multivariate model with 4 rejected forecasts out of 5. The forecasts based on the simple EWMA1994, RM94 and RM94-RM models are not rejected.

For the 10% VaR we see from table 3 that there is a slight edge to the forecasts generated using univariate models with 11 rejections out of 28 compared to the 6 rejections out of 12 for the multivariate models. For the univariate models, we see that there is a slight advantage to using intraday models with 6 rejections out of 16 forecasts, compared to 5 rejected forecasts out of 12 for interday based forecasts. Again, RV and RK generate similar forecasts. Similar results are found for multivariate models with 2 rejections out of 5 for the intraday based forecasts and 4 out rejections of 7 for the interday based forecasts. In terms of the conditionally specified distribution, we again see that no univariate model using FHS generate a rejected forecast. Compared to the repeated poor performance when combining multivariate models with FHS, with 4 rejections out of 5 forecasts. The forecasts using CAViaR models are not rejected. The forecasts based on the simple EWMA1994, RM94 and RM94-RM models are not rejected.

Univariate models (Rejected = 7, Tested = 28)

Distribution	Rejected	Tested
Gaussian	2	10
t	3	7
FHS	1	9

Realized Measure	Rejected	Tested
RV	0	8
RK	0	8

Data frequency	Rejected	Tested
Interday	7	12
Intraday	0	16

Specification (intraday)	Rejected	Tested
GARCH-X	0	6
HEAVY	0	4
HAR	0	6

Specification (interday)	Rejected	Tested
GARCH	1	3
GARCH-GJR	3	3
APARCH	2	3
EWMA1994	0	1
CAViaR	1	2

Multivariate models (Rejected = 4, Tested = 12)

Distribution	Rejected	Tested
Gaussian	0	7
FHS	4	5

Data frequency	Rejected	Tested
Interday	3	7
Intraday	1	5

Specification (intraday)	Rejected	Tested
HEAVY	1	2
RM94-RM	0	2
RE-cDCC	0	1

Specification (interday)	Rejected	Tested
BEKK	1	2
CCC	1	2
DCC	1	2
RM94	0	1

Table 1: Backtest results for the 1% VaR grouped as univariate/multivariate and by distribution, data frequency and specification. The backtests were carried out using the generalized Markov test of conditional coverage due to Pajhede (2017) with $k = 10$ lags and using the bootstrap method of Dufour (2006). All calculations were done using the backtest toolbox of Pajhede (2017). The data series consists of daily data on 44 stocks from the S&P500 index obtained from the TAQ database covering the period 03/01/2011 to 31/12/2015. Full dataset is available from the authors upon request.

Univariate models (Rejected = 16, Tested = 28)

Distribution	Rejected	Tested
Gaussian	9	10
t	7	7
FHS		9

Realized Measure	Rejected	Tested
RV	5	8
RK	5	8

Data frequency	Rejected	Tested
Interday	6	12
Intraday	10	16

Specification (intraday)	Rejected	Tested
GARCH-X	4	6
HEAVY	2	4
HAR	4	6

Specification (interday)	Rejected	Tested
GARCH	2	3
GARCH-GJR	2	3
APARCH	2	3
EWMA1994	0	1
CAViaR	0	2

Multivariate models (Rejected = 5, Tested = 12)

Distribution	Rejected	Tested
Gaussian	2	7
FHS	4	5

Data frequency	Rejected	Tested
Interday	3	7
Intraday	2	5

Specification (intraday)	Rejected	Tested
HEAVY	2	2
RM94-RM	0	2
RE-cDCC	1	1

Specification (interday)	Rejected	Tested
BEKK	1	2
CCC	1	2
DCC	1	2
RM94	0	1

Table 2: Backtest results for the 5% VaR grouped as univariate/multivariate and by distribution, data frequency and specification. The backtests were carried out using the generalized Markov test of conditional coverage due to Pajhede (2017) with $k = 10$ lags and using the bootstrap method of Dufour (2006). All calculations were done using the backtest toolbox of Pajhede (2017). The data series consists of daily data on 44 stocks from the S&P500 index obtained from the TAQ database covering the period 03/01/2011 to 31/12/2015. Full dataset is available from the authors upon request.

6.3 Model Confidence Set Results

The MCS results are presented in detail in Appendix B. The results largely agree with the results using the backtests of the previous section.

For the 1% VaR only models using intraday data is included in the MCS, of these, 4 out of 7 are univariate and 3 multivariate. 3 out of 4 forecasts included are based on models using FHS.

For the 5% VaR, the MCS now contains 25 out of the 40 models. We note that all the models of the previous MCS for the 1% VaR is included in the MCS for the 5% VaR. The larger set can be interpreted either as that the data is not very informative about which model is best, or we could say that that all forecasts are close to optimal and are therefore indistinguishable leading to a larger set of models in the MCS. The MCS contains 17 forecasts based on univariate models and 8 based on multivariate - we note that as there is also 28 univariate forecasts compared to the 12 multivariate, this suggests similar performance of the two classes. However, of the 25 included models 20 use intraday data, which suggest the best model may lie in that category.

For the 10% VaR, the MCS now contains 27 out of the 40 models. We note that all the models of the previous MCS's is included in the MCS for the 10% VaR. The only addition is the two forecasts based on CAViaR models.

Univariate models (Rejected = 11, Tested = 28)

Distribution	Rejected	Tested
Gaussian	8	10
t	3	7
FHS	0	9

Realized Measure	Rejected	Tested
RV	3	8
RK	3	8

Data frequency	Rejected	Tested
Interday	5	12
Intraday	6	16

Specification (intraday)	Rejected	Tested
GARCH-X	2	6
HEAVY	2	4
HAR	2	6

Specification (interday)	Rejected	Tested
GARCH	2	3
GARCH-GJR	1	3
APARCH	2	3
EWMA1994	0	1
CAViaR	0	2

Multivariate models (Rejected = 6, Tested = 12)

Distribution	Rejected	Tested
Gaussian	2	7
FHS	4	5

Data frequency	Rejected	Tested
Interday	4	7
Intraday	2	5

Specification (intraday)	Rejected	Tested
HEAVY	2	2
RM94-RM	0	2
RE-cDCC	0	1

Specification (interday)	Rejected	Tested
BEKK	2	2
CCC	1	2
DCC	1	2
RM94	0	1

Table 3: Backtest results for the 10% VaR grouped as univariate/multivariate and by distribution, data frequency and specification. The backtests were carried out using the generalized Markov test of conditional coverage due to Pajhede (2017) with $k = 10$ lags and using the bootstrap method of Dufour (2006). All calculations were done using the backtest toolbox of Pajhede (2017). The data series consists of daily data on 44 stocks from the S&P500 index obtained from the TAQ database covering the period 03/01/2011 to 31/12/2015. Full dataset is available from the authors upon request.

6.4 Multi-period Forecasts

While 1 day VaR is perhaps the most commonly examined measure of risk in the academic literature, due to the ease of 1 period forecasts in most models, the required industry standard is a 10 day 1% VaR, see on Banking Supervision (1996). The regulatory framework set out by the 1996 Basel accords, see on Banking Supervision (1996) section B4.c, directly advocates forecasting a lower period and then scaling it using the well known square-root-of-time rule. In this section we examine the performance, using the backtests previously described, the 5% and 1% VaR forecasts of 10 days using the square-root-of-time rule.

Suppose we have the 1 day VaR, forecast as defined in equation (2.2), that is

$$\text{VaR}_{t|t-1}(\alpha) = \mu_p + \sigma_{p,t} F^{-1}(\alpha). \tag{6.1}$$

The square-root-of-time rule for forecasting an additional x periods is then implemented simply by calculating $\sigma_{p,t+x} = \sigma_{p,t}\sqrt{x}$, and using that as the volatility in equation (2.2)[10]. This approach has been somewhat criticized in the academic literature, see eg. Diebold et al. (1997) and Wang et al. (2011), because the implicit assumption of constant variance and no autocorrelations in the innovations are unlikely to be true and leads to an overestimation of risk. Kole et al. (2015) finds that while scaling performs adequate, and better than using direct forecasts, ie. using a model for biweekly data is worse than scaling a daily models forecast, the best performance is not-surprisingly found by using an iterated forecasting scheme, see Marcellino et al. (2005), where the forecast is iterated one step until the desired period is reached. Although this may be difficult in practice for some models.

We present the results of the backtests and model confidence set in appendix C. The results are quite clear in that every single forecast is either rejected or has so few violations that no test could be calculated to evaluate the forecast - which should also be considered a rejection. However, on closer inspection one can see that for all VaR coverage levels several forecasts actually pass the UC criteria. That is, scaling by the square-root-of-time can produce acceptable forecasts in terms of unconditional coverage, but will fail the IND criteria. We find no particular evidence that using the square-root-of-time rule should lead one to overestimate the levels of risk as argued in eg. Diebold et al. (1997).

7 Conclusion

The purpose of this paper has been to draw broad generalizations about what works best when forecasting VaR for large portfolios. In this respect we have noted certain consistencies in our results in the many models examined and at different VaR coverage levels: the choice of a univariate or a multivariate model seem to not be particularly important, both types of model can generate good forecasts and neither is consistently better than the other. Using intraday data appears to produce superior forecasts, especially in univariate models. There appears to be no difference in the accuracy of forecasts produced using realized variance based on a 10 minute sampling frequency compared to using a

[10]When the forecasting method does not depend on a volatility estimate we instead scale the VaR directly using the square root of time.

realized kernel. In terms of conditional distributions, filtered historical simulation in conjunction with univariate models work excellent, producing consistently good forecasts. However, for multivariate models filtered historical simulation performs poorly compared to using a Gaussian distribution. We also find that simple fixed parameter models, the EWMA, RM94 and the RM94-RM, all have excellent forecasting performance across different VaR levels. Likewise the two CAViaR models have good performance, with only a single rejection at the 1% VaR level.

Lastly, we find that using the square-root-of-time rule is consistently rejected by backtests, especially due to the independence criteria. That is, using the square-root-of-time rule one will produce forecasts which may not account for changing levels of risk. However, the unconditional coverage criteria is only rejected for about half of the examined models and we find no particular evidence that using the rule should lead one to overestimate levels of risk.

References

Andersen, T. G. and L. Benzoni (2008). Realized volatility. Technical report.

Andersen, T. G. and T. Bollerslev (1998, November). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review 39*(4), 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003, March). Modeling and Forecasting Realized Volatility. *Econometrica 71*(2), 579–625.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Technical report.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009, November). Realized kernels in practice: trades and quotes. *Econometrics Journal 12*(3), C1–C32.

Barone-Adesi, G. and K. Giannopoulos (2001). Non parametric var techniques. myths and realities. *Economic Notes 30*(2), 167–181.

Bauwens, L., G. Storti, and F. Violante (2012, December). Dynamic conditional correlation models for realized covariance matrices. CORE Discussion Papers 2012060, Universite catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Berkowitz, J. and J. M. O'Brien (2001). How accurate are Value-at-Risk models at commercial banks? Finance and Economics Discussion Series 2001-31, Board of Governors of the Federal Reserve System (U.S.).

Bollerslev, T. (1986, April). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (1990, August). Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *The Review of Economics and Statistics 72*(3), 498–505.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988, February). A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy 96*(1), 116–31.

Boudt, K., S. Laurent, A. Lunde, and R. Quaedvlieg (2014, 02). Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity. CREATES Research Papers 2014-05, Department of Economics and Business Economics, Aarhus University.

Braione, M. and N. K. Scholtes (2014, November). Construction of value-at-risk forecasts under different distributional assumptions within a BEKK framework. CORE Discussion Papers 2014059, Universite catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–62.

Clements, M. P., A. B. GalvÃ£o, and J. H. Kim (2006). Quantile Forecasts of Daily Exchange Rate Returns from Forecasts of Realized Volatility. The Warwick Economics Research Paper Series (TWERPS) 777, University of Warwick, Department of Economics.

Corsi, F. (2009, Spring). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Creal, D., S. J. Koopman, and A. Lucas (2013, 08). Generalized Autoregressive Score Models With Applications. *Journal of Applied Econometrics 28*(5), 777–795.

Diebold, F. X., A. Hickman, A. Inoue, and T. Schuermann (1997, July). Converting 1-Day Volatility to h-Day Volatitlity: Scaling by Root-h is Worse Than You Think. Center for Financial Institutions Working Papers 97-34, Wharton School Center for Financial Institutions, University of Pennsylvania.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993, June). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics 133*(2), 443–477.

Engle, R. (2002, July). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics 20*(3), 339–50.

Engle, R. F. (1982, July). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. and K. F. Kroner (1995, February). Multivariate Simultaneous Generalized ARCH. *Econometric Theory 11*(01), 122–150.

Engle, R. F. and S. Manganelli (2004, October). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics 22*, 367–381.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993, December). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance 48*(5), 1779–1801.

Hansen, P. R. and A. Lunde (2006, April). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics 24*, 127–161.

Hansen, P. R., A. Lunde, and J. M. Nason (2011, 03). The Model Confidence Set. *Econometrica 79*(2), 453–497.

Kole, E., T. Markwat, A. Opschoor, and D. van Dijk (2015, January). Forecasting Value-at-Risk under Temporal and Portfolio Aggregation. Tinbergen Institute Discussion Papers 15-140/III, Tinbergen Institute.

Koopman, S. J., B. Jungbacker, and E. Hol (2005, June). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance 12*(3), 445–475.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. Finance and Economics Discussion Series 95-24, Board of Governors of the Federal Reserve System (U.S.).

Luc Bauwens, S. L. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics 23*(3), 346–354.

Lunde, A. and P. R. Hansen (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics 20*(7), 873–889.

Marcellino, M., J. Stock, and M. Watson (2005). A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. Technical report.

McAleer, M. and B. da Veiga (2008). Single-index and portfolio models for forecasting value-at-risk thresholds. *Journal of Forecasting 27*(3), 217–235.

Merton, R. C. (1980, February). On Estimating the Expected Return on the Market: An Exploratory Investigation. NBER Working Papers 0444, National Bureau of Economic Research, Inc.

Noureldin, D., N. Shephard, and K. Sheppard (2011, February). Multivariate High-Frequency-Based Volatility (HEAVY) Models. Economics Papers 2011-W01, Economics Group, Nuffield College, University of Oxford.

on Banking Supervision, B. C. (1996, January). AMENDMENT TO THE CAPITAL ACCORD TO INCORPORATE MARKET RISKS. Technical report, Basle Committee on Banking Supervision.

Pajhede, T. (2017). Backtesting value-at-risk: The generalized markov test. Forthcoming in journal of forecasting.

Pedersen, R. S. and A. Rahbek (2014). Multivariate Variance Targeting in the BEKK-GARCH Model. *Econometrics Journal 17*(01), 24–55.

Persand, G. and C. Brooks (2003). Volatility forecasting for risk management. *Journal of Forecasting 22*(1), 1–22.

Rabemananjara, R. and J. M. Zakoian (1993, Jan.-Marc). Threshold Arch Models and Asymmetries in Volatility. *Journal of Applied Econometrics 8*(1), 31–49.

Roll, R. (1984, September). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance 39*(4), 1127–39.

Santos, A. A. P., F. J. Nogales, and E. Ruiz (2013, March). Comparing Univariate and Multivariate Models to Forecast Portfolio Value-at-Risk. *Journal of Financial Econometrics 11*(2), 400–441.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*(2), 197–231.

Shephard, N., K. Sheppard, and R. F. Engle (2008, September). Fitting vast dimensional time-varying covariance models. Economics Series Working Papers 403, University of Oxford, Department of Economics.

Sheppard, K. (2006, January). Realized Covariance and Scrambling. Working paper, Department of Economics, University of Oxford.

Wang, J.-N., J.-H. Yeh, and N. Y.-P. Cheng (2011, May). How accurate is the square-root-of-time rule in scaling tail risk: A global study. *Journal of Banking & Finance 35*(5), 1158–1169.

A Backtest Results

Model	Coverage PCT	UC	IND	CC
APARCHFHS	0.53	0.15	0.71	0.10
APARCH	0.40	0.04	0.82	0.04
APARCHt	0.13	0.00	0.93	0.01
BEKKFHS	0.13	0.00	0.93	0.01
BEKK	1.06	0.94	0.86	0.95
CAViaR-Symmetric	0.26	0.01	0.82	0.03
CAViaR-Asymmetric	0.79	0.52	0.02	0.09
CCCFHS	0.13	0.00	1.00	0.00
CCC	0.79	0.53	0.46	0.46
DCCFHS	0.13	0.00	1.00	0.00
DCC	0.79	0.57	0.47	0.47
EWMA1994	0.79	0.52	0.47	0.45
GARCHFHS	1.06	0.87	0.24	0.57
GARCHXFHSRK	0.79	0.51	0.47	0.46
GARCHXFHSRV	0.66	0.36	0.54	0.48
GARCHXRK	0.79	0.56	0.48	0.46
GARCHXRV	0.40	0.04	0.73	0.14
GARCHXtRK	0.53	0.20	0.66	0.27
GARCHXtRV	0.40	0.06	0.75	0.11
GARCH	0.66	0.37	0.62	0.26
GARCHt	0.26	0.01	0.84	0.03
GJRGARCHFHS	1.19	0.68	0.00	0.00
GJRGARCH	0.53	0.16	0.05	0.02
GJRGARCHt	0.13	0.00	0.93	0.01
HARFHSRK	0.66	0.30	0.62	0.26
HARFHSRV	0.66	0.31	0.51	0.52
HARRK	0.79	0.51	0.47	0.46
HARRV	0.53	0.16	0.73	0.11
HARtRK	0.66	0.28	0.62	0.26
HARtRV	0.40	0.05	0.74	0.12
HEAVYRCFHS	2.64	0.00	0.01	0.00
HEAVYRC	1.06	0.91	0.19	0.40
HEAVYUnivariateRKFHS	0.79	0.54	0.47	0.47
HEAVYUnivariateRK	0.79	0.56	0.47	0.46
HEAVYUnivariateRVFHS	0.53	0.16	0.63	0.31
HEAVYUnivariateRV	0.53	0.18	0.66	0.29
RM94RCFHS	1.58	0.12	0.94	0.31
RM94RC	1.06	0.95	0.89	0.97
RM94	0.79	0.54	0.47	0.45
ReDCCRC	0.53	0.15	0.67	0.26

Table 4: Backtest results for the 1 day 1% VaR. UC , IND and CC indicates the p-values for generalized Markov backtests of their respective hypotheses as defined in section 5.1 and using $k = 10$ lags. A blank value indicates that the test could not be calculated because there was no hits, which should be interpreted as a rejection of UC. RK and RV refer to models using realized kernels and realized volatility respectively. FHS and t refers to models using filtered historical simulation and the t distribution for forecasting, models without FHS or t use the Gaussian distribution. See sections 3 and 4 for further details.

Model	Coverage PCT	UC	IND	CC
APARCHFHS	4.49	0.53	0.07	0.13
APARCH	2.77	0.00	0.14	0.00
APARCHt	1.58	0.00	0.20	0.00
BEKKFHS	1.85	0.00	0.99	0.00
BEKK	3.96	0.17	0.11	0.10
CAViaR-Symmetric	3.69	0.09	0.84	0.13
CAViaR-Asymmetric	5.94	0.27	0.05	0.08
CCCFHS	1.06	0.00	0.31	0.00
CCC	3.43	0.03	0.53	0.07
DCCFHS	1.06	0.00	0.31	0.00
DCC	3.43	0.03	0.52	0.07
EWMA1994	4.35	0.38	0.81	0.53
GARCHFHS	4.49	0.55	0.86	0.65
GARCHXFHSRK	5.28	0.70	0.09	0.24
GARCHXFHSRV	5.01	0.95	0.90	0.99
GARCHXRK	3.03	0.01	0.33	0.01
GARCHXRV	2.64	0.00	0.14	0.00
GARCHXtRK	3.43	0.04	0.22	0.04
GARCHXtRV	2.90	0.00	0.21	0.01
GARCH	2.64	0.00	0.88	0.00
GARCHt	1.58	0.00	0.06	0.00
GJRGARCHFHS	5.01	0.95	0.23	0.47
GJRGARCH	3.56	0.06	0.23	0.05
GJRGARCHt	2.11	0.00	0.02	0.00
HARFHSRK	5.28	0.68	0.27	0.51
HARFHSRV	5.28	0.69	0.86	0.95
HARRK	2.90	0.01	0.17	0.00
HARRV	2.77	0.00	0.26	0.00
HARtRK	3.43	0.03	0.22	0.05
HARtRV	3.03	0.01	0.35	0.02
HEAVYRCFHS	4.62	0.66	0.00	0.01
HEAVYRC	3.83	0.13	0.06	0.04
HEAVYUnivariateRKFHS	5.28	0.73	0.19	0.40
HEAVYUnivariateRK	2.90	0.01	0.48	0.01
HEAVYUnivariateRVFHS	5.01	0.94	0.75	0.94
HEAVYUnivariateRV	2.64	0.00	0.14	0.00
RM94RCFHS	3.96	0.18	0.44	0.25
RM94RC	5.01	0.94	0.91	0.99
RM94	4.35	0.41	0.81	0.53
ReDCCRC	2.90	0.00	0.60	0.01

Table 5: Backtest results for the 1 day 5% VaR. The second clumn, Coverage PCT, indicate the mean of the hit-sequences in percent. UC , IND and CC indicates the p-values for generalized Markov backtests of their respective hypotheses as defined in section 5.1 and using $k = 10$ lags. A blank value indicates that the test could not be calculated because there was no hits, which should be interpreted as a rejection of UC. RK and RV refer to models using realized kernels and realized volatility respectively. FHS and t refers to models using filtered historical simulation and the t distribution for forecasting, models without FHS or t use the Gaussian distribution. See sections 3 and 4 for further details.

Model	Coverage PCT	UC	IND	CC
APARCHFHS	10.03	0.97	0.27	0.56
APARCH	7.52	0.01	0.85	0.04
APARCHt	6.60	0.00	0.95	0.00
BEKKFHS	3.83	0.00	0.24	0.00
BEKK	6.86	0.00	0.00	0.00
CAViaR-Symmetric	9.50	0.64	0.81	0.81
CAViaR-Asymmetric	11.48	0.19	0.94	0.48
CCCFHS	3.96	0.00	0.78	0.00
CCC	7.92	0.05	0.40	0.08
DCCFHS	3.96	0.00	0.79	0.00
DCC	7.92	0.05	0.40	0.08
EWMA1994	8.18	0.08	0.56	0.15
GARCHFHS	9.37	0.56	0.26	0.41
GARCHXFHSRK	9.89	0.93	0.78	0.94
GARCHXFHSRV	9.76	0.84	0.78	0.94
GARCHXRK	6.99	0.00	0.38	0.01
GARCHXRV	6.99	0.00	0.77	0.01
GARCHXtRK	8.05	0.06	0.91	0.14
GARCHXtRV	8.05	0.06	0.79	0.17
GARCH	7.78	0.04	0.30	0.05
GARCHt	5.28	0.00	0.88	0.00
GJRGARCHFHS	10.69	0.52	0.82	0.91
GJRGARCH	7.78	0.03	0.78	0.07
GJRGARCHt	5.67	0.00	0.29	0.00
HARFHSRK	10.16	0.89	0.67	0.91
HARFHSRV	10.16	0.86	0.87	0.98
HARRK	7.12	0.00	0.64	0.02
HARRV	6.86	0.00	0.66	0.01
HARtRK	7.92	0.05	1.00	0.11
HARtRV	7.92	0.06	0.89	0.14
HEAVYRCFHS	8.44	0.15	0.01	0.01
HEAVYRC	7.12	0.00	0.00	0.00
HEAVYUnivariateRKFHS	9.89	0.92	0.74	0.92
HEAVYUnivariateRK	7.12	0.01	0.49	0.02
HEAVYUnivariateRVFHS	9.89	0.92	0.73	0.93
HEAVYUnivariateRV	6.99	0.00	0.78	0.01
RM94RCFHS	8.05	0.07	0.73	0.10
RM94RC	9.37	0.55	0.77	0.68
RM94	8.18	0.09	0.56	0.15
ReDCCRC	8.44	0.14	0.73	0.27

Table 6: Backtest results for the 1 day 10% VaR. The second clumn, Coverage PCT, indicate the mean of the hit-sequences in percent. UC , IND and CC indicates the p-values for generalized Markov backtests of their respective hypotheses as defined in section 5.1 and using $k = 10$ lags. A blank value indicates that the test could not be calculated because there was no hits, which should be interpreted as a rejection of UC. RK and RV refer to models using realized kernels and realized volatility respectively. FHS and t refers to models using filtered historical simulation and the t distribution for forecasting, models without FHS or t use the Gaussian distribution. See sections 3 and 4 for further details.

B Model Confidence Set Results

Model	1%	5%	10%
APARCHFHS	No	No	No
APARCH	No	No	No
APARCHt	No	No	No
BEKKFHS	No	No	No
BEKK	No	No	No
CAViaR-Symmetric	No	No	Yes
CAViaR-Asymmetric	No	No	Yes
CCCFHS	No	No	No
CCC	No	No	No
DCCFHS	No	No	No
DCC	No	No	No
EWMA1994	No	Yes	Yes
GARCHFHS	No	Yes	Yes
GARCHXFHSRK	No	Yes	Yes
GARCHXFHSRV	Yes	Yes	Yes
GARCHXRK	No	Yes	Yes
GARCHXRV	No	Yes	Yes
GARCHXtRK	No	Yes	Yes
GARCHXtRV	No	Yes	Yes
GARCH	No	Yes	Yes
GARCHt	No	No	No
GJRGARCHFHS	No	Yes	Yes
GJRGARCH	No	Yes	Yes
GJRGARCHt	No	No	No
HARFHSRK	No	Yes	Yes
HARFHSRV	Yes	Yes	Yes
HARRK	Yes	Yes	Yes
HARRV	Yes	Yes	Yes
HARtRK	No	Yes	Yes
HARtRV	No	Yes	Yes
HEAVYRCFHS	No	No	No
HEAVYRC	No	No	No
HEAVYUnivariateRKFHS	No	Yes	Yes
HEAVYUnivariateRK	No	Yes	Yes
HEAVYUnivariateRVFHS	No	Yes	Yes
HEAVYUnivariateRV	No	Yes	Yes
RM94RCFHS	Yes	Yes	Yes
RM94RC	Yes	Yes	Yes
RM94	No	Yes	Yes
ReDCCRC	Yes	Yes	Yes

Table 7: Model confidence results for the 1 day 1%, 5% and 10% VaR. We use 90% confidence sets. All calculations were done using the Oxford MFE toolbox with a block length of 2 and 10, 000 replications.

C Backtest Results for Multi period Forecasts

Model	Coverage PCT	UC	IND	CC
APARCHFHS	0.67	0.31	0.00	0.00
APARCH	0.53	0.18	0.00	0.00
APARCHt	0.13	0.00	0.93	0.01
BEKKFHS	0.00	0.00		
BEKK	0.27	0.02	0.01	0.00
CAViaR-Symmetric	0.00	0.00		
CAViaR-Asymmetric	0.67	0.31	0.00	0.00
CCCFHS	0.00	0.00		
CCC	0.13	0.00	0.92	0.01
DCCFHS	0.00	0.00		
DCC	0.13	0.00	0.93	0.01
EWMA1994	0.53	0.16	0.00	0.00
GARCHFHS	0.40	0.08	0.00	0.00
GARCHXFHSRK	0.00	0.00		
GARCHXFHSRV	0.00	0.00		
GARCHXRK	0.00	0.00		
GARCHXRV	0.00	0.00		
GARCHXtRK	0.00	0.00		
GARCHXtRV	0.00	0.00		
GARCH	0.13	0.00	0.92	0.01
GARCHt	0.00	0.00		
GJRGARCHFHS	0.93	0.99	0.00	0.00
GJRGARCH	0.53	0.20	0.00	0.00
GJRGARCHt	0.27	0.03	0.01	0.00
HARFHSRK	0.00	0.00		
HARFHSRV	0.00	0.00		
HARRK	0.00	0.00		
HARRV	0.00	0.00		
HARtRK	0.00	0.00		
HARtRV	0.00	0.00		
HEAVYRCFHS	0.00	0.00		
HEAVYRC	0.27	0.02	0.00	0.00
HEAVYUnivariateRKFHS	0.00	0.00		
HEAVYUnivariateRK	0.00	0.00		
HEAVYUnivariateRVFHS	0.00	0.00		
HEAVYUnivariateRV	0.00	0.00		
RM94RCFHS	1.74	0.08	0.00	0.00
RM94RC	1.47	0.26	0.00	0.00
RM94	0.53	0.17	0.00	0.00
ReDCCRC	0.00	0.00		

Table 8: Backtest results for the 10 day 1% VaR. The second clumn, Coverage PCT, indicate the mean of the hit-sequences in percent. UC , IND and CC indicates the p-values for generalized Markov backtests of their respective hypotheses as defined in section 5.1 and using $k = 10$ lags. A blank value indicates that the test could not be calculated because there was no hits, which should be interpreted as a rejection of UC. RK and RV refer to models using realized kernels and realized volatility respectively. FHS and t refers to models using filtered historical simulation and the t distribution for forecasting, models without FHS or t use the Gaussian distribution. See sections 3 and 4 for further details.

Model	Coverage PCT	UC	IND	CC
APARCHFHS	4.94	0.94	0.00	0.00
APARCH	3.47	0.04	0.00	0.00
APARCHt	2.54	0.00	0.00	0.00
BEKKFHS	0.93	0.00	0.00	0.00
BEKK	3.07	0.01	0.00	0.00
CAViaR-Symmetric	3.07	0.01	0.00	0.00
CAViaR-Asymmetric	6.28	0.13	0.00	0.00
CCCFHS	0.67	0.00	0.16	0.00
CCC	4.14	0.26	0.00	0.00
DCCFHS	0.66	0.00	0.15	0.00
DCC	3.96	0.16	0.00	0.00
EWMA1994	3.34	0.03	0.00	0.00
GARCHFHS	5.07	0.92	0.00	0.00
GARCHXFHSRK	4.94	0.97	0.00	0.00
GARCHXFHSRV	4.94	0.96	0.00	0.00
GARCHXRK	3.20	0.02	0.00	0.00
GARCHXRV	2.94	0.01	0.00	0.00
GARCHXtRK	3.74	0.12	0.00	0.00
GARCHXtRV	3.34	0.02	0.00	0.00
GARCH	3.07	0.01	0.00	0.00
GARCHt	1.20	0.00	0.00	0.00
GJRGARCHFHS	5.07	0.89	0.00	0.00
GJRGARCH	3.60	0.07	0.00	0.00
GJRGARCHt	2.00	0.00	0.00	0.00
HARFHSRK	4.81	0.81	0.00	0.00
HARFHSRV	4.54	0.54	0.00	0.00
HARRK	2.94	0.00	0.00	0.00
HARRV	3.07	0.01	0.00	0.00
HARtRK	3.60	0.07	0.00	0.00
HARtRV	3.47	0.04	0.00	0.00
HEAVYRCFHS	0.00	0.00		
HEAVYRC	3.07	0.01	0.00	0.00
HEAVYUnivariateRKFHS	4.94	0.98	0.00	0.00
HEAVYUnivariateRK	3.34	0.03	0.00	0.00
HEAVYUnivariateRVFHS	4.81	0.83	0.00	0.00
HEAVYUnivariateRV	2.94	0.01	0.00	0.00
RM94RCFHS	4.27	0.36	0.00	0.00
RM94RC	4.94	0.99	0.00	0.00
RM94	3.20	0.02	0.00	0.00
ReDCCRC	2.54	0.00	0.00	0.00

Table 9: Backtest results for the 10 day 5% VaR. The second clumn, Coverage PCT, indicate the mean of the hit-sequences in percent. UC , IND and CC indicates the p-values for generalized Markov backtests of their respective hypotheses as defined in section 5.1 and using $k = 10$ lags. A blank value indicates that the test could not be calculated because there was no hits, which should be interpreted as a rejection of UC. RK and RV refer to models using realized kernels and realized volatility respectively. FHS and t refers to models using filtered historical simulation and the t distribution for forecasting, models without FHS or t use the Gaussian distribution. See sections 3 and 4 for further details.

Model	Coverage PCT	UC	IND	CC
APARCHFHS	11.21	0.29	0.00	0.00
APARCH	7.88	0.05	0.00	0.00
APARCHt	6.14	0.00	0.00	0.00
BEKKFHS	4.01	0.00	0.00	0.00
BEKK	7.48	0.02	0.00	0.00
CAViaR-Symmetric	8.68	0.21	0.00	0.00
CAViaR-Asymmetric	10.55	0.59	0.00	0.00
CCCFHS	4.81	0.00	0.00	0.00
CCC	8.54	0.18	0.00	0.00
DCCFHS	4.75	0.00	0.00	0.00
DCC	8.44	0.14	0.00	0.00
EWMA1994	8.14	0.08	0.00	0.00
GARCHFHS	10.28	0.78	0.00	0.00
GARCHXFHSRK	11.62	0.15	0.00	0.00
GARCHXFHSRV	10.81	0.43	0.00	0.00
GARCHXRK	8.01	0.06	0.00	0.00
GARCHXRV	8.14	0.09	0.00	0.00
GARCHXtRK	9.61	0.76	0.00	0.00
GARCHXtRV	9.75	0.85	0.00	0.00
GARCH	7.61	0.02	0.00	0.00
GARCHt	5.87	0.00	0.00	0.00
GJRGARCHFHS	11.21	0.28	0.00	0.00
GJRGARCH	8.14	0.08	0.00	0.00
GJRGARCHt	6.28	0.00	0.00	0.00
HARFHSRK	11.08	0.35	0.00	0.00
HARFHSRV	11.08	0.34	0.00	0.00
HARRK	8.41	0.13	0.00	0.00
HARRV	8.01	0.06	0.00	0.00
HARtRK	10.01	0.97	0.00	0.00
HARtRV	10.01	0.99	0.00	0.00
HEAVYRCFHS	0.00	0.00		
HEAVYRC	7.61	0.02	0.00	0.00
HEAVYUnivariateRKFHS	11.62	0.15	0.00	0.00
HEAVYUnivariateRK	8.28	0.11	0.00	0.00
HEAVYUnivariateRVFHS	11.21	0.29	0.00	0.00
HEAVYUnivariateRV	8.01	0.07	0.00	0.00
RM94RCFHS	9.21	0.48	0.00	0.00
RM94RC	12.55	0.03	0.00	0.00
RM94	8.14	0.08	0.00	0.00
ReDCCRC	8.28	0.11	0.00	0.00

Table 10: Backtest results for the 10 day 10% VaR. The second clumn, Coverage PCT, indicate the mean of the hit-sequences in percent. UC , IND and CC indicates the p-values for generalized Markov backtests of their respective hypotheses as defined in section 5.1 and using $k = 10$ lags. A blank value indicates that the test could not be calculated because there was no hits, which should be interpreted as a rejection of UC. RK and RV refer to models using realized kernels and realized volatility respectively. FHS and t refers to models using filtered historical simulation and the t distribution for forecasting, models without FHS or t use the Gaussian distribution. See sections 3 and 4 for further details.

A Conditionally Beta Distributed Time-Series Model With Application to Monthly US Corporate Default Rates*

Thor Pajhede†

January 31, 2017

Abstract

We consider an observation driven, conditionally beta distributed model for variables restricted to the unit interval. The model includes both explanatory variables and autoregressive dependence in the mean and precision parameters using the mean-precision parametrization of the beta distribution suggested by Ferrari and Cribari-Neto (2004). Our model is a generalization of the βARMA model proposed in Rocha and Cribari-Neto (2009), which we generalize to allow for covariates and an ARCH type structure in the precision parameter. We also highlight some errors in their derivations of the score and information which has implications for the asymptotic theory. Included simulations suggest that standard asymptotics for estimators and test statistics apply. In an empirical application to Moody's monthly US 12-month issuer default rates in the period $1972-2015$, we revisit the results of Agosto et al. (2016) in examining the conditional independence hypothesis of Lando and Nielsen (2010). Empirically we find that; (1) the current default rate influences the default rate of the following periods even when conditioning on explanatory variables. (2) The 12 month lag is highly significant in explaining the monthly default rate. (3) There is evidence for volatility clustering in the default rate data.

Keywords: Beta regression, credit risk, default rates, contagion.

JEL codes: C12, C50, C32, C22.

*The author acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. We thank Anders Rahbek, Heino Bohn Nielsen, Lena Hasselberg, Silvana Acosta, Rasmus Søndergaard and Jurgen Doornik for helpful and insightful comments.

†Department of Economics, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K, Denmark. CREATES, University of Aarhus. E-mail: thor.nielsen@econ.ku.dk

1 Introduction

Since the recent financial crisis there has been a strong interest in improving our understanding of corporate defaults. A focus of this interest is whether the clustering in defaults commonly observed, is mainly caused by defaults increasing the probability of default in other firms (the contagion hypothesis), or whether these clusters are due to common risk factors, specifically business cycle and financial, affecting all companies (the conditional independence or systematic risk hypothesis). This question has already been explored by several authors; see, for example, Das et al. (2006), Lando and Nielsen (2010), and Agosto et al. (2016). For investors and regulatory authorities the systemic components of credit portfolios are of interest to ensure financial stability of either a portfolios return or the economy as a whole. While from an academic standpoint it is interesting because assuming conditional independence can be useful to assume in derivations, see Lando and Nielsen (2010).

We propose a conditionally beta distributed time series model (CBTS), which is a generalization of the βARMA model of Rocha and Cribari-Neto (2009). The CBTS allows for covariates and autoregressive dependence in both the mean and precision parameters using the parametrization of the beta distribution suggested by Ferrari and Cribari-Neto (2004). The use of a conditional beta distribution for the default rate allows one to examine the impact on both the location and scale of the distribution, whereas the Poison distribution has only one parameter to match both the mean and the variance, the beta distribution has two.

However, similar to the GARCH-X type of models, see Han and Kristensen (2014), as shown in section 3 inference is quite involved. Section 4 presents a simulation study which suggests that the maximum likelihood estimator is asymptotically Gaussian and that likelihood ratio tests are asymptotically χ^2 distributed under the null.

We apply our model to Moody's monthly US 12-month issuer default rates in the period $1973 - 2015$. The specification for the mean and precision include macroeconomic and financial variables, intended to capture the common or correlated risk factors faced by the companies. We find that while explanatory variables do explain some of the time variation in the default rate, dependence remains in the mean and the precision parameters, possibly implying contagion effects. Further, we find evidence of volatility clustering in the default rate, which we believe to be a phenomenon not previously observed in default rates. We also find that the 12 month lag is highly significant in explaining default rates which appears to be a new result when modeling aggregate defaults and might indicate previously unknown seasonality. We also find that realized volatility which was found to be highly significant in explaining corporate defaults by Agosto et al. (2016) is not significant for the mean if dummies are included for October of 1987, September 2008 and October 2008, but might be for the precision parameter.

Previously, Sean et al. (1999) applied a Poisson model to default counts as a way to forecast the default rate, since the number of companies that can default is known 12 months in advance. Similarly, Agosto et al. (2016) examine the contagion hypothesis by modeling default counts.

However, using the default rate rather rather than the default count can avoid certain drawbacks of count models. Specifically, as the numbers of companies monitored that are capable of defaulting, known as the exposure for Poison models, see Cameron and Trivedi (2013), is not constant, this may create spurious dependence in the default counts. Regardless of whether the default of a company increases the probability of additional defaults, i.e. potentially presenting misleading evidence in favor of the contagion hypothesis. Instead, By dividing the number of defaults with the number of firms, i.e. using the default rate, this is handled in a straightforward way - but restricts the variable to be modeled to the unit interval.

A possible solution therefore is to apply a regression after having log-transformed the default rates as done in, for example, Giesecke et al. (2010). However, transformed values of proportions and rates often exhibit problematic characteristics, see Ferrari and Cribari-Neto (2004). Further, the interest lies in the default rate, not a logarithmic transformation of it - therefore it seems logical to model the default rate directly.

The paper is organized as follows. Section 2 introduces the CBTS model. Section 3 considers some derivations in the model, we highlight certain difficulties related to inference in both our model and the model of Rocha and Cribari-Neto (2009). Section 4 conducts a simulation study to evaluate the finite sample accuracy of the derived asymptotics for the maximum likelihood (ML) estimator as well as the empirical size for the likelihood ratio (LR) tests when using its asymptotic distribution. Section 5 is an empirical application of the model to Moody's monthly US 12-month issuer default rates in the period $1973-2015$, we consider the impact of covariates and discuss contagion effects. Section 6 concludes.

2 The Conditionally Beta Time Series (CBTS) Model

The beta distribution is a continuous distribution on the unit interval governed by two shape parameters and is widely used to model variables restricted to the unit interval, e.g. rates and proportions. The probability density function (PDF) of the beta distribution is given by

$$f(y) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \qquad 0 \leq y \leq 1,$$

where $p > 0$, $q > 0$ and $\Gamma(\cdot)$ is the gamma function. The shape of the PDF is highly flexible, allowing for a U, bell or J (with right or left tail) shaped curve as well as nesting the uniform distribution, see Ferrari and Cribari-Neto (2004) for figures of possible shapes. The mean and variance for a beta distributed random variable, y, is given by

$$E(y) = \frac{p}{p+q} \qquad and \qquad Var(y) = \frac{pq}{(p+q)^2(p+q+1)}$$

Following Ferrari and Cribari-Neto (2004) the distribution is reparametrized by setting $p = \mu\phi$ and $q = (1-\mu)\phi$ such that

$$E(y) = \mu \qquad and \qquad Var(y) = \frac{\mu(1-\mu)}{1+\phi} \tag{2.1}$$

where $\mu = p/(p+q)$ and $\phi = p+q$; here $0 < \mu < 1$ and $\phi > 0$ where ϕ can be regarded as a *precision* parameter since a larger ϕ forces a smaller $Var(y)$ for a fixed μ. We denote this as the $Beta(\mu, \phi)$ distribution with density given by

$$f(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \qquad 0 \leq y \leq 1.$$

Let $y = (y_1, ..., y_T)'$, be a time series whose distribution we model as a function of its own past, y_{t-m}, $m \geq 1$, and in terms of some additional covariates $x_t = (x_{1t}, ..., x_{kt})' \in R^r$. We now model y_t as a conditional beta distribution with time-varying conditional mean, μ_t, and conditional precision, ϕ_t, which are measurable functions of past y_t and known covariates. Specifically, let the model be given by,

$$y_t|\mathcal{F}_{t-1} \underset{i.i.d}{\sim} Beta(\mu_t, \phi_t), \quad \mathcal{F}_{t-1} = \sigma(y_{t-m}, x_{t-m+1} : m \geq 1) \tag{2.2}$$

where the conditional density, $f(y_t|y_{t-m}, x_{t-m+1} : m \geq 1)$, is given by

$$f(y_t|y_{t-m}, x_{t-m+1} : m \geq 1) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\Gamma((1-\mu_t)\phi_t)} y_t^{\mu_t\phi_t-1}(1-y_t)^{(1-\mu_t)\phi_t-1}, \qquad 0 < y < 1$$

It is assumed that the time-varying conditional mean is related to the linear predictor, through a twice differentiable strictly monotonic link function $g_1 : (0,1) \mapsto R$, e.g. the logit function $g_1(x) = \log(\frac{x}{1-x})$. That is, we follow the $\beta ARMA$ model of Rocha and Cribari-Neto (2009) in defining $g_1(\mu_t)$ as a function of a set of regressors, x_t, and an ARMA component, τ_t, such that the general expression for the mean is

$$\begin{aligned} g_1(\mu_t) = \eta_{1t} &= x_t'\beta_1 + \tau_t \\ &= \alpha_1 + x_t'\beta_1 + \sum_{i \leq Q_1} \delta_i \left(g_1(y_{t-i}) - x_{t-i}'\beta_1\right) + \sum_{j \leq P_1} \gamma_j \left(y_{t-j} - \mu_{t-j}\right) \end{aligned} \tag{2.3}$$

where $\beta_1 = (\beta_{1,1}, ..., \beta_{1,k_1})$, for notational convenience we also define $\delta = (\delta_1, ..., \delta_{q_1})$ and $\gamma = (\gamma_1, ..., \gamma_{p_1})$ which are the vectors of moving average and autoregressive parameters respectively. Q_1 and P_1 are the sets largest lag of AR and MA included.

From Equation (2.1) it follows that the conditional variance is naturally time-varying as it is a function of the time varying μ_t. To allow for a more flexible variance, we follow Smithsom and Verkuilen (2006) and let the time varying conditional precision be related to a set of regressors in a linear predictor, η_{2t}, through a twice differentiable strictly monotonic link function $g_2 : R^+ \mapsto R$, e.g. the log function $g_2(x) = \log(x)$. Further, to allow for dependence in the precision we also include lagged standardized squared errors. We will refer to the last term as the ARCH component of the model due to the inspiration owed to the ARCH model of Engle (1982).

$$g_2(\phi_t) = \eta_{2t} = \alpha_2 + z_t'\beta_2 + \sum_{j \leq P_2} \kappa_j \epsilon_{t-j}^2 \quad , \quad \epsilon_{t-j} = \frac{(y_{t-j} - \mu_{t-j})}{\sqrt{\frac{\mu_{t-j}(1-\mu_{t-j})}{1+\phi_{t-j}}}}, \tag{2.4}$$

where $\beta_2 = (\beta_{2,1}, ..., \beta_{2,k_2})$, for notational convenience we also define $\kappa = (\kappa_1, ..., \kappa_{p_2})$ which is the vector of the ARCH parameters. P_2 is the largest ARCH lag. As a natural generalization one could consider also including lagged values of ϕ_t or η_{2t}, resulting in a GARCH like structure in the precision. However this would significantly complicate asymptotic inference.

Employing a dependence structure when specifying ϕ_t is new to the beta regression literature and is chosen for its ease of implementation and interpretation. To motivate this particular specification note that $E(y_{t-j}|\mathcal{F}_{t-j-1}) = \mu_{t-j}$ and that $Var(y_{t-j}|\mathcal{F}_{t-j-1}) = \frac{\mu_{t-j}(1-\mu_{t-j})}{1+\phi_{t-j}}$, we therefore have that $E(\epsilon_{t-j}^2|\mathcal{F}_{t-j-1}) = Var(\epsilon_{t-j}|\mathcal{F}_{t-j-1}) = 1$. With larger values of ϵ_{t-j}^2 indicating an uncharacteristically large deviation of y_{t-j} from μ_{t-j}. We can interpret a negative κ_j as indicating volatility clustering. Since ϵ_{t-j}^2 is $\mathcal{F}_{t-j-1}$ measurable it is straightforward to calculate the likelihood.

We refer to the model given by equations (2.3) and (2.4) as the conditional beta time series model or simply as a CBTS(p_1, q_1, p_2) model. The model has a decreasing variance for a mean near the extremes (0 and 1), but allows for greater flexibility than a fixed precision model could. The parameter vector is $\theta = (\alpha_1, \beta_1, \gamma, \delta, \alpha_2, \beta_2, \kappa) \in \Theta = R^{1+k_1+p_1+q_1+1+k_2+p_2}$.

3 Asymptotic Theory in the CBTS Model

Standard arguments for likelihood estimators are based on the verification of the limiting behavior of the likelihood function through the usual Taylor expansions of the log-likelihood and hence the first, second and third derivatives of the log-likelihood, see eg. Jensen and Rahbek (2004) for standard regularity conditions. Given such regularity conditions, the estimators are consistent, asymptotically Gaussian and moreover testing can be based on χ^2 inference via likelihood ratio statistics. We discuss here briefly the inherent difficulties in establishing these, see also Han and Kristensen (2014) where the conceptually similar GARHC-X model is considered for the $GARCH - X(1,1)$ case. We expect that the likelihood estimators are indeed asymptotically Gaussian under mild conditions, but were not able to establish formally the regularity conditions in terms of conditions on the true parameters θ_0 of the model. Consequently, we supplement our considerations below with a detailed simulation study of the asymptotic distributions of the likelihood estimators, $\hat{\theta}_T$ in the next section.

First consider the score and its variance. These may be used directly to facilitate numerical optimization of the likelihood. However, the non-linearity of the model leads to complex expressions which renders it difficult to derive closed form expressions or formally state the regularity conditions for the model as mentioned.

The conditional beta-type log-likelihood function conditional on $m = max(P_1, P_2, Q_1)$ observations fixed is given by

$$L_T(\theta) := \sum_{t=m+1}^{T} l_t(\theta)$$

where, simplifying the notation of $l_t(\theta)$ as l_t, we have

$$l_t = \log(\Gamma(\phi_t)) - \log(\Gamma(\mu_t\phi_t)) - \log(\Gamma((1-\mu_t)\phi_t)) + \log(y_t)(\mu_t\phi_t - 1) + \log(1-y_t)((1-\mu_t)\phi_t - 1).$$

The score is given by,

$$S_T(\theta) := \sum_{t=1}^{T} s_t(\theta) = \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta}$$

We then have the total derivative with respect to θ as

$$\frac{\partial L_t}{\partial \theta} = \frac{\partial L_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta} + \frac{\partial L_t}{\partial \phi_t}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta} \tag{3.1}$$

$\frac{\partial L_t}{\partial \mu_t}$ and $\frac{\partial L_t}{\partial \phi_t}$ are standard, see Ferrari and Cribari-Neto (2004), and where $\frac{\partial \eta_{1t}}{\partial \theta} = \left(\frac{\partial \eta_{1t}}{\partial \alpha_1}, ...\right)'$ and $\frac{\partial \eta_{2t}}{\partial \theta} = \left(\frac{\partial \eta_{2t}}{\partial \alpha_1}, ...\right)'$ are non-standard and supplied in Appendix B.

Taking the conditional expectation of the score contributions and using that μ_t, ϕ_t, $\frac{\partial \eta_{1t}}{\partial \theta}$, $\frac{\partial \eta_{2t}}{\partial \theta}$, $\frac{\partial \mu_t}{\partial \eta_{1t}}$ and $\frac{\partial \phi_t}{\partial \eta_{2t}}$ are $\mathcal{F}_{t-1}$ measurable it holds that

$$\begin{aligned} E\left(s_t(\theta)\,|\mathcal{F}_{t-1}\right) &= E\left(\frac{\partial L_t}{\partial \theta}\,|\mathcal{F}_{t-1}\right) \\ &= E\left(\frac{\partial L_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta} + \frac{\partial L_t}{\partial \phi_t}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta}\,|\mathcal{F}_{t-1}\right) \\ &= E\left(\frac{\partial L_t}{\partial \mu_t}\,|\mathcal{F}_{t-1}\right)\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta} + E\left(\frac{\partial L_t}{\partial \phi_t}\,|\mathcal{F}_{t-1}\right)\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta}. \end{aligned}$$

From Lemma 2 in the Appendix it follows that $E\left(\left.\frac{\partial L_t}{\partial \mu_t}\right|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right) = 0$ and $E\left(\left.\frac{\partial L_t}{\partial \phi_t}\right|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right) = 0$ so that $E\left(\left.s_t(\theta)\right|_{\theta=\theta_0}|\mathcal{F}_{t-1}\right) = 0$. That is, the score contribution is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$ when evaluated in the true parameter values. Thus provided stationarity and ergodicity of $\{y_t\}$ as well as finite higher order moments, standard arguments would imply asymptotic normality of the score provided contraction conditions apply to the recursions of $\frac{\partial \eta_{1t}}{\partial \theta}$ and $\frac{\partial \eta_{2t}}{\partial \theta}$. While we expect this to hold we were unable to derive explicit conditions. To illustrate the difficulty in conducting inference consider the simple CBTS(1,1,1) model. To calculate the score contribution we need the two vectors $\frac{\partial \eta_{1t}}{\partial \theta}$ and $\frac{\partial \eta_{2t}}{\partial \theta}$. Using the notational convention that $\prod_{j=1}^{0} = 1$, we can find the following expression for $\frac{\partial \eta_{1t}}{\partial \alpha_1}$, the first element of $\frac{\partial \eta_{1t}}{\partial \theta}$, as the alternating series

$$
\begin{aligned}
\frac{\partial \eta_{1t}}{\partial \alpha_1} &= 1 - \gamma \frac{\partial \mu_{t-1}}{\partial \alpha_1} \\
&= 1 - \gamma \left(\frac{\partial \mu_{t-1}}{\partial \eta_{1t-1}} \frac{\partial \eta_{1t-1}}{\partial \alpha_1} \right) \\
&= \ldots \\
&= \sum_{i=0}^{t} (-1)^i \left[\gamma^i \prod_{j=1}^{i} (1 - \mu_{t-j}) \mu_{t-j} \right],
\end{aligned}
$$

where we have used that $\frac{\partial \mu_{t-1}}{\partial \alpha_1} = \frac{\partial \mu_{t-1}}{\partial \eta_{1t-1}} \frac{\partial \eta_{1t-1}}{\partial \alpha_1}$.

This result differs from the score for the $\beta ARMA$ model of Rocha and Cribari-Neto (2009) and invalidates the asymptotic theory derived for estimators, diagnostic and test statistics suggested in that paper[1]. Similar derivations for the other parameters and higher order derivatives are typically much more complex, and for more general models even $\frac{\partial \eta_{1t}}{\partial \alpha_1}$ becomes difficult to derive in any sort of closed form.

4 Simulation Study

In the previous section it was shown that deriving formal asymptotic theory is quite difficult. In this section we perform a simulation study to evaluate the asymptotics for the ML estimator as well as the empirical size for the LR tests when assuming usual inference, that is, χ^2 asymptotics for LR tests, are valid. We use sample sizes $T = 50$, 100, 200, 500 and 1,000 with $N = 1,000$ Monte Carlo replications for each sample size.

In the following two subsections we consider the following two data generating processes (DGP) for the covariate, let x_t be generated from an AR(1) model given by

$$x_t = \kappa + \psi x_{t-1} + \epsilon_t, \quad \epsilon_t \underset{i.i.d.}{\sim} N(0, \sigma^2) \tag{4.1}$$

We use $\sigma^2 = 0.05$ and with AR parameter $\psi = 0.5$ or $\psi = 0.95$. The two DGPs are respectively somewhat persistent or highly persistent, as commonly seen in macroeconomic and financial time series. The intercept, κ, is set such that $E(x_t) = \frac{\kappa}{1-\psi} = 1$.

We let y_t be generated by the CBTS model of equation (2.2) with mean and precision specifications given by

$$g_1(\mu_t) = \alpha_1 + x_t'\beta_1 + \gamma(y_{t-1} - \mu_{t-1}) + \delta \left(g(y_{t-i}) - x_{t-i}'\beta_1 \right) \tag{4.2}$$

$$g_2(\phi_t) = \alpha_2 + x_t'\beta_2 + \kappa\epsilon_{t-1}^2, \qquad \epsilon_{t-1} = \frac{(y_{t-1} - \mu_{t-1})}{\sqrt{\frac{\mu_{t-1}(1-\mu_{t-1})}{1+\phi_{t-1}}}} < z \tag{4.3}$$

where $g_1(\cdot)$ is the logit function and $g_2(\cdot)$ is the exponential function. We use the parameter values $\alpha_1 = -2$, $\beta_1 = 0.5$, $\gamma = 0.5$, $\delta = 0.5$, $\alpha_2 = 8$, $\beta_2 = 0.5$ and $\kappa = -0.5$. The parameters are chosen such that the simulated

[1] As a result, the authors of the original paper are now preparing a corrigendum to their original paper. The mistake of that particular paper is a result of neglecting the recursive elements of the score, this result then permeates throughout the paper.

y_t has a level around 3%, with some dependence, influence from the covariate and volatility clustering. The level is similar to that of the default rate for speculative issuers examined in section 5. In the very rare case that y_t gets so close to 0 that its value is set to 0 by the computer, we drop the simulated path and simulate a new one. The log-likelihood is maximized numerically, with initial values for α_1 and β_1 based on OLS estimates as suggested in Ferrari and Cribari-Neto (2004) but including x_{t-1} and y_{t-1} as regressors to account for some dependence. We also initialize by matching α_1 and α_2 to the first two moments of the data. When calculating test statistics we also initialize in the unrestricted MLE but add the restrictions of the test statistic. Maximization is carried out using the interior-point method available in Matlab 2015B with the analytical scores derived in section 3[2].

4.1 Finite Sample Performance of ML Estimator Asymptotics

In this subsection we perform a simulation study to illustrate the finite sample properties of the MLE when simulating the model given by equations (4.2)-(4.3). Figure 4.1 (A)-(D) report histogram and kernel density estimates for sample sizes $T = 200$, 500 and 1000 of the estimators along with the asymptotic distributions probability density function when simulating the covariate using equation 4.1 with $\psi = 0.5$.

From Figure 4.1 it appears that the kernel estimates are reasonably close to the fitted normal distribution. Results were unchanged when using $\psi = 0.95$. Close examination of the simulation data revealed that the normal approximation was actually worsened by a few outliers (less than 0.5% of the data), with the remaining 99.5% of the data appearing to follow a Gaussian distribution[3] quite closely.

4.2 Finite Sample Performance of LR Test Asymptotics

In the following subsection we perform a simulation study to illustrate the empirical size and power of the LR tests when using the asymptotic distribution derived in the previous section.

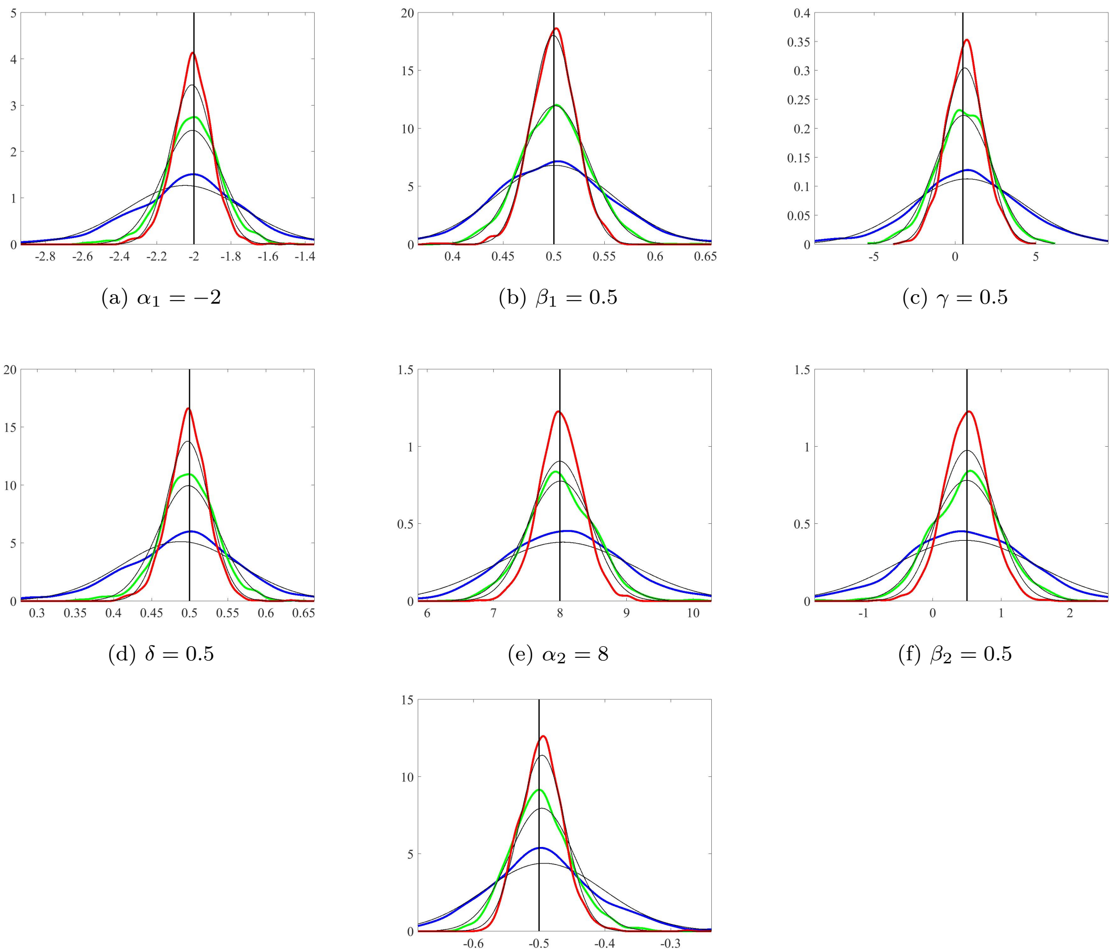
4.2.1 Empirical Size

We consider the hypotheses $H_0: \theta_i = 0$ for $i = 3, ..., 7$ with θ_i the $i'th$ element of θ. Becayse this is a single restriction, the LR test is expected to be asymptotically $\chi^2(1)$ if usual asymptotics apply and we therefore examine if this is the case for the simulations. The results of the simulations are presented in Table 1 for sample sizes $T = 50$, 100, 200, 500 and $1,000$ with $N = 1,000$ replications for each sample size. For each sample size we report the empirical rejection frequency using the 90%, 95% and 99% critical value of the $\chi^2(1)$ distribution as well as the P-value of the Kolmogorov-Smirnov test for the hypothesis that the test statistics are $\chi^2(1)$ distributed.

[2]Simulations not shown indicate using numerical derivatives does not significantly affect the results.

[3]These outliers do no appear to be due to a failure of the maximization procedure which was reinitialized in several different areas of the parameter space and using several different optimization methods.

The results suggest that $\chi^2(1)$ can be used as a good approximation for the distribution of the LR test for most of the parameters when 200 or more observations are used in conjunction with a significance level of $90 - 95\%$, the exceptions being tests on δ and κ parameters. The critical values $90 - 95\%$ of the asymptotic distribution produce a size close to the intended level for tests on all parameters, except δ, when using 500 or more observations. The dependence of the explanatory variable, as measured by the AR parameter ψ, does not appear to influence the empirical size of the test statistics. The 99% critical values are generally slightly oversized.



(a) $\alpha_1 = -2$ (b) $\beta_1 = 0.5$ (c) $\gamma = 0.5$

(d) $\delta = 0.5$ (e) $\alpha_2 = 8$ (f) $\beta_2 = 0.5$

(g) $\kappa = -0.5$

Figure 4.1: Kernel density estimates of the simulated distributions of the estimated parameters for the CBTS model described in section 4. The covariate, x_t was simulated using the model of equation (4.1) with the AR parameter set to 0.5 and an unconditional mean of 1. We display kernel density estimates for sample sizes $T = 200$ (blue), 500 (green) and 1000 (red) with $N = 1,000$ Monte Carlo replications. The vertical black line indicates the true parameter value and the thin black curves are the PDF of normal distributions with mean and variance fitted to the data.

T / CV	$H_0: \gamma = 0$ 90%	95%	99%	KS	$H_0: \delta = 0$ 90%	95%	99%	KS	$H_0: \kappa = 0$ 90%	95%	99%	KS
50	26.8	19.5	10.6	0	52.9	44.3	29.8	0	48.1	41.5	29.9	0
100	14.2	7.8	2.6	0	34.4	25.8	14.4	0	19.7	13.2	5.8	0
200	10.6	5.7	1.7	0.67	23.6	15.7	6.4	0	14.1	8.1	2.3	0
500	10.1	4.5	1.0	0.34	16.8	9.9	4.3	0	11.6	6.1	1.9	0.04
1,000	11.6	6.1	1.5	0.14	16.4	10.9	3.6	0	10.2	5.6	1.1	0.13

T / CV	$H_0: \beta_1 = 0$ 90%	95%	99%	KS	$H_0: \beta_2 = 0$ 90%	95%	99%	KS
50	23.3	16.8	8.4	0	21.9	15.6	7.0	0
100	13.8	8.4	1.8	0	14.6	8.3	1.7	0
200	11.6	6.2	2	0.06	11.2	6.5	1.9	0.43
500	10.0	4.7	1.0	0.53	10.7	6.2	1.5	0.1
1,000	10.4	5.3	1.1	0.93	9.9	4.3	1.1	0.83

(a) Results using $\psi = 0.5$

T / CV	$H_0: \gamma = 0$ 90%	95%	99%	KS	$H_0: \delta = 0$ 90%	95%	99%	KS	$H_0: \kappa = 0$ 90%	95%	99%	KS
50	26.2	19.8	10.5	0	53.8	44.7	29.7	0	52.4	46.6	34.6	0
100	14.3	8.2	2.9	0	32.0	24.1	12.6	0	20.6	13.1	5.9	0
200	11.6	6.4	1.3	0.49	25.8	17.5	7.7	0	13.0	7.0	2.3	0
500	10.8	4.6	0.9	0.20	17.4	10.6	4.1	0	12.6	6.4	1.7	0
1,000	10.9	6.1	1.4	0.43	15.9	9.3	3.1	0	11.3	6.3	2.4	0.11

T / CV	$H_0: \beta_1 = 0$ 90%	95%	99%	KS	$H_0: \beta_2 = 0$ 90%	95%	99%	KS
50	25.3	18.0	8.6	0	22.3	15.5	7.1	0
100	14.8	8.1	1.9	0	13.4	8.5	2.5	0
200	11.8	6.6	2.2	0.5	14.4	8.1	2.1	0.02
500	12.5	5.7	0.7	0.26	11.0	6.4	1.6	0.62
1,000	11.5	6.0	1.5	0.85	11.9	7.1	2.2	0.68

(b) Results using $\psi = 0.95$

Table 1: Empirical rejection frequency (ERF) in percent for the LR test for either of the hypothesis listed in each table using the 90%, 95% or 99% critical values (CV) of the $\chi^2(1)$ distribution. Also shown is the P-value of the Kolmogorov-Smirnov test for the hypothesis that the data is $\chi^2(1)$ distributed. The covariate, x_t was simulated using the model of equation (4.1) with the AR parameter set to 0.5 and an unconditional mean of 1. We display ERF's for sample sizes (T) 50, 100, 200, 500 and 1,000 with $N = 1,000$ Monte Carlo replications.

5 Empirical Application to US Default Rates $1973 - 2015$

In this section we examine Moody's monthly US 12-month issuer default rates[4] in the period from January 1973 until September 2015 ($T = 514$). The data is available for both the non-speculative grade issuers and speculative grade issuers. We first examine the default rate for the non-speculative issuers before turning to the default rate of the speculative grade issuers in subsection 5.2.

Primarily, we wish to examine if the ARMA component, τ_t, is included in the mean when correcting for other variables indicating evidence in favor of the contagion hypothesis, if no such component is significant it indicates evidence in favor of the conditional independence hypothesis.

The secondary goal is to examine which variables are important for explaining the mean and precision parameters for the default rate, whether these are the same when considering non-speculative and speculative issuers and how stable the parameters have been over time.

Lastly, we wish to compare our findings to those of a number of papers; Agosto et al. (2016) who use monthly US default counts in the 1982-2011 period, and Simons and Rolwes (2009) who use quarterly default rates from the Netherlands from 1983-2006.

The non-speculative and speculative default rates are shown in Figure 5.2. From the plots it can be seen that there is a large degree of persistence in the default rates, but also that they vary over time. The non-speculative default rate is as low as 0.09% from December 1979 until March1980 and as high as 7.73% in November of 2009. Similarly, the speculative default rate varies from a minimum of $0,43\%$ from January 1980 until Marts 1980 to a high of 14.71% in November 2009. From the figure it is also discernible that large increases in the default rates have been associated with recessions in the past, as indicated by the shaded time periods.

The choice of covariates in explaining the default rate largely follows that of Lando and Nielsen (2010) and Agosto et al. (2016). We include the following financial and macroeconomic variables in our study: Baa Moody's rated 10-year Treasury spread (SP)[5], 6 month change in Industrial Production Index (IP)[6], the Chicago Fed National Activity Index (NA) [7] released by the Federal Reserve Bank of Chicago. We use NA rather than the Leading Index, released by federal reserve bank of St. Louis, as NA has been published for the entirety of our sample period. We also include the recession indicator released by the National Bureau of Economic Research (RI)[8] and monthly realized volatility (RV) of the S&P 500 index, calculated using daily returns obtained from Bloomberg.

Motivated by Simons and Rolwes (2009) where it was found that quarterly default rates in the Netherlands are influenced by oil prices and interest rates, we include the 12 month changes in percent for oil prices (WTI) and corporate bond yields (CBY). Both variables represent an expenditure for most companies therefore a change

[4]The default rate is available from Moody's webpage from the Monthly Default Report in the Research & Ratings section.
[5]research.stlouisfed.org/fred2/series/BAA10YM
[6]research.stlouisfed.org/fred2/series/INDPRO
[7]chicagofed.org/research/data/cfnai/current-data
[8]research.stlouisfed.org/fred2/series/USREC

could affect the companies ability to repay their loans. We measure the oil price by the West Texas Intermediate[9] and use Moody's Seasoned Baa Corporate Bond yield[10] for the interest rate.

Further, we also include the 12 month return on the S&P500 index[11] (SP12), because in the structural framework of credit risk, see Merton (1973), an increase in the underlying asset i.e. the companies value, should lead to a decrease in default probability[12].

Lastly, since it has been argued that leverage cycles may be important in determining defaults, see eg. Geanakoplos (2009), Geanakoplos and Fostel (2008) and Brave and Butters (2012), we also include the Chicago Fed National Financial Conditions Leverage Sub-index[13] (FCL). FCL is a weighted average of 33 indicators of debt and equity measures in the US financial system, see Brave and Butters (2012) for details. The index is constructed to have an average of zero and a standard deviation of one with positive (negative) values indicate tighter (looser) than average conditions in money markets, debt and equity markets. As the Index is released on a weekly basis, we average over the weeks to get a monthly value.

It should be noted that some care should be taken when interpreting the estimates of the model. Firstly, no variable exists in a vacuum, for example the Recession Indicator is sure to be negatively correlated with National Activity, making ceteris paribus interpretation of either meaningless. Secondly, while we have attempted to use reasonable measures for oil and interests rate changes, it could be the case that companies have hedged their risk at some time period but are still exposed to changes in oil and interests over a different time period. Similarly, if many oil companies finance their operations through bonds it may be that a falling oil price, which one might expect would lead to fewer defaults as companies have less costs in their production actually has the opposite effect when dropping below the production costs of some producers causing them to default. In their November 2015 announcement Moody's write "We note that over a third of corporate defaults have been from commodity sectors so far this year, with the majority from oil and gas", which might be expected following the oil glut and subsequent price drops in 2014-2016, similar to the oil glut of the 1980's. Lastly one should be mindful of reverse causality, for example one might expect an increased leverage to signal an increase in defaults. However, it may well be that lenders are only willing to lend at an increased leverage when there are few defaults which then builds up until default rates rise, this could then actually cause a negative correlation between leverage and defaults.

Figure 5.2 displays the covariates, including a shading indicating a recession as defined in the RI variable. There do not appear to be any trends in the variables, but a degree of persistence is found in all of them. Fitting an AR(1) model to each series we find autoregressive parameters ranging from 0.97 for IP to 0.45 for RV. Large

[9] WTI can be found at research.stlouisfed.org/fred2/series/MCOILWTICO but is only available from 1986, prior to this we use the spot oil price of West Texas Intermediate, available at https://research.stlouisfed.org/fred2/series/OILPRICE. The two oil prices are very highly correlated.

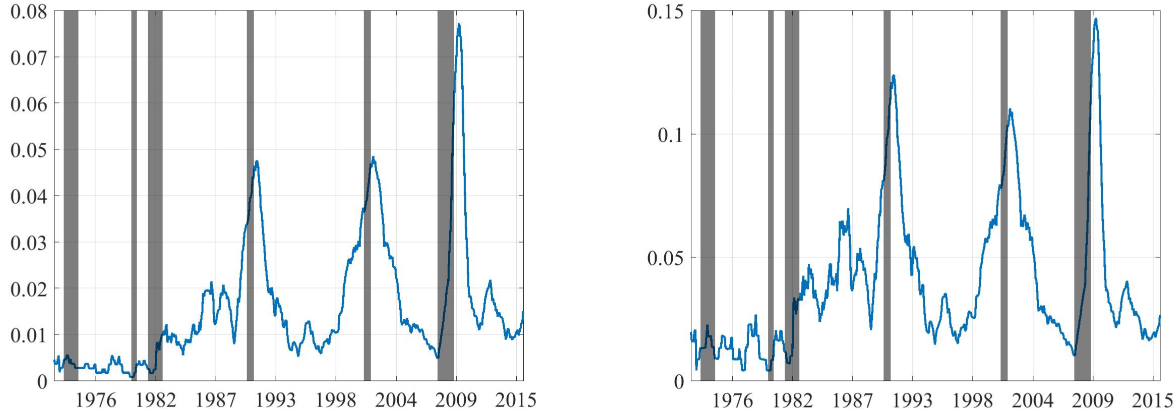
[10] https://research.stlouisfed.org/fred2/series/BAA

[11] Obtained from Bloomberg.

[12] As the company should be able to roll their debt using the increased value of the company, therefore a general increase in stocks value would be expected to decrease default rates.

[13] research.stlouisfed.org/fred2/series/NFCILEVERAGE

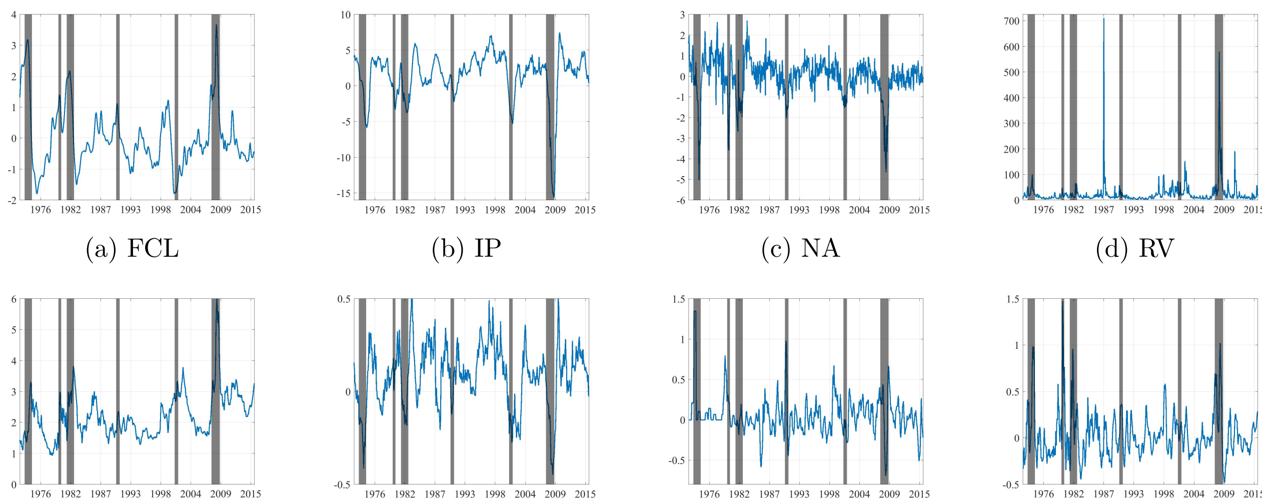
outliers in October of 1987, September 2008 and October 2008 dominates the RV variable, this is due to the crash of 1987 and the financial crisis. We may wish to include dummy variables for these observations. Further, FC shows a tendency to increase in most recessions despite being designed to be uncorrelated with economic conditions. Lastly, there is correlation between the covariates, none less than 0.24 in absolute value and the following with correlations greater than 0.5 in absolute value; RI and NA (-0.67), IP and RI (-0.54), IP and SP (-0.60), SP and NA (-0.52), RI and FC (0.60).



(a) Non-Speculative Issuers

(b) Speculative Issuers

Figure 5.1: Plot of default rates for non-speculative and speculative graded issuers, with shading indicating a recession as defined by the recession indicator released by the National Bureau of Economic Research (RI).



(a) FCL

(b) IP

(c) NA

(d) RV

(e) SP

(f) SP12

(g) WTI

(h) CBY

Figure 5.2: Plot of covariates with shading indicating a recession as defined by the recession indicator released by the National Bureau of Economic Research (RI). The variables are: The Chicago Fed National Financial Conditions Leverage Sub-index (FCL), change in Industrial Production Index (IP), Chicago Fed National Activity Index (NA), realized volatility (RV), Baa Moody's rated 10-year Treasury spread (SP), 12 month return on the S&P500 index (SP12), 6 month changes in percent for oil prices (WTI) and corporate bond yields (CBY).

5.1 Full-sample Analysis

In this subsection we conduct an analysis for the full sample. Figure 5.3 sub-figure (a) shows the Akaike information criterion (AIC) for CBTS(1,1,1) through CBTS(18,18,18). From the plot it appears that little gain is obtained by including lags $4 - 11$, but including lag 12 improves the model. Based on the AIC, and since optimizing the log-likelihood is both time consuming and difficult for larger models, we use the following model as the starting point for our analysis

$$g_1(\mu_t) = \alpha_1 + x_t'\beta_1 + \sum_{j\in\{1,2,3,12\}} \gamma_j(y_{t-j} - \mu_{t-j}) + \sum_{i\in\{1,2,3,12\}} \delta_i \left(g_1(y_{t-i}) - x_{t-i}'\beta_1\right), \tag{5.1}$$

$$g_2(\phi_t) = \alpha_2 + z_t'\beta_2 + \sum_{j\in\{1,2,3,12\}} \kappa_j \epsilon_{t-j}^2 \tag{5.2}$$

Where x_t and z_t both contain all the covariates described in 5. We will refer to the model of equations (5.1) and (5.2) as the full model. Table 2 shows the estimation results for the full model. We have slightly changed the notation of the AR, MA and ARCH lags to highlight that parameters between 3 and 12 are set to 0.

We then iteratively reduce the full model by removing the least significant variables, excepting the intercepts, using LR tests and re-estimating the model until a significance level of 10% is reached for all remaining variables. After this procedure we have a model which we will refer to as the reduced model, parameter estimates are presented in Table 3 for the reduced model.

The fit of the reduced model is evaluated in Figure 5.3 sub-figures (b)-(f) by examining the weighted residuals suggested by Espinheira et al. (2008) which in a well specified beta regression model are approximately $N(0,1)$ distributed. From the sub-figures it appears that the reduced model has a good fit to the data.

Examining the estimated model we see that even when including all the covariates many lags are significant. We find that there is evidence of volatility clustering, but this is mainly due to the 2 and 3 month ARCH term with the 1 month ARCH term actually having a positive estimate. The model appears to have several insignificant variables with only RV being significant for the mean specification. For the precision specification RV and WTI are significant. The 12 month lag is highly significant for both the mean and the precision specifications. The 3 month AR and MA terms are both highly significant and negative. The fitted distribution is at all points bell shaped rather than J-shaped[14].

As expected, a large number of variables was removed from both the mean and precision. For the mean only RV and the WTI are significant. The parameter values suggest that increased volatility in the financial markets could cause an increase in the default rate while a drop in oil prices will tend to cause a decrease.

For the precision we see that NA and RV will actually decrease the conditional variance of the default rate whereas an increase in IP will increase the conditional variance of the default rate.

[14] This can not be seen directly from the parameter estimates but for all points in time the requirement was checked manually.

No dependence terms for the mean where removed at the 10% significance level, although the second MA term is close with a P-value of 9.3%. Again we see a large negative dependence in the default rate to the past years default rate, i.e. lag 12 is negative and highly significant for both AR and MA parameters. The ARCH component seem to indicate a volatility clustering effect as indicated by the negative parameter values for the 2, 3 and 12 month lags, with the puzzling existence of a positive parameter for the 1 lag ARCH parameter. This last parameter however is dwarfed by the other lags and is only significant at the 8.9% level.

It was noted earlier that there were some large outliers in RV at October of 1987, September 2008 and October 2008. Corresponding to the crash of 1987 and the onset of the recent financial crisis. In Table D we present the estimation obtained by including dummies for the RV outliers in the full model, as it was done previously we reduce until a 10% significance level is reached, the results are presented in Table 7.

From Table 7 we see that there are only two explanatory variables in the models mean specification, FCL and WTI. That is, the RV variable is replaced by the FCL variable. In Agosto et al. (2016) when examining parameter stability it can similarly be seen that RV is only significant around 2008. We speculate that the link between financial market volatility and corporate defaults is mostly present in cases of extreme volatility and that the evidence for a general link is limited.

5.2 Analysis of Speculative Grade Default Rate

We redo the empirical application but use the default rate for companies designated as speculative (a credit rating of $Ba1$ or worse). Our procedure is similar as the one used for the non-speculative grade in initially estimating the larger model which is then iteratively reduced, the parameter estimates of the full and reduced models are presented in Appendix C.

Similar to the non-speculative default rate model we see that RV is highly significant for the mean, however FCL is now also significant. However, for the precision parameters far more parameters are now significant; IP, SP, NA, RV, SP12 and CBY. We note that IP has a different sign on its estimate than what was found for the non-speculative default rate, but the remaining significant parameters have the same sign. In the MA component we now see more negative signs, but also a smaller 12 month lag. As for the non-speculative default rate we find highly significant 12 month effects. The ARCH terms are only significant at the 3 month lag where there is an indication of volatility clustering.

Parameters for mean			
Parameter	Estimate	LR test	P-value
α_1	-0.07168	13.70347	0.0021***
IP	0.00020	0.00185	0.96567
RI	0.00328	0.05519	0.81427
SP	-0.01013	0.95591	0.32822
NA	0.00576	1.74532	0.18647
FCL	-0.02043	1.43759	0.23053
RV	0.00017	8.16847	0.00426**
SP12	0.02404	0.20830	0.64810
CBY	0.01305	1.44000	0.23014
WTI	-0.19491	4.32757	0.03750

MA-1	3.32602	3.27553	0.07032
MA-2	3.15727	3.27682	0.07027
MA-3	4.22991	4.4957	0.03400*
MA-12	-20.16512	60.44146	0.00000 ***
AR-1	1.03316	204.57602	0.00000 ***
AR-2	0.10618	2.42971	0.11906
AR-3	-0.10185	2.9844	0.08412
AR-12	-0.05521	27.39353	0.00000 ***

Parameters for precision			
Parameter	Estimate	LR test	P-value
α_2	9.47078	1860.73064	0.00000 ***
IP	-0.05878	2.57947	0.10826
RI	-0.04070	0.06553	0.79796
SP	-0.05443	0.18758	0.66494
NA	0.24416	5.22490	0.02227*
FCL	-0.05744	0.43869	0.50775
RV	0.00989	7.26361	0.00704**
SP12	0.41791	0.20830	0.64810
CBY	0.13528	1.44000	0.23014
WTI	-0.30037	4.32757	0.03750*

ARCH-1	0.09784	5.59168	0.01805*
ARCH-2	-0.06810	6.64973	0.00992**
ARCH-3	-0.16588	7.80750	0.00520**
ARCH-12	-0.12907	22.94570	0.00000 ***

Table 2: Parameter estimates for the full model for the non-speculative grade default rate, likelihood ratio tests and P-values.

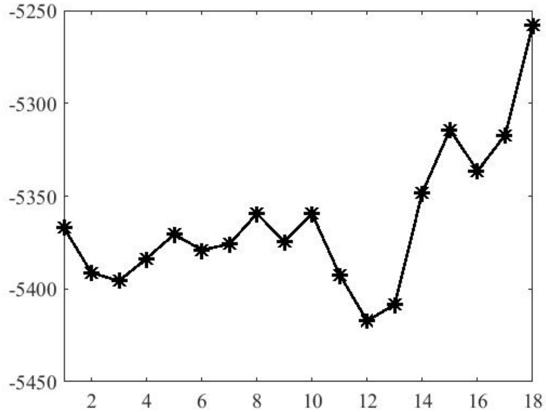
Parameters for mean

Parameter	Estimate	LR test	P-value
α_1	-0.07278	15.35459	0.00009***
RV	0.00016	8.60110	0.00336**
WTI	-0.18559	4.35506	0.03690*
MA-1	4.22262	3.67163	0.05535
MA-2	4.28348	2.82136	0.09302
MA-3	3.92035	5.04191	0.02474*
MA-12	-19.18987	56.16531	0.00000 ***
AR-1	1.01104	208.73117	0.00000 ***
AR-2	0.11293	3.88781	0.04864*
AR-3	-0.08925	3.74313	0.05302
AR-12	-0.05252	37.82379	0.00000 ***

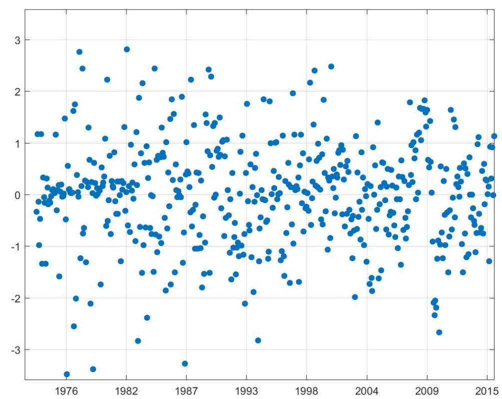
Parameters for precision

Parameter	Estimate	LR test	P-value
α_2	9.50469	3718.50885	0.00000 ***
IP	-0.06415	4.91345	0.02665*
NA	0.33692	17.26600	0.00000 ***
RV	0.00687	7.82486	0.00515**
ARCH-1	0.05570	2.88428	0.08945
ARCH-2	-0.08647	8.92701	0.00281**
ARCH-3	-0.18166	21.53402	0.00000 ***
ARCH-12	-0.14500	26.92891	0.00000 ***

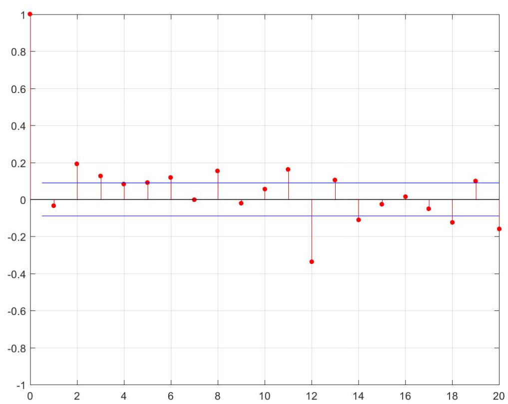
Table 3: Reduced parameter estimates for the reduced model for the non-speculative grade default rate, likelihood ratio tests and P-values.



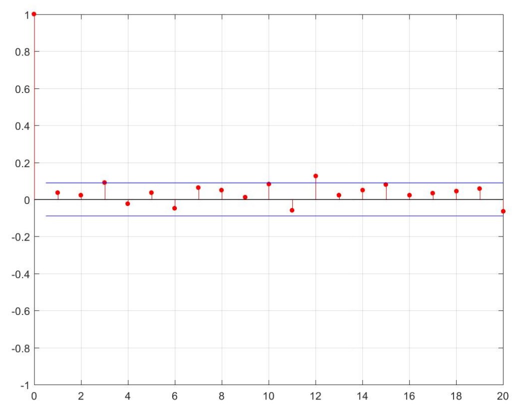
(a) AIC



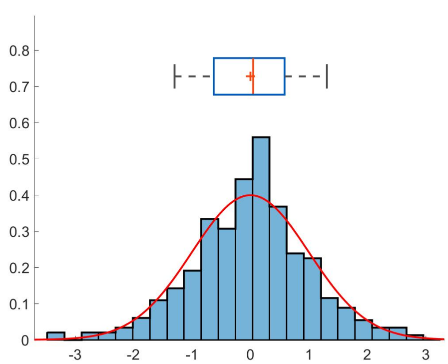
(b) Residuals



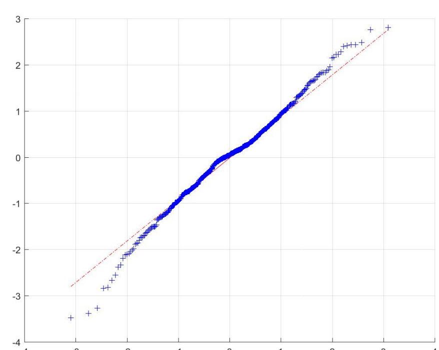
(c) ACF of residuals



(d) ACF of squared residuals



(e) Histogram, boxplot of residuals and N(0,1) PDF



(f) QQ-plot of residuals

Figure 5.3: Sub-figure (a) displays the AIC for CBTS(1,1,1) through CBTS(18,18,18). Sub-figures (b)-(f) display plots evaulating the weighted residuals suggested by Espinheira et al. (2008) for the reduced model.

6 Concluding Remarks

We have proposed an extension to the Beta-ARMA model of Rocha and Cribari-Neto (2009) which allows for dependence and explanatory variables in both the mean and the precision parameters. We have discussed some issues related to inference and note that there exists an error in the results of Rocha and Cribari-Neto (2009). Simulations presented suggest that standard inference applies for realistic sample sizes for at least some parameter values.

We suggest that working with default counts may be biased towards the contagion hypothesis and that working with default rates using our model solves this problem. We apply our model to Moody's monthly US 12-month speculative and non-speculative issuer default rates in the period from December 1972 until September 2015, including several explanatory variables in both the mean and precision parameters. From residuals our model appears to be well specified. After removing insignificant variables we find evidence in favor of an ARMA component to the mean, thus presenting evidence in favor of the contagion hypothesis.

Our results suggest there may exist volatility clustering in the default rates and that the 12 month lag is significant for both the mean and precision parameters, as it enters with a negative parameters. This suggest that a large number of defaults might decrease the default rate 1 year later but that it also increases the variance of the default rate 1 year later. Both appear to be novel results in the literature.

We initially confirm the observation of Agosto et al. (2016) that RV is significant in explaining corporate defaults, but this becomes insignificant when including dummies for September and October 2008, thus suggests a relationship only exists in the most extreme of cases and not even always then since a dummy for September of 1987 was not found to be significant

References

Agosto, A., G. Cavaliere, D. Kristensen, and A. Rahbek (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX). *Journal of Emperical Finance (Forthcoming).*

Brave, S. and R. A. Butters (2012, June). Diagnosing the Financial System: Financial Conditions and Financial Stress. *International Journal of Central Banking 8*(2), 191–239.

Cameron, A. C. and P. K. Trivedi (2013). *Regression Analysis of Count Data.* Number 9781107667273 in Cambridge Books. Cambridge University Press.

Das, S., D. Duffie, N. Kapadia, and L. Saita (2006, January). Common Failings: How Corporate Defaults are Correlated. NBER Working Papers 11961, National Bureau of Economic Research, Inc.

Engle, R. F. (1982, July). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), 987–1007.

Espinheira, P., S. Ferrari, and F. Cribari-Neto (2008). On beta regression residuals. *Journal of Applied Statistics 35*(4), 407–419.

Ferrari, S. and F. Cribari-Neto (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics 31*(7), 799–815.

Geanakoplos, J. (2009, July). The Leverage Cycle. Cowles Foundation Discussion Papers 1715, Cowles Foundation for Research in Economics, Yale University.

Geanakoplos, J. and A. Fostel (2008, September). Leverage Cycles and the Anxious Economy. *American Economic Review 98*(4), 1211–44.

Giesecke, K., F. A. Longstaff, S. Schaefer, and I. Strebulaev (2010, March). Corporate bond default risk: A 150-year perspective. Working Paper 15848, National Bureau of Economic Research.

Han, H. and D. Kristensen (2014, July). Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics 32*(3), 416–429.

Jensen, S. r. T. and A. Rahbek (2004, December). Asymptotic Inference For Nonstationary Garch. *Econometric Theory 20*(06), 1203–1226.

Lando, D. and M. S. Nielsen (2010, July). Correlation in corporate defaults: Contagion or conditional independence? *Journal of Financial Intermediation 19*(3), 355–372.

Merton, R. C. (1973). On the pricing of corporate debt: the risk structure of interest rates. Technical report.

Rocha, A. and F. Cribari-Neto (2009). Beta autoregressive moving average models. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research 18*(3), 529–545.

Sean, K., S. Jorge, and H. David (1999). Predicting default rates: A forecasting model for moody's issuer-based default rates. *Moodys Investors Service 1*(1).

Simons, D. and F. Rolwes (2009, September). Macroeconomic efault Modeling and Stress Testing. *International Journal of Central Banking 5*(3), 177–204.

Smithsom, M. and J. Verkuilen (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods 11*(1), 54–71.

A Moments of Beta Distributed Variables

Given $x \sim Beta(\mu, \phi)$, we have the following results for moments of log transformations, see eg. Ferrari and Cribari-Neto (2004).

$$E(\log(x)) = \psi(\mu\phi) - \psi(\phi),$$

$$E(\log(1-x)) = \psi((1-\mu)\phi) - \psi(\phi),$$

$$E(\log(X)^2) = [\psi(\mu\phi) - \psi(\phi)]^2 + \psi(\mu\phi) - \psi'(\phi),$$

$$E(\log(1-X)^2) = [\psi((1-\mu)\phi) - \psi(\phi)]^2 + \psi'((1-\mu)\phi) - \psi'(\phi),$$

$$E(\log(X)\log(1-X)) = [\psi(\mu\phi) - \psi(\phi)]\,[\psi((1-\mu)\phi) - \psi(\phi)] - \psi'(\phi),$$

$$var\left(\log\left(\frac{x}{1-x}\right)\right) = \psi'(\mu\phi) + \psi'((1-\mu)\phi).$$

B Useful Lemmas for the Score and Information

Using the expressions for moments of a beta distributed variable supplied in Appendix A and with $\psi(.)$ denoting the digamma function the following lemma 1 can be shown

Lemma 1. *Define the following;* $y_t^* := \log(\frac{y_t}{1-y_t})$, $y_t^{**} := \log(1-y_t)$, $y_t^{***} := \log(y_t) log(1-y_t)$ *and* $y_t^{****} := \log(1-y_t)^2 = (y_t^{**})^2$. *We then have their conditional expectations, and the conditional variance of* y_t^*, *as*

$$\mu_t^* := E\left(y_t^* \left|\mathcal{F}_{t-1}\right.\right) = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$$

$$\mu_t^{**} := E\left(y_t^{**} \left|\mathcal{F}_{t-1}\right.\right) = \psi((1-\mu_t)\phi_t) - \psi(\phi_t)$$

$$\mu_t^{***} := E\left(y_t^{***} \left|\mathcal{F}_{t-1}\right.\right) = \left[\psi(\mu_t\phi_t) - \psi(\phi_t)\right]\left[\psi((1-\mu_t)\phi_t) - \psi(\phi_t)\right] - \psi'(\phi_t)$$

$$\mu_t^{****} := E\left(y_t^{****} \left|\mathcal{F}_{t-1}\right.\right) = \left[\psi((1-\mu_t)\phi_t) - \psi(\phi_t)\right]^2 + \psi'((1-\mu_t)\phi_t) - \psi'(\phi_t)$$

$$\sigma_t^{2*} := E\left(\left[(y_t^* - \mu_t^*)\right]^2 \left|\mathcal{F}_{t-1}\right.\right) = \psi'(\mu_t\phi_t) - \psi'((1-\mu_t)\phi_t)$$

$$\sigma_t^{**2} =:= E\left((y_t^{**} - \mu_t^{**})^2\right) = \psi'\left((1-\mu_t)\phi_t\right) - \psi'\left(\phi_t\right)$$

The partial derivatives of $L_t(\theta)$ with respect to μ_t, ϕ_t and their second and product moments are supplied in the following Lemma

Lemma 2. *With* y_t^*, y_t^{**}, μ_t^* *and* μ_t^{**} *as defined in Lemma 1 we have*

$$\begin{aligned}
\frac{\partial L_t(\theta)}{\partial \mu_t} &= -\phi_t\psi(\mu_t\phi_t) + \phi_t\psi\left((1-\mu_t)\phi_t\right) + \phi_t log(y_t) - \phi_t log(1-y_t) \\
&= \phi_t\left(\log\left(\frac{y_t}{1-y_t}\right) - \psi\left(\mu_t\phi_t\right) + \psi\left((1-\mu_t)\phi_t\right)\right) \\
&= \phi_t\left(y_t^* - \mu_t^*\right)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial L_t}{\partial \phi_t} &= \psi(\phi_t) - \mu_t\psi\left(\mu_t\phi_t\right) - (1-\mu_t)\psi\left((1-\mu_t)\phi_t\right) + \mu_t\log(y_t) + (1-\mu_t) + \log(1-y_t) \\
&= \mu_t\left(\log\left(\frac{y_t}{1-y_t}\right) - \psi\left(\mu_t\phi_t\right) + \psi\left((1-\mu_t)\phi_t\right)\right) + \psi(\phi_t) - \psi((1-\mu_t)\phi_t) + \log(1-y_t) \\
&= \mu_t\left(y_t^* - \mu_t^*\right) + y_t^{**} - \mu_t^{**}
\end{aligned}$$

Using that $\mu_t^{***} - \mu_t^{****} - \mu_t^*\mu_t^{**} = \psi'((1-\mu_t)\phi_t)$ *it follows that*

$$\begin{aligned}
E\left(\left[\frac{\partial L_t}{\partial \mu_t}\right]^2 \left|\mathcal{F}_{t-1}\right.\right) &= E\left[\phi\left(y_t^* - \mu_t^*\right)\left|\mathcal{F}_{t-1}\right.\right]^2 \\
&= \phi_t^2\sigma_t^{*2}
\end{aligned}$$

and

$$
\begin{aligned}
E\left(\left[\frac{\partial L_t}{\partial \phi_t}\right]^2 |\mathcal{F}_{t-1}\right) &= E\left(\left[\mu_t\left(y_t^*-\mu_t^*\right)+\left(y_t^{**}-\mu_t^{**}\right)\right]^2 |\mathcal{F}_{t-1}\right) \\
&= E\left(\mu_t^2\left(y_t^*-\mu_t^*\right)^2+\left(y_t^{**}-\mu_t^{**}\right)^2+2\mu_t\left(y_t^*-\mu_t^*\right)\left(y_t^{**}-\mu_t^{**}\right)|\mathcal{F}_{t-1}\right) \\
&= \mu_t^2\sigma_t^{*2}+\sigma_t^{**2}+2\mu_t E\left[\left(y_t^*-\mu_t^*\right)\left(y_t^{**}-\mu_t^{**}\right)|\mathcal{F}_{t-1}\right] \\
&= \mu_t^2\sigma_t^{*2}+\sigma_t^{**2}+2\mu_t(-\psi'((1-\mu_t)\phi_t))
\end{aligned}
$$

and

$$
\begin{aligned}
E\left(\frac{\partial L_t}{\partial \mu_t}\frac{\partial L_t}{\partial \phi_t}|\mathcal{F}_{t-1}\right) &= E\left(\phi_t\left(y_t^*-\mu_t^*\right)\left[\mu_t\left(y_t^*-\mu_t^*\right)+\left(y_t^{**}-\mu_t^{**}\right)\right]|\mathcal{F}_{t-1}\right) \\
&= \phi_t\mu_t\sigma_t^{*2}+\phi_t\left(\psi'(\phi_t)-\psi'((1-\mu_t)\phi_t)\right)
\end{aligned}
$$

The partial derivatives of μ_t and ϕ_t are given in the following Lemma 3

Lemma 3. *Using the logit and log link-functions for g_1 and g_2 we have*

$$\frac{\partial \mu_t}{\partial \eta_{1t}} = \frac{1}{g_1'(\mu_t)} = \mu_t(1-\mu_t) = \mu_t - \mu_t^2$$

$$\frac{\partial \phi_t}{\partial \eta_{2t}} = \frac{1}{g_2'(\phi_t)} = \phi_t.$$

Using the results of Lemmas 2 and 3, we find the following expression for the expectation of the variance of the score

$$
\begin{aligned}
E\left(\left[\frac{\partial L_t}{\partial \theta}\right]^2 |\mathcal{F}_{t-1}\right) &= E\left(\left[\frac{\partial L_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta}+\frac{\partial L_t}{\partial \phi_t}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta}\right]^2 |\mathcal{F}_{t-1}\right) \\
&= E\left(\left[\frac{\partial L_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta}\right]^2 |\mathcal{F}_{t-1}\right)+E\left(\left[\frac{\partial L_t}{\partial \phi_t}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta}\right]^2 |\mathcal{F}_{t-1}\right) \\
&+ 2E\left(\frac{\partial L_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta}\frac{\partial L_t}{\partial \phi_t}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta}|\mathcal{F}_{t-1}\right) \\
&= E\left(\left[\frac{\partial L_t}{\partial \mu_t}\right]^2 |\mathcal{F}_{t-1}\right)\left(\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta}\right)^2+E\left(\left[\frac{\partial L_t}{\partial \phi_t}\right]^2 |\mathcal{F}_{t-1}\right)\left(\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta}\right)^2 \\
&+ E\left(\frac{\partial L_t}{\partial \mu_t}\frac{\partial L_t}{\partial \phi_t}|\mathcal{F}_{t-1}\right)2\frac{\partial \mu_t}{\partial \eta_{1t}}\frac{\partial \eta_{1t}}{\partial \theta}\frac{\partial \phi_t}{\partial \eta_{2t}}\frac{\partial \eta_{2t}}{\partial \theta} \\
&= \phi_t^2\left(\psi'(\mu_t\phi_t)+\psi'((1-\mu_t)\phi_t)\right)\left(\mu_t-\mu_t^2\right)^2\left(\frac{\partial \eta_{1t}}{\partial \theta}\right)^2 \\
&+ \left[\mu_t^2\sigma_t^{2*}+\psi'((1-\mu)\phi_t)-\psi'(\phi_t)+2\mu_t\psi'((1-\mu_t)\phi_t)\right]\phi_t^2\left(\frac{\partial \eta_{2t}}{\partial \theta}\right)^2 \qquad \text{(B.1)} \\
&+ \phi_t\left[\mu_t\sigma_t^{2*}-\psi'((1-\mu_t)\phi_t)\right]2\left(\mu_t-\mu_t^2\right)\frac{\partial \eta_{1t}}{\partial \theta}\phi_t\frac{\partial \eta_{2t}}{\partial \theta} \qquad \text{(B.2)}
\end{aligned}
$$

Parameters for mean			
Parameter	Estimate	LR test	P-value
α_1	-0.00130	0.03581	0.84992
IP	-0.01029	1.11074	0.29192
RI	-0.01242	0.35386	0.55194
SP	0.01255	1.13511	0.28669
NA	-0.00567	2.35080	0.12522
FCL	-0.4859	4.59063	0.03215*
RV	0.00013	13.31484	0.00026***
SP12	0.00719	0.15860	0.69045
CBY	0.02293	0.94508	0.33097
WTI	0.00000	0.00002	0.99639
MA-1	-1.69665	3.49131	0.06169
MA-2	-1.51656	4.47343	0.03443*
MA-3	-0.93493	2.07778	0.14946
MA-12	-12.28802	53.68793	0.00000***
AR-1	1.14039	193.75087	0.00000***
AR-2	-0.00829	6.00053	0.01430*
AR-3	-0.10148	1.71556	0.14946
AR-12	-0.03209	27.43364	0.00000***

Parameters for precision			
Parameter	Estimate	LR test	P-value
α_2	7.00872	315.08911	0.00000***
IP	0.07335	9.25022	0.00235**
RI	0.38818	2.11923	0.14546
SP	0.30635	39.55042	0.00000***
NA	0.26151	5.35775	0.02063*
FCL	-0.04463	0.83535	0.36073
RV	0.01470	10.40583	0.00126**
SP12	1.09474	6.32007	0.01194*
CBY	0.14764	6.27895	0.01222*
WTI	-0.05449	0.10919	0.74107
ARCH-1	0.03973	2.28595	0.13055
ARCH-2	-0.03387	1.57222	0.20988
ARCH-3	-0.13774	30.92964	0.00000***
ARCH-12	-0.07559	3.26573	0.07074

Table 4: Parameter estimates for the full model for the speculative default rate, likelihood ratio tests and P-values.

C Tables of Estimates Speculative

Parameters for mean			
Parameter	Estimate	LR test	P-value
α_1	-0.06387	12.76857	0.00035***
FCL	-0.03913	10.24851	0.00137**
RV	0.00014	14.14489	0.00017***
MA-1	2.07273	6.41911	0.01129*
MA-2	2.53084	7.13915	0.00754**
MA-12	-8.96960	75.45613	0.00000 ***
AR-1	1.03528	937.09653	0.00000 ***
AR-12	-0.05484	38.84132	0.00000 ***

Parameters for precision			
Parameter	Estimate	LR test	P-value
α_2	7.33560	1664.05496	0.00000 ***
SP	0.30966	6.55506	0.01046*
NA	0.31951	27.30663	0.00000 ***
RV	0.01013	12.92601	0.00032***
CBY	0.11099	4.29210	0.03829*
ARCH-1	0.06801	4.49505	0.03399*
ARCH-2	-0.08994	8.23242	0.00411**
ARCH-3	-0.18084	51.64435	0.00000 ***

Table 5: Reduced parameter estimates for the reduced model for the speculative grade default rate, likelihood ratio tests and P-values.

Parameters for mean

Parameter	Estimate	LR test	P-value
α_1	-0.07600	9.93414	0.00162 **
IP	-0.00766	2.63333	0.10464
RI	-0.00902	0.59649	0.43992
SP	0.01523	0.86813	0.43992
NA	0.00038	0.04733	0.82778
FCL	-0.05144	5.53702	0.01862 *
RV	-0.00019	116.25579	0.00000 ***
SP12	0.00707	0.10453	0.74646
CBY	-0.01423	0.00884	0.92511
WTI	-0.02291	2.04122	0.15309
30-Sep-1987	0.01385	0.00215	0.96304
30-sep-2008	0.17699	18.30465	0.00002 ***
31-Oct-2008	0.10234	110.47038	0.00000 ***
MA-1	12.29137	23.72818	0.00000 ***
MA-2	7.16837	3.67438	0.05525
MA-3	6.62618	32.45402	0.00000 ***
MA-12	-14.66305	56.95383	0.00000 ***
AR-1	0.93980	246.68468	0.00000 ***
AR-2	0.21309	73.72456	0.00000 ***
AR-3	-0.10855	17.64756	0.00000 ***
AR-12	-0.06311	30.01065	0.00000 ***

Parameters for precision

Parameter	Estimate	LR test	P-value
α_2	10.05947	461.47984	0.00000 ***
IP	-0.14189	15.03152	0.00011 ***
RI	-0.20372	0.65714	0.41757
SP	-0.32817	7.49649	0.00618 **
NA	0.34872	21.74455	0.00000 ***
FCL	-0.18620	7.67287	0.00561 **
RV	0.02363	18.41214	0.00002 ***
SP12	-0.24438	0.20083	0.65405
CBY	0.00000	0.00007	0.99348
WTI	0.41058	3.39631	0.06534
30-Sep-1987	-18.08655	6.38315	0.01152 *
30-sep-2008	-0.00000	0.00001	0.99807
31-Oct-2008	3.78217	3.77848	0.05192
ARCH-1	0.05030	2.56128	0.10951
ARCH-2	-0.10525	20.25920	0.00001 ***
ARCH-3	-0.14736	19.39046	0.00001 ***
ARCH-12	-0.14350	41.81648	0.00000 ***

Table 6: Parameter estimates for the full model for the non-speculative grade default rate, likelihood ratio tests and P-values.

D Tables of Estimates With Dummies

Parameters for mean

Parameter	Estimate	LR test	P-value
α_1	-0.05767	14.76943	0.00012***
FCL	-0.05662	12.33340	0.00044***
WTI	-0.22164	5.22036	0.02232*
30-Sep-1987	0.12218	12.95025	0.00032***
30-sep-2008	0.10397	14.15309	0.00017***
31-Oct-2008	0.07022	10.36111	0.00129**
MA-12	-23.58777	81.07336	0.00000 ***
AR-1	1.07085	394.15715	0.00000 ***
AR-2	0.13811	9.82311	0.00172**
AR-3	-0.17830	16.78098	0.00000 ***
AR-12	-0.04540	38.07460	0.00000 ***

Parameters for precision

Parameter	Estimate	LR test	P-value
α_2	8.95324	4676.07860	0.00000 ***
FCL	-0.21444	6.15989	0.01307*
RV	0.01212	14.53732	0.00014***
SP12	0.69690	5.59366	0.01803*
CBY	0.10123	3.73678	0.05323
31-Oct-2008	6.35810	5.98634	0.01442*
ARCH-1	0.11752	8.81574	0.00299**
ARCH-2	-0.05076	7.38878	0.00656**
ARCH-3	-0.08194	6.90678	0.00859**
ARCH-12	-0.08887	15.75281	0.0007***

Table 7: Parameter estimates for the reduced model for the non-speculative grade default rate, likelihood ratio tests and P-values.