

# Risk and Uncertainty



Rogier Quaedvlieg

Maastricht University School of Business and Economics

Maastricht, Netherlands

A thesis submitted for the degree of

*Doctor of Philosophy at Maastricht University*

May 13, 2016

© 2016 Rogier Quaedvlieg

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

The cover is based on a photo by Chase Lindberg on flickr.  
This book was typeset using L<sup>A</sup>T<sub>E</sub>X.

Printed by **Datawyse / Universitaire Pers Maastricht**

# Risk and Uncertainty

DISSERTATION

to obtain the degree of Doctor at Maastricht University,  
by the authority of Prof. dr. L.L.G. Soete, Rector Magnificus,  
in accordance with the decision of the Board of Deans,  
to be defended in public  
on **Friday, May 13th 2016**, at **10:00**

by

**Rogier Quaedvlieg**

## **Promotors**

Prof. dr. Peter Schotman

Prof. dr. Sébastien Laurent, Aix-Marseille University

## **Assessment committee**

Prof. dr. Jean-Pierre Urbain, Chairman

Prof. dr. Siem Jan Koopman

Prof. dr. Olivier Scaillet

Dr. Stefan Straetmans

This research was financially supported by The Netherlands Organisation for Scientific Research (NWO). I thank the Global Risk Institute and INQUIRE Europe for additional support.

*I refuse to answer that question on the grounds that I don't know the answer*  
- Douglas Adams

# Preface

Writing this thesis was a lot of fun. It gave me the opportunity to meet many brilliant people, to travel the world, and most importantly, to learn a couple of interesting things. My dissertation would not have happened if it were not for some people, and I would not have enjoyed it as much if it were not for some others.<sup>1</sup>

First, I want to thank my supervisors, Sébastien Laurent and Peter Schotman. Sébastien, I would not have written this thesis without you. Starting from the supervision of my Bachelor Thesis, to my Master Thesis, and finally to my PhD Thesis, your supervision got more intense, our meetings more frequent and the output a lot more interesting. I cannot thank you enough for your suggestion to pursue a PhD. It appears that was the right choice.

Peter, despite the fact that your name is on just one of my papers, you have been invaluable in every project I have worked on. Your constant guidance and knowledge about almost everything has thought me how little I still know. Your ability to grasp things I have been looking at for weeks in the span of about two minutes, all the while calling yourself ‘a little slow sometimes’ is inspiring. I hope to someday have your level of intuition.

Next I would like to thank my hosts at CREATES and Duke University, Asger Lunde, Tim Bollerslev and Andrew Patton. I immensely enjoyed my visits to Aarhus and Durham and I am looking forward to our continued collaboration. Tim in particular, I would like to take this opportunity again to thank you for opening so many doors for me. You have become like a third supervisor, in pushing me to become a better and more critical researcher and in guiding me through the job market.

I would like to thank the reading committee, Jean-Pierre Urbain, Siem Jan Koopman, Olivier Scaillet and Stefan Straetmans for their time and reading of the thesis. Obviously the papers in this thesis would not have been possible without my other co-authors, Kris Boudt, Christophe Hurlin, Orimar Sauri and Stephan Smeekes. I enjoyed our collaboration and I have learned a lot from all of you.

A special thanks to the ladies of the secretariat, Carina, Cecile, Els and Francien. You often provided a much needed distraction and fantastic support throughout. Thanks also to Rob, Jaap, Paulo and Jeroen for providing opportunities, help, feedback and interesting discussions. The same goes for the remainder of the potentially page-filling number of

---

<sup>1</sup>The intersection of these sets is non-empty.

colleagues that have come and gone over the past years. It has been fun and I hope to stay in touch with many of you.<sup>2</sup>

This is the part where I thank my girlfriend for the continuous support throughout all the hardship that is a PhD. Judy, you've been fortunate that you did not have to provide a lot of support, although I have no doubt that you would have, and you would have been great at it. However, I've been much more fortunate to have you beside me, and I want to thank you for all the joy you've brought me over the past years in matters mostly unrelated to research.

And last but not least, my parents who have always been there for me, against all odds kept showing interest in my work, and most importantly gave me every opportunity to get the most out of my education and every other aspect of life. Moeke, dank dat je mijn grootste fan bent en dat constant laat blijken. Fatzo, dank dat je me vertelde een echt vak te gaan studeren en het niet bij business te houden.

---

<sup>2</sup>You know who you are. Printing costs are per page...

# Contents

Preface	ii
List of Figures	viii
List of Tables	ix
<b>1 Introduction</b>	<b>1</b>
<b>2 Risk Measure Inference</b>	<b>4</b>
2.1 Introduction . . . . .	4
2.2 Framework and Risk Measure Definitions . . . . .	6
2.3 Hypotheses of Interest and Test . . . . .	9
2.3.1 Comparison Test of Risk Measures . . . . .	9
2.3.2 Bucketing Procedure . . . . .	11
2.3.3 Procedure Implications . . . . .	12
2.3.4 FWE and FDR . . . . .	12
2.3.4.1 FWE controlling method . . . . .	13
2.3.4.2 FDR controlling method . . . . .	13
2.4 Bootstrap Implementation . . . . .	14
2.4.1 Bootstrap Algorithm . . . . .	15
2.4.2 Bootstrap Validity . . . . .	15
2.5 Simulation Study . . . . .	16
2.5.1 Simulation Design . . . . .	16
2.5.2 Pairwise Comparison Test . . . . .	18
2.5.3 Bucketing Procedure . . . . .	19
2.6 Empirical Application . . . . .	20
2.6.1 Time Series Properties of Risk Measures . . . . .	22
2.6.2 Buckets . . . . .	24
2.7 Conclusion . . . . .	27
<b>3 CholCov</b>	<b>32</b>
3.1 Introduction . . . . .	32
3.2 Theoretical setup . . . . .	34
3.2.1 Decomposition . . . . .	35
3.3 The Estimator . . . . .	36
3.3.1 Dealing with asynchronicity . . . . .	36
3.3.2 Ranking stocks based on liquidity . . . . .	38
3.3.3 The CholCov . . . . .	40
3.3.4 Using CholCov in practise . . . . .	42

3.4	Asymptotic Results . . . . .	44
3.4.1	Preliminaries . . . . .	45
3.4.2	The case of synchronous observations . . . . .	46
3.4.3	The case of non-synchronous observations . . . . .	47
3.4.4	A CLT for the CholCov . . . . .	48
3.5	Monte Carlo simulation . . . . .	50
3.5.1	Setup . . . . .	51
3.5.2	Comparison of CholCov with the standard (composite) MRC estimators . . . . .	52
3.5.3	Normality of the CholCov estimator in finite samples with stochastic volatility. . . . .	54
3.6	Empirical Illustration I: Value-at-Risk forecasting . . . . .	55
3.6.1	Data . . . . .	56
3.6.2	Methodology . . . . .	57
3.6.3	Results . . . . .	60
3.7	Empirical Illustration II: Forecasting betas . . . . .	62
3.8	Conclusions . . . . .	65
<b>4</b>	<b>Exploiting the Errors</b>	<b>78</b>
4.1	Introduction . . . . .	78
4.2	Realized Volatility-Based Forecasting and Measurement Errors . . . . .	81
4.2.1	Realized Variance High-Frequency Distribution Theory . . . . .	81
4.2.2	The ARQ model . . . . .	82
4.2.3	The HARQ model . . . . .	84
4.2.4	Simulation Study . . . . .	87
4.3	Modeling and Forecasting Volatility . . . . .	90
4.3.1	Data . . . . .	90
4.3.2	In-Sample Estimation Results . . . . .	92
4.3.3	Out-of-Sample Forecast Results . . . . .	93
4.3.4	Dissecting the Superior Performance of the HARQ Model . . . . .	96
4.3.5	Longer Forecast Horizons . . . . .	98
4.4	Robustness . . . . .	100
4.4.1	Alternative Realized Variance Estimators . . . . .	100
4.4.2	Alternative Quarticity Estimators . . . . .	104
4.4.3	Alternative HARQ Specifications . . . . .	105
4.4.4	Alternative Q-Models . . . . .	107
4.5	Conclusion . . . . .	110
<b>5</b>	<b>(Un)Reliable Realized Risk Measures</b>	<b>120</b>
5.1	Introduction . . . . .	120
5.2	Dynamic Modeling with Heteroskedastic Measurement Error . . . . .	123

5.2.1	Notation and Definitions . . . . .	124
5.2.2	Vector Autoregressive Models with Measurement Error . . . . .	125
5.2.3	HAR Models . . . . .	127
5.2.4	EWMA Filters . . . . .	129
5.2.5	HEAVY Models . . . . .	130
5.3	Forecasting the Covariance Matrix . . . . .	130
5.3.1	Data . . . . .	130
5.3.2	In-Sample Results . . . . .	131
5.3.3	Out-of-Sample Results . . . . .	133
5.4	Global Minimum Variance Portfolios . . . . .	135
5.4.1	Setup . . . . .	136
5.4.2	HARQ for portfolio allocation . . . . .	138
5.4.3	Simulation Results: Fundamental and Spurious Turnover . . . . .	139
5.4.4	Empirical Results: Minimum Variance and Tracking Error Portfolios	140
	5.4.4.1 Portfolio Allocation Results . . . . .	141
	5.4.4.2 No Shortselling . . . . .	143
	5.4.4.3 Weekly and Monthly Rebalancing . . . . .	144
5.5	Beta Forecasting . . . . .	146
5.5.1	HARQ for Beta Forecasting . . . . .	148
5.5.2	Simulation Results: The merits of dimension reduction . . . . .	149
5.5.3	In-Sample Results . . . . .	151
5.5.4	Out-of-Sample Results . . . . .	151
5.5.5	Economic Evaluation of Beta Forecasts . . . . .	152
	5.5.5.1 Beta Hedging . . . . .	152
	5.5.5.2 Regression Based Comparison . . . . .	153
5.6	Conclusion . . . . .	156
<b>6</b>	<b>Score-Driven Nelson Siegel</b>	<b>162</b>
6.1	Introduction . . . . .	162
6.2	Nelson Siegel model . . . . .	164
	6.2.1 Factor Hedging with NS . . . . .	165
	6.2.2 Estimation on yields versus returns . . . . .	166
6.3	DCS-NS Model . . . . .	168
	6.3.1 Time-varying volatility . . . . .	170
6.4	Empirical Analysis . . . . .	170
	6.4.1 Data . . . . .	170
	6.4.2 In-sample Results . . . . .	171
6.5	Hedging long-term liabilities . . . . .	173
	6.5.1 Setup . . . . .	173
	6.5.2 Hedging Results . . . . .	175
6.6	Conclusion . . . . .	177

References	179
Research Impact	192
Biography	195

# List of Figures

2.1	MES of JPM and GS . . . . .	23
2.2	Significant difference MES . . . . .	24
3.1	Number of observations available for each element in estimation of $\Sigma$ . . . . .	38
3.2	Duration sorting versus observation count sorting. . . . .	40
3.3	Daily Number of Observations . . . . .	57
4.1	Estimation Error of RV . . . . .	86
4.2	HAR vs. HARQ . . . . .	87
4.3	Distribution of HARQ/HAR ratio . . . . .	89
4.4	Individual Stock Loss Ratios . . . . .	98
5.1	Dynamic Shrinkage . . . . .	138
5.2	Beta Estimates and Forecasts for General Electric . . . . .	148
6.1	Time-variation in $\lambda$ . . . . .	165
6.2	Yield Curves . . . . .	171
6.3	Time Series of $\lambda_t$ for various models. . . . .	173
6.4	Hedging Long-Term Liabilities . . . . .	174

# List of Tables

2.1	Rejection frequencies of the single test of equal VaR . . . . .	19
2.2	Simulation Results Bucketing Procedure . . . . .	21
2.3	Bucket Allocation – Top 10 . . . . .	25
2.4	Number of estimated buckets . . . . .	26
3.1	Results Simulation . . . . .	53
3.2	Results Simulation II: Condition numbers . . . . .	54
3.3	Rejection rates of the normality assumption for the realized covariances based CholCov estimates . . . . .	55
3.4	Dynamic Quantile Test P-Values . . . . .	61
3.5	Loss function MCS results . . . . .	62
3.6	Beta Comparisons . . . . .	64
4.1	Simulation Results . . . . .	88
4.2	Summary Statistics . . . . .	91
4.3	In-Sample Estimation Results . . . . .	92
4.4	Out-of-Sample Forecast Losses . . . . .	95
4.5	Stratified Out-of-Sample Forecast Losses . . . . .	97
4.6	In-Sample Weekly and Monthly Model Estimates . . . . .	99
4.7	Weekly Out-of-Sample Forecast Losses . . . . .	101
4.8	Monthly Out-of-Sample Forecast Losses . . . . .	101
4.9	HAR Models based on Noise-Robust $RV$ s versus HARQ Model . . . . .	103
4.10	HARQ versus HAR Models based on Noise-Robust $RV$ s . . . . .	104
4.11	Alternative $IQ$ estimators. . . . .	106
4.12	Alternative HARQ Specifications . . . . .	107
4.13	Alternative Q-Model In-Sample Estimates . . . . .	108
4.14	Alternative Q-Model Out-of-Sample Forecast Losses . . . . .	109
4.15	MSE Ratios - Rolling Window . . . . .	113
4.16	MSE Ratios - Increasing Window . . . . .	114
4.17	QLIKE Ratios - Rolling Window . . . . .	115
4.18	QLIKE Ratios - Increasing Window . . . . .	116
4.19	Average Parameter Estimates Across Stocks . . . . .	117

4.20	HAR(Q) Models Including $RV^{(all)}$ . . . . .	118
4.21	HARQ Ratios with the Same Left- and Right-Hand-Side Volatility Measure	119
5.1	In-Sample Estimates . . . . .	133
5.2	Out-of-Sample Forecast Results . . . . .	134
5.3	Simulation Results: Fundamental and Spurious Turnover . . . . .	141
5.4	Unconstrained Portfolio Results . . . . .	142
5.5	No Short-sale Portfolio Results . . . . .	144
5.6	Longer Horizon Portfolio Results . . . . .	145
5.7	Simulation Results: Beta Forecasts . . . . .	150
5.8	In-Sample Parameter Estimates Betas . . . . .	151
5.9	Out-of-Sample Forecast Results . . . . .	152
5.10	Regression Based Comparison of Betas . . . . .	155
5.11	Realized Kernel Summary Statistics . . . . .	159
6.1	In-Sample Parameter Estimates . . . . .	172
6.2	Daily hedging error descriptives . . . . .	176

# Chapter 1

## Introduction

Econometrics is a powerful tool to extract economically relevant information from a wide variety of data. It allows us to quantify relationships, estimate important theoretical quantities and forecast economic variables. However, one should always be careful, as all of econometrics relies on mere estimates, which are sometimes very precise and reliable, but often poor and potentially far away from the ‘true’ underlying value we are trying to estimate. It is therefore vital to strive to minimize the uncertainty in the estimates. However, knowing that they can never be perfect, we have to make sure that we, at least, appropriately account for the uncertainty which is still inherent in the estimates.

All measures are imperfect and there are many potential sources of variation. The largest source of uncertainty is due to the fact that we work with limited samples, and we may not have a lot of information to precisely pin down the values of the parameters. Increasing the sample size might help. However, data is not always available, either due to the fact that it was not recorded, or due to the fact that older data, or larger cross-sections are not representative if parameters change.

This dissertation focuses on the general problem of dealing with uncertainty when modeling financial risk from a variety of angles. The chapters all stress the fact that estimation uncertainty and the dynamic properties of parameters need to be taken seriously. Chapter 2 aims to quantify the uncertainty in settings where previously we had no idea about the magnitude. Chapters 3 and 6 propose methods to reduce uncertainty, either by more effectively using data, or by taking time-variation in parameters into account. Chapters 4 and 5 propose models that deal with time-variation in the uncertainty, and show that better economic decisions can be made when we exploit this heterogeneity. These three facets all contribute to the better interpretation of data, or to better decision making with potentially large economic impact. The new estimation techniques and models proposed in this thesis therefore lead to better understanding and analysis of financial problems in asset pricing and risk management.

## 1. INTRODUCTION

---

In the first study, chapter 2, we quantify and analyze the uncertainty in forecasts of conditional risk measures. The paper is motivated by the numerous empirical studies that proposed different measures of systemic risk since the financial crisis of 2008. These papers tend to rely on point estimates to be able to compare and rank the different financial institutions. Single point estimates have the benefit of allowing for a very simple ranking of firms. However, we have to be careful in attributing too much weight to them. The risk measures are typically the result of complicated econometric methods, many of which we do not know the properties of.

The chapter proposes a bootstrap-based algorithm that allows us to quantify the uncertainty in the point estimates, with the goal of statistically testing for differences between risk measure forecasts. Second, the chapter provides a method to obtain a grouped ranking of firms, which takes into account the degree of estimation uncertainty. This way we obtain a set of firms which we cannot statistically distinguish, but we can say with a certain confidence level that the firms are more risky than all others. We find that many of the measures are estimated with a large degree of uncertainty which makes it difficult to statistically distinguish which firms are the most risky.

The third chapter proposes a new method to reduce the uncertainty in high-frequency covariance estimates. High-frequency returns theoretically allow us to get arbitrarily precise estimates of daily (co)variances. However, there is a major obstacle in the estimation of multivariate quantities. When looking at trades at high-frequencies, prices are only observed when stocks trade. Different stocks trade at different times which leads to asynchronous data. In order to be able to estimate the covariance between two assets, their observed prices are typically synchronized. Whichever method is used to put the data on a common time-grid, a large amount of data loss is almost inevitable. When more than two series need to be synchronized at once, the data loss is far greater, as all assets need to have traded at least once to obtain a new observation.

To deal with this problem, the chapter introduces a method to obtain a positive definite estimate of the covariance matrix without having to synchronize all series jointly. As adding an additional series to the synchronization scheme will always result in a sample with less (or at most an equal amount) of observations, efficiency gains can be made by estimating elements on a subset of series. By estimating individual elements, one will typically lose one of the defining properties of the covariance matrix, its positive definiteness, when estimating large-dimensional covariance matrices. By factorizing the covariance matrix using the Cholesky decomposition, each element of the matrix can be estimated on a sample that is only partly synchronized, but the total estimate is ensured to be positive definite. This way we obtain an estimation scheme, called the CholCov, which exploits the different trading intensities to obtain a more efficient estimator that is positive definite by construction. The efficiency gains of the estimator are analyzed both analytically and in simulations.

---

Chapter 4 takes univariate high-frequency volatility estimates as given, but proposes a new dynamic model which adapts to the degree of estimation error in them. Using estimated quantities in dynamic models leads to so-called attenuation bias, where the autoregressive parameters are biased towards zero. For forecasting, the attenuation serves a purpose, as part of the variation in the past is noise, and is not informative for the future.

Based on the asymptotic theory of the Realized Variance estimator, we know the exact form of the variance of the estimation error, which significantly varies from day to day. As such, some days the estimates are precise and other days the estimates are unreliable. We develop a new dynamic model which allows for dynamic attenuation of the autoregressive parameters. The parameters are high when the estimate is precise and low when estimation uncertainty is high. The model allows for increased short-term dynamics when possible, while limiting the impact of noisy estimates on the process. We show that the model's forecasts consistently beat a set of benchmark models.

In chapter 5 I develop a multivariate analogue of chapter 4 and investigate the economic impact of the new dynamic models in portfolio construction and beta forecasting. In the multivariate setting the dynamic models endogenously provide a dynamic alternative to the basic principle used to deal with estimation error in the unconditional setting of shrinking the covariance matrix. I show that taking into account the measurement error using these dynamic models can have a great economic impact. In terms of portfolio allocation the models lead to significantly more stable portfolios, resulting in less turnover and therefore lower transaction costs. I show that a risk-averse investor would be willing to pay up to 180 basis points per year to use my models. Moreover, beta forecasts derived from models incorporating the measurement error lead to improved hedging decisions.

In the final chapter, we investigate the Nelson-Siegel model for the term structure of bond yields. The Nelson-Siegel model is a factor model in which the factor loadings for different maturities are governed by a single parameter. In the empirical literature, this parameter is often not even estimated, but simply set to a certain value inspired by economic intuition. The chapter demonstrates that this important parameter is, in fact, time-varying and proposes a dynamic model to estimate its path.

The term structure of yields is of great importance for a variety of applications. In this chapter we use the model for a hedging exercise of long-term liabilities. Any liability can be hedged by forming a portfolio with opposite factor loadings, which in the Nelson-Siegel model is governed by this single parameter. The dynamic model offers superior hedging performance and reduces the hedging error standard deviation by almost 50% during the financial crisis. Much of the improvement is due to the model for the shape parameter with some further reduction achieved by a GARCH model for the residual risk.

# Chapter 2

## Risk Measure Inference\*

### 2.1 Introduction

Financial risk management is fundamentally based on the comparison of risk measures across different assets, portfolios, or financial institutions. Examples include the comparison of total risk of two portfolios measured by their volatility, of tail risk measured by the Value-at-Risk (VaR) or the Expected Shortfall (ES), of systematic risk measured by the beta, or the comparison of systemic risk scores of two financial institutions, and many others. Comparing unconditional risk measures can be done using a variety of parametric or non-parametric tests. However, most risk measures are expressed conditionally on an information set and the corresponding forecasts are generally issued from a dynamic parametric or semi-parametric model. For instance, a (M-)GARCH model can be used to produce conditional VaR or ES forecasts, or a DCC can be used to estimate a dynamic conditional beta (Engle, 2015). As a consequence, the conditional distribution of the estimated risk measure is generally unknown and depends on the estimation procedure used.

In this paper, we propose a general testing methodology that takes into account estimation uncertainty to statistically test for equality of conditional risk measures for different assets, portfolios, or firms at a *single point* in time. We propose two types of tests. The first one is a bootstrap-based comparison test of two risk measures. This test can be applied to a wide class of conditional risk measures and (semi-)parametric models. For example, it can be used to compare conditional measures of volatility, VaR, or ES for two assets or two portfolios at a particular time. It can also be used to test the relative level of systemic risk for two banks on a given day. Additionally, it can be used to test the equality of two conditional risk measures (for instance two VaRs) issued from two different models (e.g. GARCH and RiskMetrics) for the same asset or the same portfolio.

The second test is a procedure that allocates a large set of assets, portfolios or firms

---

\*This chapter is co-authored with Christophe Hurlin (Université d'Orléans), Sébastien Laurent (Aix-Marseille University) and Stephan Smeekes (Maastricht University) and is published at the Journal of Business & Economic Statistics.

into groups of elements that are statistically indistinguishable from each other in terms of riskiness, given a conditional risk measure. This method, inspired by the Model Confidence Set (MCS) of Hansen, Lunde, and Nason (2011) can be applied to any type of risk measure. However, it is particularly well suited to identify buckets of Global Systemically Important Banks (G-SIBs) that have similar contribution to systemic risk. The intuition is in line with what the Financial Stability Board (FSB) does each year when it publishes its five-bucket list of G-SIBs in order to set extra capital requirement (Basel Committee on Banking Supervision, 2013). By doing so, the FSB recognizes the inevitable estimation uncertainty in their estimated riskiness and do not fully rely on point estimates.

Many measures of systemic risk have been proposed in the academic literature over the past years, the most well-known being the *Marginal Expected Shortfall* (MES) and the *Systemic Expected Shortfall* (SES) of Acharya, Pedersen, Philippon, and Richardson (2010), the *Systemic Risk Measure* (SRISK) of Acharya, Engle, and Richardson (2012) and Brownlees and Engle (2012), and the *Delta Conditional Value-at-Risk* ( $\Delta\text{CoVaR}$ ) of Adrian and Brunnermeier (2014). These measures are designed to summarize the systemic risk contribution of each financial institution into a single figure. The appeal is that there exists a ranking of financial institutions according to their systemic risk measures that can be displayed in real time with a daily or weekly frequency (see for instance the V-Lab website of the Volatility Institute, NYU Stern). However, claiming that firm A is more risky than firm B because its systemic risk measure is higher, implies that risk is estimated without error. This is certainly not the case, since these measures typically rely on dynamic parametric models that require sophisticated estimation techniques. Even if the model is correctly specified, replacing the true parameters of the dynamic model by their estimates has an impact on the estimation accuracy of the risk measure itself. Indeed, there is convincing evidence that systemic risk measures are subject to substantial estimation risk (e.g. Danielsson, James, Valenzuela, and Zer, 2011; Guntay and Kupiec, 2015). If this is taken into account, it is unlikely that one can discern such an absolute ranking.

To the best of our knowledge, there is only one alternative test for equality of systemic risk measures. Castro and Ferrari (2014) propose a method for testing whether two firms differ in terms of their  $\Delta\text{CoVaR}$ . However, their approach is specific to  $\Delta\text{CoVaR}$  and to the linear quantile regression. In contrast, our method is more general as it works with any conditional risk measure (SRISK, SES, VaR, ES, etc.) and is not specific to any particular estimation method.

Our study is related to the literature on estimation risk in dynamic risk models, which is generally assessed through asymptotic confidence intervals. For instance, Chan, Deng, Peng, and Xia (2007) and Francq and Zakoïan (2015) derive the asymptotic confidence intervals for the conditional VaR estimator in the specific context of heavy-tailed GARCH models. Gouriéroux and Zakoïan (2013) consider a different approach based on an Estimation adjusted VaR (EVAR). Alternatively, several papers propose resampling methods

to carry out inference on risk measures. Hartz, Mittnik, and Paoletta (2006) introduce a bootstrap approach to correct the estimation bias and to improve the VaR forecasting ability of the normal-GARCH model. Robio (1999), Reeves (2005), Christoffersen and Gonçalves (2005) and Pascual, Romo, and Ruiz (2006) propose a more general approach to assess the estimation error of volatility, VaR and ES forecasts. Their resampling techniques allow the computation of bootstrap-based confidence intervals around the risk forecasts issued from historical simulation methods or GARCH-type models. Finally, Escanciano and Olmo (2010, 2011) implement robust backtests for the VaR, using resampling methods.

Unlike previous studies, we do not focus on the inference for a *single* financial asset. Our testing strategy is designed to *compare* the riskiness of two or more assets, given the estimation risk of the corresponding risk measures. In that sense, our study can also be related to the literature on forecast comparison tests (Diebold and Mariano, 1995; West, 1996, 2006). However, our null hypothesis, and therefore our test, differs in some important ways. First, in most cases, we do not compare two models, but the riskiness of two assets, portfolios, or financial institutions, measured with the same measure and the same model. Second, we do not compare a forecast to an *ex-post* observation. Finally, and most importantly, we test for equality of two or more conditional risk measures at time  $t$ , for which we have only one estimate each. We do not test the equality of these measures over the full sample.

There are also some similarities with the literature devoted to the volatility forecast comparison, in case our test is used to compare the forecasts of the same risk measure issued from two alternative models (Hansen and Lunde, 2006a; Patton, 2011b). However, our comparison test does not require the use of a proxy variable since it is not designed to determine the ‘best’ model.

The remainder of the paper is structured as follows. Section 2.2 introduces a general definition for the conditional risk measures and gives some examples. Section 2.3 presents two types of tests: a comparison test of two risk measures, and a bucketing procedure. The bucketing procedure is a multiple testing problem, making it important to control the number of false rejections. For that, we consider two alternative methods based on the False Discovery Rate (FDR) and the Family Wise Error Rate (FWE). Section 2.4 discusses the bootstrap implementation and Section 2.5 presents some Monte Carlo simulation results for both tests. In Section 2.6, we propose an empirical application for three systemic risk measures, namely the MES, the SRISK and the  $\Delta\text{CoVaR}$ , based on a panel of 94 US financial institutions. Finally, Section 7 concludes and suggests extensions.

## 2.2 Framework and Risk Measure Definitions

Consider an asset, a portfolio, or a firm indexed by  $i$  and a  $\mathcal{F}_{i,t-1}$ -conditional risk measure (denoted RM) issued from a dynamic parametric or semi-parametric model, where  $\mathcal{F}_{i,t-1}$

denotes the information set available at time  $t - 1$ . Formally, we define RM at time  $t$  as follows:

$$RM_{i,t} = f_i(\theta_i, \omega; X_{i,t-1}), \quad (2.1)$$

where  $f_i(\cdot)$  denotes a functional form that depends on (i) the risk measure itself (for instance, the VaR) and (ii) the parametric or semi-parametric model used to produce the corresponding forecast (for instance, a GARCH model).  $X_{i,t-1}$  is a set of variables belonging to  $\mathcal{F}_{i,t-1}$ ,  $\theta_i$  is the vector of model parameters and  $\omega$  is a vector of parameters specific to the risk measure itself. The latter parameters are determined by the user. For instance, in the case of the VaR, it corresponds to the risk level, generally fixed to 1% or 5% by convention. The framework can easily be extended to test the equality of risk measure forecasts for a horizon  $h > 1$ , by considering the information set  $\mathcal{F}_{t-h}$  rather than  $\mathcal{F}_{t-1}$ .

The notation for  $RM_{i,t}$  encompasses a wide class of (semi-)parametric models and conditional risk measures. For instance,  $RM_{i,t}$  can be a measure of price variation (conditional volatility), a systematic risk measure (conditional beta), a tail risk measure (VaR, ES), or a systemic risk measure (MES, SRISK,  $\Delta\text{CoVaR}$ ). The model could be a univariate or a multivariate GARCH model, a quantile or a linear regression model, etc. Thus, this notation can be viewed as a generalization of that used by Gouriéroux and Zakoïan (2013) for parametric VaR models.

As examples of the notation we consider (i) a conditional VaR based on a Student-GARCH model, (ii) the conditional MES of Acharya, Pedersen, Philippon, and Richardson (2010) and Brownlees and Engle (2012), (iii) the SRISK of Acharya, Engle, and Richardson (2012) and Brownlees and Engle (2012) and (iv) the  $\Delta\text{CoVaR}$  of Adrian and Brunnermeier (2014). These are also the risk measures used throughout the paper.

**Example 1 (VaR-GARCH)** Consider a demeaned return process  $r_{i,t}$  associated with an asset indexed by  $i$ . Assuming a t-GARCH(1,1) model for  $r_{i,t}$ , the corresponding conditional VaR for a coverage rate  $\tau \in [0, 1]$  can be expressed as a linear function of the conditional volatility  $\sigma_{i,t}$  of the returns as follows:

$$f_i^{\text{VaR}}(\theta_i, \omega; X_{i,t-1}) = -t_\nu^{-1}(\tau) \sqrt{\frac{v-2}{v}} \sigma_{i,t},$$

with  $\sigma_{i,t}^2 = \gamma_i + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2$ .  $t_\nu^{-1}(\tau)$  denotes the  $\tau$ -quantile of the standardized Student cdf with  $\nu$  degrees of freedom. As such  $\theta_i = (\gamma_i, \alpha_i, \beta_i, \nu)'$ ,  $\omega = \tau$  and  $X_{i,t-1} = \{\underline{r}_{i,t-1}\}$ , where  $\underline{r}_{i,t-1}$  is the set of return observations for firm  $i$  up to time  $t - 1$ .

**Example 2 (MES)** The MES measures how firm  $i$ 's risk taking adds to the financial system risk (measured by the ES). Let us denote the market return as  $r_{m,t} = \sum_{i=1}^n w_{i,t} r_{i,t}$ , with  $w_{i,t}$  the value-weight of firm  $i = 1, \dots, n$  at time  $t$ , and  $r_{i,t}$  the demeaned firm returns. The conditional MES is defined by the first derivative  $-\partial \mathbb{E}_{t-1}(r_{m,t} \mid r_{m,t} < C) / \partial w_{i,t}$ ,

## 2. RISK MEASURE INFERENCE

where  $C$  is a threshold. If the vector process  $(r_{i,t} \ r_{m,t})'$  follows a GARCH-DCC, Brownlees and Engle (2012) show that:

$$f_i^{MES}(\theta_i, \omega; X_{i,t-1}) = -\sigma_{i,t} \rho_{im,t} \mathbb{E}_{t-1}(\epsilon_{m,t} | \epsilon_{m,t} < C/\sigma_{m,t}) \\ - \sigma_{i,t} \sqrt{1 - \rho_{im,t}^2} \mathbb{E}_{t-1}(\epsilon_{i,t} | \epsilon_{m,t} < C/\sigma_{m,t}),$$

where  $\sigma_{i,t}^2 = \gamma_i + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2$ ,  $\rho_{im,t} = Q_{im,t} / \sqrt{Q_{ii,t} Q_{mm,t}}$  with  $Q_{ij,t}$  the  $(i, j)^{th}$  element of the so-called pseudo correlation matrix  $Q_t$ , and  $Q_t = (1 - \alpha_C - \beta_C) \bar{Q} + \alpha_C \epsilon_{t-1} \epsilon'_{t-1} + \beta_C Q_{t-1}$ , with  $\epsilon_{i,t} = r_{i,t} / \sigma_{i,t}$ . Brownlees and Engle (2012) consider a non-parametric estimator (Scaillet, 2004, 2005) for the tail expectations of the standardized returns  $\epsilon_t$ . Then, we have  $\theta_i = (\gamma_i, \gamma_m, \alpha_i, \alpha_m, \beta_i, \beta_m, \bar{Q}, \alpha_C, \beta_C)'$ ,  $\omega = C$  and  $X_{i,t-1} = \{\underline{r}_{i,t-1}, \underline{r}_{m,t-1}\}$ .

**Example 3 (SRISK)** The SRISK is defined as the expected capital shortfall of a given financial institution  $i$ , conditional on a crisis affecting the whole financial system. Acharya, Engle, and Richardson (2012) define the SRISK as follows:

$$f_i^{SRISK}(\theta_i, \omega; X_{i,t-1}) = \max(0; k D_{i,t-1} - (1 - k) W_{i,t-1} (1 - LRME S_{i,t})),$$

where  $D_{i,t}$  and  $W_{i,t}$  denote the book value of total liabilities and the market value of the financial institution respectively, and  $k$  is a prudential capital ratio.  $LRME S_{i,t}$  denotes the long-run MES, i.e. the expectation of the firm equity multi-period return conditional on the systemic event. The LRME S can be approximated as  $LRME S_{i,t} = 1 - \exp(-18ME S_{i,t})$ , where  $ME S_{i,t}$  is the estimate of the MES for firm  $i$  at time  $t$  as defined in Example 2 (Acharya, Engle, and Richardson, 2012). Then, we have  $\omega = (C, k)'$  and  $X_{i,t-1} = \{\underline{r}_{i,t-1}, \underline{r}_{m,t-1}, D_{i,t-1}, W_{i,t-1}\}$ . The vector  $\theta_i$  is similar to that obtained in Example 2. The individual SRISK is generally expressed as a percentage of the aggregate SRISK:

$$f_i^{\%SRISK} = f_i^{SRISK} / \sum_{j=1}^n f_j^{SRISK}.$$

**Example 4 ( $\Delta$ CoVaR)** The  $\Delta$ CoVaR is a systemic risk measure based on the CoVaR, i.e. the conditional VaR of market returns given an event  $\mathbb{C}(r_{i,t})$  observed for firm  $i$ :

$$\Pr\left(r_{m,t} \leq CoVaR_{i,t}^{m|\mathbb{C}(r_{i,t})} \mid \mathbb{C}(r_{i,t})\right) = \alpha. \quad (2.2)$$

The  $\Delta$ CoVaR is the difference between the VaR of the financial system conditional on the distress firm  $i$  and the VaR of the system conditional on the median state of that same firm. Adrian and Brunnermeier (2014) suggest using  $r_{i,t} = VaR_{i,t}(\tau)$  as conditioning event and estimating the CoVaR using a quantile regression model,  $r_{m,t} = \mu_\tau + \gamma_\tau r_{i,t}$ . We then get:

$$f_i^{\Delta CoVaR}(\theta_i, \omega; X_{i,t-1}) = \gamma_\tau \sigma_{i,t} (F^{-1}(\tau) - F^{-1}(0.5)), \quad (2.3)$$

where  $F^{-1}(\tau)$  is the  $\tau$ -quantile of the standardized returns. Hence  $\theta_i = \{\gamma_i, \alpha_i, \beta_i, \gamma_\tau\}$ ,

$\omega = \tau$  and  $X_{i,t-1} = \{r_{i,t-1}, r_{m,t-1}\}$ .

Notice that the functional form  $f_i(\cdot)$  in Equation (2.1) is indexed by  $i$ . Indeed, even if we consider the same risk measure for two assets  $i$  and  $j$ , one may use two different parametric models to produce the corresponding forecasts. For instance, the notation allows for the comparison of the conditional VaR for Bank of America obtained from a GARCH model, and the conditional VaR for Citigroup, obtained using an internal model based on RiskMetrics. On the contrary, if the functional form  $f_i(\cdot)$  is equivalent to that of  $f_j(\cdot)$ , it means that both firms use the same type of parametric model to produce the risk forecasts. However, in all cases, the vectors of parameters  $\theta_i$  and  $\theta_j$  are generally different for  $i \neq j$ .

## 2.3 Hypotheses of Interest and Test

We propose a general framework to statistically test for equality of conditional risk measures obtained for, at least, two different assets, portfolios or financial institutions *at a particular point in time*. In this section, we present two types of tests: (i) a comparison test of two risk measures and (ii) a bucketing procedure. The latter is a form of sequential testing that allocates assets/firms to multiple buckets of equal risk.

### 2.3.1 Comparison Test of Risk Measures

We wish to test whether two assets or firms indexed by  $i$  and  $j$  respectively, present the same level of risk *at time*  $t$  with respect to the conditional risk measure  $RM_t$ . Such a risk comparison test may be useful in many contexts. For instance, it allows a fund manager to test the equality of two assets' volatilities on a particular date, in order to implement a risk parity investment strategy. It also allows a risk manager to test if the VaR of portfolio  $i$  is equal to the VaR of another portfolio  $j$ , on a given day. A third example of when this would be useful is when a regulator wishes to compare the SRISK of bank  $i$ , say Bank of America, and the SRISK of bank  $j$ , say Citigroup, on a single day, e.g. on September 15th 2008, given the information set available prior to this date.

If there is no model uncertainty, i.e. if the functional forms  $f_i(\cdot)$  and  $f_j(\cdot)$  are known, this test consists of comparing  $RM_{i,t} = f_i(\theta_i, \omega; X_{i,t-1})$  to  $RM_{j,t} = f_j(\theta_j, \omega; X_{j,t-1})$ , where  $\theta_i$  and  $\theta_j$  denote the true value of the parameters. Given the common information set  $\mathcal{F}_{t-1} = \mathcal{F}_{i,t-1} \cup \mathcal{F}_{j,t-1}$  for both assets, the two conditional risk measures are observed. Then, the null hypothesis of equal risk at time  $t$  can be defined as:

$$H_{0,t} : RM_{i,t} = RM_{j,t}. \quad (2.4)$$

The null hypothesis is indexed by  $t$ , to stress the fact that we are testing the equality of two conditional risk measures on a *single date*  $t$  given the information set  $\mathcal{F}_{t-1}$ . Contrary

## 2. RISK MEASURE INFERENCE

---

to the forecast comparison tests (Diebold and Mariano, 1995; West, 1996) for instance, we do not test for  $RM_{i,t} = RM_{j,t}$  over the full sample  $t = 1, \dots, T$ , or over a sequence of out-of-sample forecasts. Thus, the alternative hypothesis  $H_{1,t} : RM_{i,t} \neq RM_{j,t}$  means that the risk of asset  $i$  is different from the risk of asset  $j$  at time  $t$  given  $\mathcal{F}_{t-1}$ , according to the risk measure  $RM_t$ .

The need for statistical inference comes the fact that  $RM_{i,t}$  and  $RM_{j,t}$  are not observed, since the parameters  $\theta_i$  and  $\theta_j$  are generally unknown and replaced by their estimators  $\hat{\theta}_i$  and  $\hat{\theta}_j$ . So, the null hypothesis is based on the true risk measure implied by  $f_i(\cdot)$ ,  $RM_{i,t} = f_i(\theta_i, \omega; X_{i,t-1})$ , while the estimated value  $\widehat{RM}_{i,t} = f_i(\hat{\theta}_i, \omega, X_{i,t-1})$  is affected by estimation risk. Our test boils down to the question of whether  $\left[ f_i(\hat{\theta}_i, \omega; X_{i,t-1}) - f_j(\hat{\theta}_j, \omega; X_{j,t-1}) \right]$  is large enough relative to parameter estimation error coming from  $\{\hat{\theta}_i, \hat{\theta}_j\}$  to reject the null.

Relating our setup to that of the forecast comparison literature, note that the tests of Diebold and Mariano (1995) ignore parameter uncertainty, which is justified asymptotically as the forecast error dominates parameter uncertainty for increasing sample size. On the other hand, West (1996) explicitly considers both uncertainty arising from the forecast errors, which occurs due to the approximation of the unconditional expectation of the loss function by a sample mean, and uncertainty from parameter estimation. As noted above, as we consider conditional risk measures at a single date, we do not take sample averages to approximate the unconditional expectation considered in West (1996). Therefore compared to his setup all uncertainty in our case comes from the estimation of the parameters.

Testing the null hypothesis  $H_{0,t}$  is challenging, as the conditional distribution of the estimated risk measure  $\widehat{RM}_{i,t}$  is generally unknown and may be difficult to obtain depending on the model used to estimate the risk measure. Typically, the estimates are obtained using (M-)GARCH models, whose estimates' distribution is widely unknown. Furthermore, even in the cases where the distribution is known (Chan, Deng, Peng, and Xia, 2007; Gouriéroux and Zakoïan, 2013), the joint distribution of  $\widehat{RM}_{i,t}$  and  $\widehat{RM}_{j,t}$  is almost surely not, except for the trivial, but unlikely case of independence between the two risk measures. As a consequence, traditional testing methods are not directly applicable and a new testing procedure is needed. To achieve this, we use the assumed data generating process (DGP) to bootstrap the conditional risk measures and obtain their distribution at time  $t$ . We propose the following two-sided test statistic:

$$T(\alpha) \equiv \frac{|\hat{x}_{ij,t}|}{c_{ij,t}^*(\alpha)}, \quad (2.5)$$

where  $\hat{x}_{ij,t} = \widehat{RM}_{i,t} - \widehat{RM}_{j,t}$  and  $c_{ij,t}^*(\alpha)$  is the bootstrap critical value obtained from the absolute null-value shifted bootstrap distribution of  $\hat{x}_{ij,t}$ . The use of the critical value means that the  $\alpha\%$  rejection point for all combinations  $(i, j)$  is scaled to 1. Rejection thus occurs at the  $\alpha\%$  level if  $T(\alpha) > 1$ . Ex-post, one may draw conclusions on which firm is

the riskiest based on the sign of  $x_{ij,t}$ . The bootstrap is assumed to be asymptotically valid for the risk measures considered, in the sense that it correctly reproduces the asymptotic distribution of the risk measure estimator (see Section 2.4.2).

### 2.3.2 Bucketing Procedure

When considering more than two assets, pairwise comparisons become challenging. One could test for the significance of the difference between each pair, appropriately taking into account the multiple testing problems that arise. However, without adding some additional structure, the set of rejections is unlikely to lead to a cohesive ranking. Instead, we propose an iterative bucketing procedure that can be used to obtain a grouped ranking of assets. The objective is to get a complete ranking by means of a procedure inspired by the Model Confidence Set of Hansen et al. (2011). Our procedure produces buckets of equally risky assets, in the sense that we cannot statistically distinguish the assets within one bucket in terms of their riskiness. This testing procedure can be applied to any type of conditional risk measure, but it has particular application in the context of the systemic risk where the goal is to rank the financial institutions according to their systemic risk contribution.

Consider the set of all financial institutions  $\mathcal{N}^0$ . We start with the identification of the set of most risky firms, defined at time  $t$  as:

$$\mathcal{N}_t^{(1)} \equiv \{i \in \mathcal{N}^0 : x_{ij,t} \geq 0 \ \forall j \in \mathcal{N}^0\}. \quad (2.6)$$

The goal is to find the set  $\mathcal{N}_t^{(1)}$ . This is achieved through a sequence of comparison tests where objects in  $\mathcal{N}^0$  are removed from the set under consideration if they are found to be less risky. The null we are testing is therefore

$$H_{0,t,\mathcal{N}} : x_{ij,t} = 0 \ \forall i, j \in \mathcal{N}, \quad (2.7)$$

with  $\mathcal{N} \subseteq \mathcal{N}^0$ , the subset containing the not yet eliminated firms. The null hypothesis states that all firms in the final set, after the elimination procedure, should be equally risky. For any set  $\mathcal{N}$  this can be tested using an equivalence test and an elimination rule (see Section 2.3.4.1). If the equivalence test is rejected, we use the elimination rule to remove the most significantly different firm, reducing the size of  $\mathcal{N}$ , and follow with re-applying the equivalence test. Our set of most risky firms is the subset of  $\mathcal{N}^0$  that contains  $\mathcal{N}_t^{(1)}$  with a certain probability which can be controlled. This procedure identifies the most risky set only. To obtain the full ranking, we apply the procedure on the set  $\mathcal{N}^0 \setminus \hat{\mathcal{N}}_t^{(1)}$  to obtain a second bucket,  $\hat{\mathcal{N}}_t^{(2)}$ . This process is repeated until all firms have been allocated to a bucket.

### 2.3.3 Procedure Implications

Of course, there are many different ways to obtain buckets of equally risky financial institutions, and even to rank them. However, the implications of our procedure are ideally suited to ranking systemic firms.

First, the approach is one directional, which means we only control the Type I error of the null of equal risk, in one direction as well. Since we consider a top-down approach (from the bucket of the most risky firms to the less risky ones), a false rejection leads to a firm being assigned to a less risky cluster in the next iteration. Underestimating the risk is, in our opinion, much more hazardous than the reverse, and this is controlled.

Second, the Type II error of failing to eliminate a firm results in assignment to a too risky bucket. In practice, what might happen is that a firm with a low point estimate but a high standard error might be assigned to a riskier bucket than a firm with a higher point estimate, but a low standard error. In some sense, these firms are *loose cannons*. Their return series have characteristics that make it difficult to estimate their true risk with accuracy. Again, due to the top-down approach, the resulting ranking will be prudent; in case of large uncertainty, a firm is always put in the most risky bucket.

Finally, we want to emphasize that the number of buckets is not specified ex-ante. This is the main difference with the approach proposed by the BCBS. Ex-post, the number of buckets ranges between one and the total number of firms, depending on the precision of the estimates. Therefore, our testing procedure endogenously strikes a balance between compression and accuracy of the ranking.

### 2.3.4 FWE and FDR

The bucketing procedure is clearly a multiple testing problem, and as such it is important to control the number of false rejections. We consider two alternative controlling methods that may result in different allocations (see e.g. Bajgrowicz and Scaillet, 2012).

The Family Wise Error Rate (FWE) is defined as the probability of rejecting at least one of the true null hypotheses. Controlling the FWE requires that the FWE be no bigger than the significance level  $\alpha$ , at least asymptotically. In many applications one might be willing to tolerate a larger number of false rejections if there is a large number of total rejections. Instead of allowing a fixed amount of false rejections, we tolerate a certain proportion of false rejections out of total rejections. This can be achieved by controlling the False Discovery Proportion (FDP). Let  $F$  be the number of false rejections made by a multiple testing method, and let  $R$  be the total number of rejections. The FDP is defined as  $FDP = F/R$  if  $R > 0$  and 0 otherwise. Benjamini and Hochberg (1995) suggest controlling the False Discovery Rate (FDR), the expected value of the FDP. A testing method is said to control the  $FDR$  at level  $\alpha$  if  $FDR = \mathbb{E}(FDP) \leq \alpha$ , for any sample size  $T$ . A testing method is said to control the FDR asymptotically at level  $\alpha$  if  $\lim_{T \rightarrow \infty} \sup FDR \leq \alpha$ .

The next two sections outline the methods to control either the FWE or the FDR. When the number of hypotheses to be tested becomes very large, the FWE loses a lot of power, making it difficult to reject any hypothesis at all. Romano, Shaikh, and Wolf (2008b) argue that the number of false hypotheses rejected may even tend to zero if the number of hypotheses tested increases. Common practice is to control the FWE in ‘small’ problems, and control the FDR in ‘large’ settings. What is small and what is large greatly varies by application. We will shed some light on the performance of our newly proposed test, in the simulation exercise.

### 2.3.4.1 FWE controlling method

In order to carry out the bucketing procedure we need an equivalence test and an elimination rule. In case of equivalence we have that  $x_{ij,t} = 0$  for all  $i, j \in \mathcal{N}$ . We propose the following test statistic:

$$T^{max}(\alpha) \equiv \max_{i,j \in \mathcal{N}} \frac{|\hat{x}_{ij,t}|}{c_{ij,t}^*(\alpha)}. \quad (2.8)$$

Here, the need for standardization of the statistic becomes evident, as we want to identify the firm which is most likely to be different from the rest. If there is a significant difference, an elimination rule follows naturally. We eliminate the firm  $\arg \max_{i,j \in \mathcal{N}} \hat{x}_{ij,t}/c_{ij,t}^*(\alpha)$ , or put simply, the most significantly rejected firm. Once we can no longer reject a null hypothesis, all firms are equally risky and we identified a bucket.

The FWE can be controlled by obtaining an appropriate critical value for the  $T^{max}(\alpha)$  statistic. Its critical value  $d_t^*(\alpha)$  is chosen such that

$$d_t^*(\alpha) = \inf \{x \in \mathbb{R} : P(T^{max}(\alpha) \geq x) \leq \alpha\}. \quad (2.9)$$

In practice, the probability distribution  $P$  is unknown, and we replace it with a suitable bootstrap estimate  $P^*$ , discussed in Section 2.4. The asymptotic results in White (2000) and Romano and Wolf (2005) imply that our bootstrap method controls FWE asymptotically, provided that the bootstrap is asymptotically valid. This FWE-controlling test bears clear similarities to the Reality Check of White (2000), who proposes a method to test whether one of a set of models significantly outperforms a benchmark.

### 2.3.4.2 FDR controlling method

Romano, Shaikh, and Wolf (2008a) propose a method to control the FDR in a bootstrap setting. The intuition is as follows. Consider the ordered series of test statistics, denoted  $T_{(k),t}$ , such that  $T_{(1),t} \leq \dots \leq T_{(s),t}$ , with  $H_{(k),t}$  the corresponding null hypothesis. Define  $T_{(k:l),t}$  as the  $k$ -th largest of the  $l$  test statistics  $T_{(1),t}, \dots, T_{(l),t}$ . The idea is to reject all  $H_{(s),t}, \dots, H_{(s-h^*),t}$ , where  $h^*$  is the largest integer  $h$  satisfying  $T_{(s),t} \geq c_{s,t}, \dots, T_{(s-h),t} \geq c_{s-h,t}$ . Again, controlling the FDR is a matter of choosing the appropriate critical values  $c_{k,t}$ . Romano, Shaikh, and Wolf (2008a) show that, in order to control the FDR at level

## 2. RISK MEASURE INFERENCE

---

$\alpha$ , the critical values are defined recursively as follows. Having determined  $\hat{c}_{1,t}, \dots, \hat{c}_{h-1,t}$ , compute  $\hat{c}_{h,t}$  according to:

$$\hat{c}_{h,t} = \inf \left\{ x \in \mathbb{R} : \sum_{s-h+1 \leq r \leq s} \frac{r-s+h}{r} \times P \left( T_{(h:h),t} \geq x, \dots, T_{(s-r+1:h),t} \geq \hat{c}_{s-r+1}, T_{(s-r:h),t} < \hat{c}_{s-r} \right) \leq \alpha \right\}, \quad (2.10)$$

with

$$\hat{c}_{1,t} = \inf \left\{ x \in \mathbb{R} : \frac{1}{s} P \left( T_{(1),t} \geq x \right) \leq \alpha \right\}. \quad (2.11)$$

Again, the probability distribution  $P$  will be approximated by a bootstrap counterpart.

Having obtained the critical values, starting with  $T_{(s),t}$  and working downwards, we check whether  $T_{(r),t} \geq \hat{c}_{r,t}$  and if the null is rejected, we eliminate the significantly less risky firm from the set. The firms that remain after the  $h^*$  rejected hypotheses are statistically equally risky and form a bucket. Romano, Shaikh, and Wolf (2008a) prove that this bootstrap approach asymptotically controls the FDR conditionally on the bootstrap being asymptotically valid.

### 2.4 Bootstrap Implementation

This section describes how to obtain  $c_{ij,T}^*$  and  $P^*$  at particular date  $T$ . Consider  $N$  assets or firms, and assume a general multivariate DGP for the corresponding returns,  $r_t = g(\theta, \epsilon_t | \mathcal{F}_{t-1})$ , with  $r_t$  and  $\epsilon_t$  vectors of dimension  $N$ , and  $\theta$  the set of model parameters. We assume  $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{N,t})$  to have zero mean and covariance matrix equal to the identity matrix. In this paper we assume i.i.d. innovations, such that an i.i.d. bootstrap suffices. This assumption can be relaxed to allow for, for example, serial correlation, but the bootstrap method has to be adapted to the assumption. In the case of serial correlation one could use a block bootstrap instead.

Notice that this representation allows for non-linear cross-sectional dependence across the  $\epsilon_{i,t}$  elements. We define the inverse,  $\epsilon_t = g^{-1}(\theta, r_t | \mathcal{F}_{t-1})$ , which retrieves the innovations from the observed return process. For instance, consider a single asset ( $N = 1$ ), with demeaned returns  $r_t = g(\theta, \epsilon_t | \mathcal{F}_{t-1}) = \sigma_t \epsilon_t$ , where  $\sigma_t$  follows a GARCH process with parameters  $\theta$ . Then,  $\epsilon_t = g^{-1}(\theta, r_t | \mathcal{F}_{t-1}) = r_t / \sigma_t$  simply corresponds to the standardized return.

To obtain the bootstrap distribution, we employ a multivariate version of the methodology suggested by Pascual, Romo, and Ruiz (2006) and Christoffersen and Gonçalves (2005) for GARCH forecasts. The approach is as follows. First estimate  $\theta$  on the original series  $r_t$  for  $t = 1, \dots, T - 1$ . Generate bootstrap series,  $r^*$ , using  $\hat{\theta}$ , and innovations drawn with replacement from the empirical distribution of the centered residuals. Estimate the same model on the bootstrap series, to obtain  $\hat{\theta}^*$ . The bootstrap risk measures,

$RM_{i,T}^* = f_i^*(\hat{\theta}^*, \omega; X_{i,T-1})$  is computed for each asset  $i = 1, \dots, N$ , based on the original past return series  $r_{T-1}$  and bootstrap parameter estimates  $\hat{\theta}^*$ . The use of the original return series in  $RM_{i,T}^*$ , instead of the bootstrapped ones, ensures that the current state of the returns is taken into account in the bootstrap RM forecast. As such, the bootstrap only measures the estimation uncertainty.

### 2.4.1 Bootstrap Algorithm

We propose the following algorithm:

1. Estimate the models to obtain  $\hat{\theta}$ . Use the parameter estimates to estimate  $\hat{x}_{ij,T}$ , for all pairs  $(i, j) \in \{1, \dots, N\}^2$ .
2. Compute the residuals  $\hat{\epsilon}_t = g^{-1}(\hat{\theta}, r_t | \mathcal{F}_{t-1})$  for all  $t = 1, \dots, T - 1$ .
3. Draw  $s_1, \dots, s_{T-1}$  i.i.d. from the uniform  $\mathcal{U}_{\{1, T-1\}}$  distribution and construct the bootstrap errors from the centered residuals  $\epsilon_t^{*b} = \hat{\epsilon}_{s_t}, \forall t = 1, \dots, T - 1$ .
4. Construct the bootstrap return series  $r_t^{*b} = g(\hat{\theta}, \epsilon_t^{*b} | \mathcal{F}_{t-1})$ .
5. Estimate the model on the bootstrapped series to obtain  $\hat{\theta}^{*b}$ . Compute  $\widehat{RM}_{i,T}^{*b}$  using  $f_i(\hat{\theta}^{*b}, \omega; X_{i,T-1})$  and similarly for  $\widehat{RM}_{j,T}^{*b}$  to obtain  $\hat{x}_{ij,T}^{*b}$ .
6. Repeat steps 3 to 5  $B$  times, obtaining bootstrap statistics  $x_{ij,T}^{*b}, b = 1, \dots, B$ .

Two remarks have to be made concerning this bootstrap algorithm. First, note that in Step 3, we re-sample cross-sectional vectors of residuals. The time-concordant sampling ensures that the potential cross-sectional dependence in the innovations is preserved. Second, the critical values  $c_{ij,T}^*$  and  $d_{ij,T}^*$  are obtained as the  $\alpha$ -quantiles of the ‘null-value shifted’ series  $|\hat{x}_{ij,T}^{*b} - \hat{x}_{ij,T}|$  and  $T_{ij,T}^{max*b} - T_{ij,T}^{max}$ , respectively. Romano, Shaikh, and Wolf (2008b, p. 412) argue that using these ‘null-value shifted’ series is equivalent to inverting bootstrap multiple confidence regions, and therefore a valid approach. For a detailed description on how to obtain the bootstrap critical values in the FDR procedure from the bootstrap distribution, we refer to Romano, Shaikh, and Wolf (2008a).

### 2.4.2 Bootstrap Validity

A formal proof of the asymptotic validity of the bootstrap - in the sense that the bootstrap correctly reproduces the asymptotic distribution of the risk measure estimator - is outside the scope of this paper, as the general setup for the risk measures cannot be treated uniformly with regards to the bootstrap. Bootstrap validity has to be considered for each case separately, and doing so explicitly would complicate the paper. Instead, we provide some general guidelines for checking bootstrap validity. First, the most important condition for the validity of the bootstrap is that it correctly replicates the asymptotic

distribution of the estimators of the parameters  $\theta$ . If the parametric model assumed to estimate  $\theta$  is correct, and the estimators of  $\theta$  are “well-behaved”, for instance by being  $\sqrt{T}$ -consistent and asymptotically normal, then it can typically be shown that the bootstrap is asymptotically valid for these parameters. For instance, Hidalgo and Zaffaroni (2007) and Shimizu (2013) explicitly derive the bootstrap validity for the parameters of stationary (ARMA-)GARCH models.

Our setting contains two additional difficulties. First, the distribution of the model parameter estimators is only an intermediate step in obtaining the distribution of  $\widehat{RM}_{i,t}$ . As argued by Francq and Zakoïan (2015), given the distribution of these parameter estimators, an application of the Delta method allows the derivation of the asymptotic distribution of the risk measure estimate. The same argument can be applied to the bootstrap and suggests that validity of the bootstrap parameter estimators suffices for establishing bootstrap validity of the risk measure. However, a formal proof requires one to deal with the subtleties involved with conditioning on the past for constructing the conditional risk measure. Second, we need the joint distribution of  $\widehat{RM}_{i,t}$  and  $\widehat{RM}_{j,t}$ , which may be more difficult to obtain even if the univariate distributions are known. For these two reasons we believe that formal proofs of bootstrap validity for a general class of risk measures deserve separate attention and are outside the scope of the paper. In what follows, we therefore work under the assumption that the bootstrap method chosen for a particular risk measure is appropriate. For our specific choices of bootstrap methods and risk measures, we return to this issue in the simulation study where we study their small sample performance. The results we find there do not give us a reason to doubt the validity of our bootstrap approach.

### 2.5 Simulation Study

We use Monte Carlo simulations to study the properties of both the single test and the bucketing procedure. The Monte Carlo simulation is performed on 1,000 replications and for the bootstrap we generate  $B = 999$  samples. We always compare the conditional risk measures at time  $T$  and estimate them over the sample 1 to  $T - 1$ . We apply the comparison test to the VaR, and both the single test and the bucketing procedure to the MES, as defined in Examples 1 and 2 respectively. All the results are generated using Ox version 7.00 (see Doornik, 2012) and the G@RCH package version 7.0 (Laurent, 2013).

#### 2.5.1 Simulation Design

For the VaR, we consider two assets, indexed by  $i = 1, 2$ , and the following DGP:

$$r_{i,t} = \sigma_{i,t} \epsilon_{i,t} \tag{2.12}$$

$$\epsilon_{i,t} \stackrel{i.i.d.}{\sim} ST(0, 1, \nu_i), \tag{2.13}$$

where  $\sigma_{i,t}^2$  follows a GARCH(1,1) model with parameters  $(\gamma, \alpha_1, \beta_1) = (0.05, 0.10, 0.85)$  for both return series. The innovations follow a Student distribution with zero mean, unit variance and degrees of freedom  $\nu_i$ . Under the null, the  $\tau$ -VaRs are equal for both series,  $VaR_{1,T}(\tau) = VaR_{2,T}(\tau) \iff t_{\nu_1}^{-1}(\tau) \sqrt{(\nu_1 - 2)/\nu_1} \sigma_{1,T} = t_{\nu_2}^{-1}(\tau) \sqrt{(\nu_2 - 2)/\nu_2} \sigma_{2,T}$ . To impose this equality, we simulate processes and re-scale the returns ex-post such that the volatilities at time  $T$ ,  $\sigma_{1,T}$  and  $\sigma_{2,T}$ , imply the equality of both VaRs. We consider two cases in which the degrees of freedom  $\nu_1$  and  $\nu_2$  are equal or different. In the former case, the volatility at time  $T$  is equal for both firms, in the latter case the volatility will be higher for the firm with higher degrees of freedom. For the case with equal degrees of freedom, we set  $\nu_1 = \nu_2 = 5$ . We set  $\sigma_{1,T} = 2$  and define  $\sigma_{2,T}$  relative to that as  $\Delta\sigma = \sigma_{2,T} - \sigma_{1,T}$ . We use  $\Delta\sigma = \{0.0, 0.1, 0.2\}$  to simulate under the null hypothesis ( $\Delta\sigma = 0$ ) and alternatives ( $\Delta\sigma > 0$ ). In the case of different degrees of freedom, we set  $\nu_1 = 5$  and  $\nu_2 = 7$ , where again  $\sigma_{1,T} = 2$ . We scale  $\sigma_{2,T}$  such that the VaRs at time  $T$  have the same value under the null, i.e.  $\sigma_{2,T} = \frac{t_5^{-1}(\tau)}{t_7^{-1}(\tau)} \sqrt{21/25}(\sigma_{1,T} + \Delta\sigma)$ . In all cases, the coverage rate for the VaR is fixed at 5%, i.e.  $\tau = 0.05$ .

For the MES, we consider the general DGP proposed by Brownlees and Engle (2012), i.e.

$$\begin{aligned} r_{m,t} &= \sigma_{m,t} \epsilon_{m,t} \\ r_{i,t} &= \sigma_{i,t} \left( \rho_{i,t} \epsilon_{m,t} + \sqrt{1 - \rho_{i,t}^2} \xi_{i,t} \right) \\ (\epsilon_{m,t}, \xi_{i,t}) &\sim F, \end{aligned} \tag{2.14}$$

where  $\sigma_{m,t}$  and  $\sigma_{i,t}$  follow GARCH processes, while  $\rho_{i,t}$  follows a DCC as described in Example (3).  $F$  is a general zero mean, unit variance distribution with unspecified non-linear dependence structures. For the Monte Carlo simulations, we restrict the model to a multivariate Gaussian conditional distribution and constant correlations, i.e.  $\rho_{i,t} = \rho_t$  (CCC model). Of course, both assumptions will be relaxed in the empirical application. We have done simulations using DCC correlations for a few parameter settings with a small number of replications and found very similar results to those reported here.

Since the innovations are i.i.d. and all dependence between firms and the market is captured by the correlation, then the MES can be written as:

$$MES_{i,t}(\tau) = \beta_{i,t} ES_{m,t}(\tau), \tag{2.15}$$

where  $\beta_{i,t} = \rho_i \sigma_{i,t} / \sigma_{m,t}$  denotes the conditional beta of the firm  $i$  and  $ES_{m,t}(\tau)$  is the ES of the market returns. Under the normality assumption, the ES has a closed form expression. Denote by  $\phi(\cdot)$  and  $\Phi(\cdot)$  the standard normal univariate pdf and cdf, respectively. The MES can be written as follows:

$$MES_{i,t}(\tau) = \beta_{i,t} \sigma_{m,t} \lambda(\Phi^{-1}(\tau)) = \rho_{i,t} \sigma_{i,t} \lambda(\Phi^{-1}(\tau)), \tag{2.16}$$

## 2. RISK MEASURE INFERENCE

---

where  $\lambda(z) = \phi(z)/\Phi(z)$  is the Mills ratio. Therefore, the MES solely depends on the volatility of the firm and its correlation with the market. Under these assumptions, two firms have equal MES if the product of conditional volatilities and conditional correlations with the market, at time  $T$ , is equal. We use this result to control the relative risk of simulated firms.

The GARCH parameters  $(\gamma, \alpha_1, \beta_1)'$  are set to  $(0.05, 0.10, 0, 85)'$  for each series. In order to simulate the returns under the null and the alternative, we re-scale the simulated process to control for the value of the MES at time  $T$ . For the single test, we generate the returns for two firms and the market. The market has  $\sigma_{m,T} = 1$ , and the first firm has  $\sigma_{1,T} = 2$  and  $\rho_1 = 0.4$ . We vary the volatility and correlation of the second firm. We choose  $\Delta\sigma = \{0, 0.1, 0.2\}$  and  $\Delta\rho = \{0, 0.0125, 0.0250\}$ , where  $\Delta\rho = \rho_2 - \rho_1$ . The distance between the MES of firms 1 and 2 is therefore a function of the parameters  $(\Delta\sigma, \Delta\rho)$ . For instance, setting  $(\Delta\sigma, \Delta\rho) = (0.1, 0.0125)$  results in  $MES_{1,T} = 1.650$  and  $MES_{2,T} = 1.787$ . The null hypothesis of equal MES is obtained for  $(\Delta\sigma, \Delta\rho) = (0, 0)$ .

For the bucketing procedure, we generate the returns for  $N$  firms and the market. In order to obtain firms that satisfy the null hypothesis of equal systemic risk, we give all firms within the same bucket identical variance and correlation. To illustrate the trade-off between the FWE and FDR tests, we simulate  $N = 10, 20, 40, 60, 80, 100$  firms. In each simulation there are  $c = N/5$  buckets, each containing five firms. The market again has  $\sigma_{m,T} = 1$ . All firms  $i$  in bucket 1 have  $\sigma_{i,T}^{(1)} = 2, \rho_i^{(1)} = 0.4$ . All firms  $i$  in bucket  $k = 2, \dots, c$  have  $\sigma_{i,T}^{(k)} = 2 + (k - 1)\Delta\sigma$  and  $\rho_i^{(k)} = 0.4 + (k - 1)\Delta\rho$ . The difference between two successive buckets in terms of volatility and correlation is therefore equal to that between the two firms in the single test of equal MES. We also take the same values for  $\{\Delta\sigma, \Delta\rho\}$ .

### 2.5.2 Pairwise Comparison Test

Table 2.1 reports the rejection frequencies of the null hypothesis of equal VaR and equal MES for  $T = 1,000$  and  $2,000$  observations at the 5% significance level. The empirical size of the test corresponds to the case  $\Delta\sigma = \Delta\rho = 0$ . Results suggest that for both risk measures, and for all the DGPs we consider, the test does not suffer from any size distortion. Indeed, the rejection rates are remarkably close to the nominal size even for  $T = 1,000$ .

The other entries in Table 2.1 correspond to power. We first consider the VaR. When the second VaR is 5% bigger than the first one ( $\Delta\sigma = 0.1$ ), power already exceeds 50%, and it is close to 70% when the difference is twice as big. Power is increasing with the sample size, and interestingly, power is bigger when the two series have different distributions.

Power for the MES test is comparable to the power of the single test for the VaR. The values are very close to those of the VaR, when  $\Delta\rho = 0$ . Small changes in the correlation are more difficult to precisely estimate than changes in volatility, and as such, power is much lower in the direction of increasing correlation compared to increasing volatility. But the differences do stack up: when both  $\Delta\rho$  and  $\Delta\sigma$  are large, power exceeds 90%.

**Table 2.1:** Rejection frequencies of the single test of equal VaR

		VaR					
		T=1,000			T=2,000		
$\nu_2 \setminus \Delta\sigma$		0.0	0.1	0.2	0.0	0.1	0.2
5		0.045	0.523	0.652	0.049	0.613	0.846
7		0.052	0.544	0.721	0.050	0.671	0.844

		MES					
		T=1,000			T=2,000		
$\Delta\rho \setminus \Delta\sigma$		0.0	0.1	0.2	0.0	0.1	0.2
0.0000		0.046	0.414	0.763	0.048	0.592	0.850
0.0125		0.069	0.612	0.854	0.112	0.789	0.891
0.0250		0.199	0.791	0.888	0.310	0.877	0.920

*Note:* The table contains the rejection rates of a single test of equal VaR and MES. Nominal size is 5%.

### 2.5.3 Bucketing Procedure

In order to save space, we only report the results for  $T = 2,000$  and choose a significance level for both the FDR and FWE of  $\alpha = 0.05\%$ . It is difficult to evaluate the bucketing procedure in terms of size and power. This is mainly due to the fact that an error in any of the iterations has an impact on the next steps. Indeed, the composition of the second bucket will be affected by the composition of the first one, and so on. Moreover, we may overestimate the number of buckets if, for instance, the first bucket is split up into two separate buckets, such that the third estimated bucket is in fact the second bucket implied by the DGP. As such, we do not expect to always have a one-to-one correspondence between the generated ranking and the estimated ranking.

We therefore summarize the performance of our bucketing procedure in five numbers, three based on the first bucket only, and two on the full ranking. The first two are the actual FWE and FDR, computed on the first bucket. Next we consider the power of the test, defined as the fraction of less risky firms that are successfully rejected. Finally, to assess the accuracy of the complete ranking, we present the Spearman rank correlation between the true and estimated rankings, as well as the total number of buckets found. The latter should be close to  $N/5$  when the bucketing procedure has an ideal trade-off between Type I and Type II errors.

Table 2.2 presents the results of the simulation. Each panel has one of the five performance criteria, with the results for the FWE (resp. FDR) controlling procedure in the left (resp. right) panel. First, both the FWE and FDR approaches control their respective error, as they converge to 0.05. When the difference between buckets is small or the number of firms is large, the FWE procedure tends to over-reject a little, but the FWE is relatively well controlled when the difference between buckets is large. The FDR is too high when the number of firms is small, and there is little difference between buckets.

## 2. RISK MEASURE INFERENCE

---

There is slight under-rejection when the number of firms becomes very large, but the FDR is nicely around 0.05 when the buckets are furthest apart. Of course, when the FDR is controlled, the actual FWE will be above 0.05, as the number of correct rejections is far larger than the number of true hypotheses. Similarly, the FDR of the FWE controlling procedure is generally below 0.05 for the same reason. Finally, as expected, the FDR procedure is more powerful across all specifications considered.

Note that the FWE controlling procedure further deviates from the target when  $N$  becomes large, while the FDR further deviates from the target when  $N$  is small. For the FWE, the probability of a single false rejection is controlled. As the number of hypotheses is of order  $N^2$ , for greater sets of firms a larger fraction of hypotheses is false, and it becomes more difficult to not make a single false rejection. Similarly, the FDR allows a fraction of the true hypotheses to be falsely rejected. When  $N$  is small, the number of true hypotheses is small and only a small amount of false rejections are allowed. Moreover, if the number of true rejections is small, a single mistakenly rejected hypothesis could swing the ratio to a very different number. On the other hand, when  $N$  is large, more hypotheses are true, and more false rejections are allowed, which is easier to control.

Next, consider the statistics on the complete ranking. First, the Spearman rank correlation gives an indication of how good the ranking is. Importantly, even if all firms are ranked above or at the same level as all firms that are less risky, the Spearman correlation still penalizes the bucketing procedure if they are not in the correct bucket. As such, when a bucket is split up into two estimated buckets, the rank correlation will go down. The rank correlation of the FWE buckets is generally higher for  $N = 10, 20$  and the FDR has higher rank correlation with  $N = 40$  and up. This is in line with general practise where the FDR is often used as the number of hypotheses becomes large and power of FWE controlling procedure drops. This is further evidenced by the final panel which shows the number of buckets. The FDR procedure generally estimates a greater number of buckets, as it rejects more null hypotheses by construction. The FWE generally has far too few buckets. For instance, for  $N = 100$ , even in the case where the distance between buckets is large, the average number of buckets is only 6.597. Interestingly, the FDR procedure comes very close, with an average of 19.103 buckets, when there are 20 true buckets.

### 2.6 Empirical Application

In this empirical application we apply the bucketing procedure to a panel of 94 large U.S. financial firms. The dataset we use is identical to the panel studied by Acharya, Pedersen, Philippon, and Richardson (2010), Brownlees and Engle (2012) and many other papers on similar topics. It contains daily returns and market capitalizations retrieved from CRSP and quarterly book value of equity from Compustat. The data covers the period between January 3, 2000 and December 31, 2012, for a total of 3,269 daily observations. The

**Table 2.2:** Simulation Results Bucketing Procedure

		FWE Controlling Procedure						FDR Controlling Procedure					
$N$		10	20	40	60	80	100	10	20	40	60	80	100
$\Delta\rho$	$\Delta\sigma$	FWE											
0.0125	0.0	0.098	0.144	0.250	0.291	0.332	0.375	0.186	0.225	0.324	0.358	0.404	0.436
0.0250	0.0	0.074	0.096	0.111	0.182	0.274	0.291	0.228	0.304	0.348	0.444	0.496	0.481
0.0000	0.1	0.055	0.062	0.064	0.083	0.100	0.084	0.133	0.185	0.334	0.437	0.532	0.553
0.0125	0.1	0.053	0.059	0.064	0.079	0.093	0.082	0.154	0.186	0.414	0.501	0.621	0.636
0.0250	0.1	0.052	0.054	0.062	0.071	0.083	0.078	0.185	0.208	0.406	0.583	0.587	0.653
0.0000	0.2	0.047	0.047	0.048	0.068	0.088	0.080	0.134	0.272	0.437	0.530	0.601	0.647
0.0125	0.2	0.048	0.047	0.048	0.065	0.078	0.077	0.164	0.336	0.503	0.565	0.657	0.679
0.0250	0.2	0.048	0.047	0.048	0.062	0.071	0.074	0.194	0.417	0.547	0.649	0.747	0.752
		FDR											
0.0125	0.0	0.254	0.187	0.139	0.087	0.063	0.048	0.259	0.186	0.137	0.086	0.064	0.050
0.0250	0.0	0.122	0.084	0.054	0.044	0.041	0.031	0.112	0.100	0.081	0.074	0.061	0.048
0.0000	0.1	0.123	0.053	0.013	0.009	0.008	0.005	0.160	0.076	0.067	0.040	0.036	0.029
0.0125	0.1	0.098	0.042	0.012	0.007	0.009	0.006	0.107	0.064	0.050	0.041	0.035	0.031
0.0250	0.1	0.063	0.041	0.012	0.008	0.006	0.004	0.088	0.059	0.051	0.045	0.036	0.044
0.0000	0.2	0.055	0.038	0.013	0.007	0.005	0.004	0.094	0.072	0.061	0.048	0.039	0.034
0.0125	0.2	0.053	0.044	0.010	0.005	0.003	0.003	0.081	0.061	0.055	0.050	0.041	0.042
0.0250	0.2	0.032	0.032	0.009	0.003	0.002	0.001	0.058	0.052	0.051	0.047	0.048	0.051
		Power											
0.0125	0.0	0.090	0.156	0.227	0.283	0.340	0.369	0.178	0.239	0.301	0.353	0.410	0.437
0.0250	0.0	0.106	0.166	0.499	0.513	0.675	0.591	0.212	0.248	0.370	0.482	0.789	0.737
0.0000	0.1	0.141	0.323	0.565	0.689	0.758	0.781	0.265	0.503	0.829	0.915	0.946	0.955
0.0125	0.1	0.166	0.375	0.571	0.761	0.786	0.818	0.348	0.610	0.886	0.944	0.982	0.986
0.0250	0.1	0.254	0.448	0.649	0.806	0.811	0.831	0.446	0.705	0.929	0.981	0.983	0.992
0.0000	0.2	0.236	0.541	0.720	0.788	0.825	0.842	0.407	0.826	0.940	0.961	0.974	0.978
0.0125	0.2	0.283	0.613	0.795	0.803	0.840	0.881	0.469	0.874	0.966	0.986	0.977	0.995
0.0250	0.2	0.377	0.645	0.828	0.867	0.883	0.926	0.480	0.904	0.972	0.991	0.993	1.000
		Spearman Rank Correlation											
0.0125	0.0	0.643	0.431	0.371	0.318	0.278	0.267	0.613	0.409	0.352	0.312	0.289	0.279
0.0250	0.0	0.706	0.464	0.395	0.366	0.333	0.350	0.660	0.430	0.434	0.389	0.368	0.354
0.0000	0.1	0.726	0.579	0.613	0.733	0.799	0.827	0.705	0.570	0.702	0.777	0.839	0.878
0.0125	0.1	0.798	0.644	0.676	0.800	0.847	0.836	0.797	0.587	0.789	0.791	0.899	0.934
0.0250	0.1	0.816	0.690	0.687	0.832	0.870	0.846	0.808	0.615	0.794	0.839	0.932	0.938
0.0000	0.2	0.772	0.697	0.783	0.848	0.874	0.899	0.742	0.691	0.834	0.903	0.933	0.948
0.0125	0.2	0.863	0.699	0.842	0.871	0.925	0.928	0.747	0.697	0.925	0.937	0.969	0.959
0.0250	0.2	0.910	0.722	0.869	0.918	0.937	0.978	0.769	0.702	0.956	0.985	0.999	0.989
		Number of Buckets											
0.0125	0.0	1.480	1.802	2.022	2.110	2.200	2.266	1.534	1.844	2.110	2.212	2.326	2.388
0.0250	0.0	1.502	1.914	2.248	2.435	3.063	4.140	1.560	2.214	2.564	3.524	4.623	3.899
0.0000	0.1	1.508	2.004	2.582	3.166	3.816	4.200	1.586	2.128	3.190	4.374	5.612	6.740
0.0125	0.1	1.547	2.071	2.704	4.093	4.121	6.146	1.662	2.543	4.446	4.468	9.080	10.411
0.0250	0.1	1.615	2.120	3.036	4.974	5.519	6.450	1.719	2.647	4.702	6.771	9.840	15.215
0.0000	0.2	1.622	2.202	3.240	4.166	4.996	5.678	1.748	2.790	4.916	7.094	9.304	11.362
0.0125	0.2	1.672	2.354	3.576	4.423	5.559	6.224	1.821	2.924	4.917	7.290	12.190	16.129
0.0250	0.2	1.739	2.436	3.975	5.044	5.704	6.597	1.882	3.058	5.560	9.117	12.520	19.103

*Note:* The table contains simulation results for various parameter settings  $\{\Delta\sigma, \Delta\rho\}$  and number of firms  $N$ . Each column gives the results for  $N$  firms, which are allocated to  $N/5$  buckets of five firms each. The left- and right-hand side give the results for the FWE and FDR controlling procedures respectively. The first two panels give the FWE and FDR computed on the first bucket only. The power is the fraction of firms successfully rejected for the first bucket. The fourth panel gives the Spearman Rank correlation between the true and the estimated ranking, and finally we provide the number of estimated buckets.

## 2. RISK MEASURE INFERENCE

---

market return is approximated by the CRSP market value-weighted index return. Market value is determined by CRSP daily closing prices and number of shares outstanding. Quarterly book values of total liabilities are from Compustat (LTQ). This results in a dataset containing all U.S. financial firms with a market capitalization greater than 5 billions USD as of the end of June 2007. A full list of ticker symbols and firms is given in Appendix A.

The objective of this empirical application is twofold. In a first section, we apply our pairwise comparison test for the MES. We consider a subset of financial institutions in order to emphasize the time profile of systemic risk and the need for a comparison of *conditional* risk measures. In the second section, we apply the bucketing procedure to the full sample, contrasting the FWE with the FDR approach. We estimate buckets for the MES, %SRISK and  $\Delta\text{CoVaR}$ .

The estimation of the three systemic risk measures is done according to the same methodology as that recommended by their authors. The MES is estimated using  $C = \text{VaR}_{m,t}(0.05)$  and a DCC-GJR-GARCH model (estimated by QML). We check for possible dynamics in the mean by minimizing the Schwarz Information Criteria for the individual ARMA( $m,n$ )-GJR-GARCH(1,1) models over  $m, n = 0, \dots, 3$ . We test for the presence of serial correlation in the standardized residuals and their squares, and fail to reject the null for all series. As such the bootstrap for serially uncorrelated returns described in Section 2.4 will suffice. For the %SRISK, we fix the capital ratio  $k$  at 8%, following Brownlees and Engle (2012). We only consider the series with strictly positive SRISK estimates. Finally, for the  $\Delta\text{CoVaR}$  we consider a risk level  $\tau$  equal to 5%.

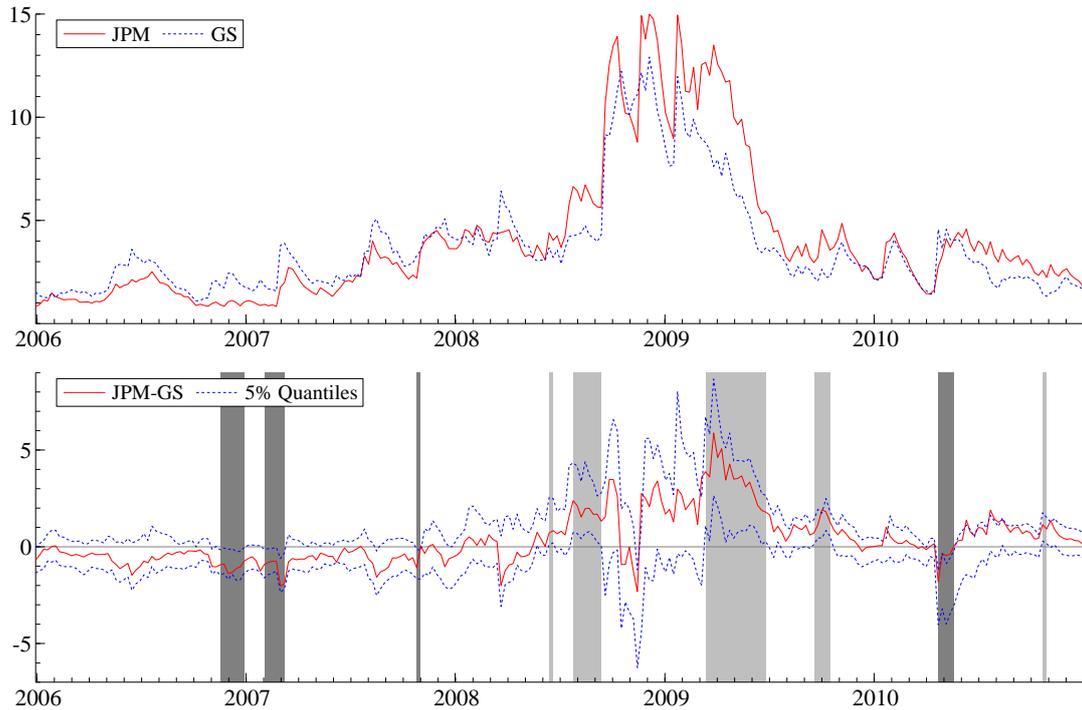
Over time different estimation techniques have been proposed for the various measures, and in this paper we only consider one technique each. We stress that different estimation techniques will have different degrees of uncertainty, and lead to different conclusions. One might obtain more power using simpler specifications, for instance using constant correlations, but then the estimates suffer from a more fundamentally misspecified model. Regardless, the global message would be the same: estimation uncertainty needs to be taken into account when comparing and ranking risk measures.

### 2.6.1 Time Series Properties of Risk Measures

In this section we restrict our analysis to the subset of the sixteen most risky firms which were designated as Global and Domestic Systemically Important Banks (G- and D-SIBs) in 2009 by the Stress Tests of the Federal Reserve and kept that status through 2014. On every Friday of our sample, we estimate the conditional MES for each firm and we obtain the estimates' distribution by means of our bootstrapping procedure. Then, for all pairs of firms, we test for equality of MES at these dates using the test statistic in Equation (2.5).

To illustrate the need for a conditional approach, we plot the MES of J.P. Morgan (JPM) and Goldman Sachs (GS), along with their difference and its 5% confidence bounds

Figure 2.1: MES of JPM and GS

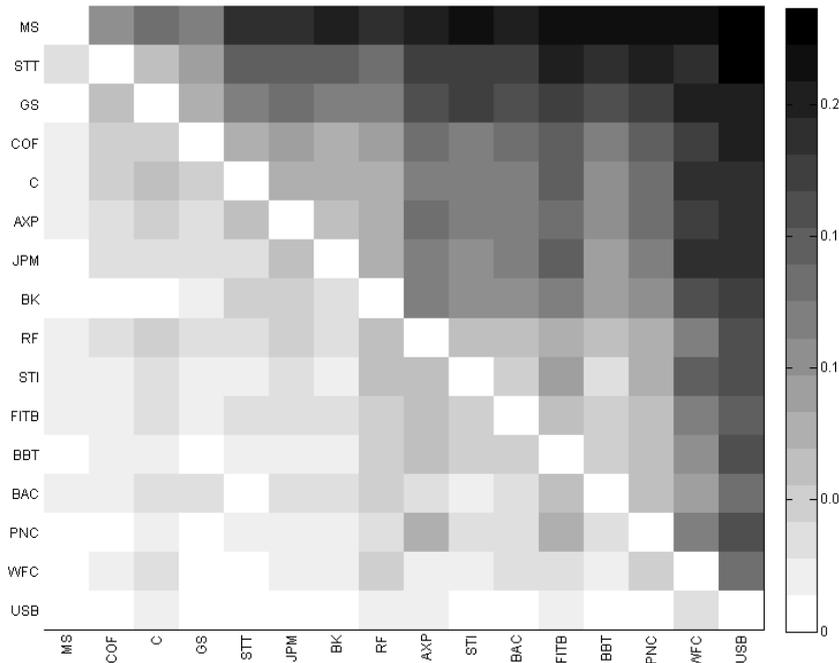


*Note:* The top panel shows the estimated MES of JPM and GS in the period 2006-2011. The MES is estimated every Friday. The bottom panel shows the difference, along with bootstrap confidence bounds. The shaded regions represent a significant difference between the two. When the shading is dark, GS has significantly higher MES than JPM, when it is light the reverse is true.

in Figure 2.1. Significant differences are marked by shaded regions, dark indicating GS is more risky than JPM and light shading indicating the reverse. This figure illustrates that the MES of the two firms are highly correlated. Until 2008 the point estimates for GS are generally higher than those for JPM, and this order is reversed after 2008. However, although the point estimates may be different, they are not frequently significantly different. GS is more risky on 8.5% of sample days, while JPM's risk exceeds GS' on just 5.9% of days, so that the parameters can only be estimated precisely enough on about 14.4% of the days to truly distinguish the two banks. Importantly, significant rejections are clustered with an autocorrelation of 0.7, meaning that the single days where one firm is more risky than the other, are rare.

The results for all other pairs are summarized in Figure 2.2. This figure plots the rejection frequencies for each pair, where the color corresponds to a value determining the frequency at which the firm on the y-axis is found to be more risky than the one on the x-axis. The heatmap shows that even the firms with highest MES are only significantly more risky (at 5%) than firms with the lowest MES in about 20-25% of the time. On average, across pairs, we find a significant difference between firms on 16.4% percent of the days. Different significance levels do not change the relative picture much, but at 10% the highest rejection frequencies approach 50%.

Figure 2.2: Significant difference MES



Note: The heatmap plots the rejection frequencies over the full sample of the hypothesis that  $H_0 : x_{ij,t} = 0$  vs  $H_1 : x_{ij,t} > 0$ , with  $i$  on the y-axis and  $j$  on the x-axis. A value of 0.25 means that the firm on y-axis had significantly higher MES than the firm on the x-axis on 25% of the days.

### 2.6.2 Buckets

In this section, we apply the bucketing procedure to the 94 financial institutions for three systemic risk measures, the MES, the %SRISK and the  $\Delta\text{CoVaR}$ , which were defined in Examples 2, 3 and 4, respectively. By applying the bucketing procedure, we test whether an absolute ranking can be distinguished. If no absolute ranking can be distinguished, we want to test whether we can, at least, identify buckets of firms that are indistinguishable from each other within the bucket, but distinguishable from firms belonging to lower ranked buckets. The three systemic risk measures are affected differently by the estimation risk, and are also likely to differ in the ordering of their point estimates (Benoit, Colletaz, and Hurlin, 2014). As a consequence, different risk measures can lead to different rankings.

We estimate the bucket allocation for the MES, %SRISK and  $\Delta\text{CoVaR}$  on eight pre-determined dates coinciding with those considered in Brownlees and Engle (2012). A firm is included in the ranking at a certain date, if the firm still exists and if there are at least 1,000 observations up until that date. Table 2.3 displays the results of the bucketing procedure, with  $\alpha = 0.05$ , for two days. The results for the remaining days are deferred to Appendix B. The firms are first ranked in terms of their bucket, and within buckets we order the firms in descending value of their risk measure estimate, even though there is no statistical evidence that their risk is statistically different. We then report the ten highest ranked firms, as is done in Brownlees and Engle (2012). For each firm, we report the point estimate, as well as the allocated bucket according to the FWE and the FDR

**Table 2.3:** Bucket Allocation – Top 10

Tick	MES			Tick	%SRISK			Tick	$\Delta\text{CoVaR}$		
	FWE	FDR	Est.		FWE	FDR	Est.		FWE	FDR	Est.
30-06-2008											
LEH	1	1	10.287	C	1	1	0.152	FITB	1	1	3.480
MBI	1	1	9.781	BAC	2	2	0.091	HBAN	1	1	2.874
CIT	1	1	8.111	JPM	2	3	0.081	LEH	1	1	2.685
WM	1	1	7.459	MER	2	3	0.078	KEY	1	1	2.372
PFG	1	1	6.563	MS	3	4	0.073	RF	1	1	2.330
ABK	1	2	7.806	FRE	4	5	0.065	C	1	1	2.277
FITB	1	2	7.733	FNM	4	6	0.063	STI	1	1	2.034
C	1	2	5.816	AIG	4	7	0.057	BBT	1	1	2.017
FRE	1	2	5.713	GS	5	7	0.056	AIG	1	1	2.008
MER	2	2	6.248	LEH	6	8	0.052	MI	1	1	1.951
30-01-2009											
STT	1	1	22.188	JPM	1	1	0.153	AFL	1	1	9.049
C	1	1	20.884	C	2	2	0.142	PNC	1	1	8.266
HBAN	1	1	20.775	BAC	3	3	0.129	STT	1	1	6.891
FITB	1	1	19.821	WFC	4	4	0.093	FITB	1	1	6.414
PNC	1	1	19.817	AIG	5	5	0.063	BAC	1	1	5.974
AFL	1	1	19.499	GS	5	6	0.061	ACAS	1	1	5.537
LNC	1	1	19.032	MS	6	7	0.046	ALL	1	1	5.487
BAC	1	1	18.491	MET	6	8	0.036	WFC	1	1	5.399
HIG	1	1	17.415	PRU	7	9	0.034	STI	1	1	5.258
PFG	1	1	17.097	HIG	8	10	0.022	C	1	1	5.139

*Note:* This table provides the ranking estimated by the FWE and FDR controlling methods, based on the MES, %SRISK and  $\Delta\text{CoVaR}$  risk measures. We show only the top 10 of firms sorted by assigned bucket.

method.

The results suggest that it is indeed difficult to find significant differences between the estimated risk measures. Although point estimates may vary considerably, they are not necessarily statistically different. In general, and in line with theory, we find that the FDR rejects more frequently, and we obtain smaller buckets compared to the FWE. In June 2008, the precision of the MES estimates allows for a division of the top 10 risky firms into two buckets. The size of the most risky bucket using FDR is five firms, compared to nine for the FWE.

The point estimates in our ranking are not monotonically decreasing. For instance, in June 2008, based on the FDR method, we find that ABK is in a lower bucket than PFG, despite a higher point estimate. This is a direct consequence of the one-directional approach, and is a feature shared with the MCS (see Hansen, Lunde, and Nason, 2011, Table V, where a model with lower average loss is excluded from the MCS). Although PFG has a lower point estimate, its estimation uncertainty is far greater. As such, the procedure cannot reject that its risk is smaller than that of for instance LEH, whereas we can reject that same hypothesis for ABK. Hence, firms with large estimation uncertainty are prudently allocated to high-risk buckets.

The procedure rejects more frequently for the %SRISK, finding a total of six or eight buckets for the top 10 firms. The reason for this is that the liabilities and the market value of the firm, introduced in the definition of the SRISK (see Example 3), add variability between the different point estimates without adding additional estimation risk.

## 2. RISK MEASURE INFERENCE

**Table 2.4:** Number of estimated buckets

Date	#Firms	FWE Controlling Procedure					Significance Level					FDR Controlling Procedure				
		30%	20%	10%	5%	1%	30%	20%	10%	5%	1%	30%	20%	10%	5%	1%
MES																
30-03-2007	83	5	5	4	4	3	7	7	6	4	3					
29-06-2007	83	5	5	3	3	3	7	5	4	4	3					
31-12-2007	81	5	5	4	4	3	13	10	8	5	5					
29-02-2008	82	6	5	5	4	4	15	12	8	5	4					
30-06-2008	82	6	5	5	5	5	11	11	7	6	5					
29-08-2008	81	8	8	7	6	5	21	16	10	8	7					
30-01-2009	73	6	6	5	4	4	20	15	9	7	5					
30-06-2010	75	5	4	3	3	3	9	7	5	5	4					
%SRISK																
30-03-2007	14	6	6	6	5	4	7	7	6	6	5					
29-06-2007	13	6	6	5	4	4	11	10	7	6	4					
31-12-2007	36	17	17	13	11	11	22	21	17	15	12					
29-02-2008	37	17	16	15	13	12	26	25	19	17	14					
30-06-2008	39	20	18	17	15	12	37	37	26	21	17					
29-08-2008	36	16	15	14	13	10	34	33	25	18	15					
30-01-2009	53	33	31	29	29	29	49	49	45	39	31					
30-06-2010	37	19	18	16	15	11	31	31	22	20	15					
$\Delta$ CoVaR																
30-03-2007	83	2	2	2	2	1	1	1	1	1	1					
29-06-2007	83	1	1	1	1	1	1	1	1	1	1					
31-12-2007	81	2	2	1	1	1	2	2	2	2	2					
29-02-2008	82	2	2	2	2	1	2	2	2	2	2					
30-06-2008	82	3	3	1	2	2	3	3	2	2	2					
29-08-2008	81	2	1	2	2	2	2	2	2	2	2					
30-01-2009	73	2	2	2	2	2	4	4	4	3	3					
30-06-2010	75	1	1	1	1	1	2	2	2	2	2					

*Note:* This table reports the sensitivity of the procedures to the level of FWE and FDR that is controlled. We show the total number of firm and the number of buckets they are assigned to.

In fact, in January 2009 we find an absolute ranking using the FDR method, where each firm has statistically different risk. Similar to the MES, in our sample it is difficult to statistically distinguish firms based on  $\Delta$ CoVaR. The  $\Delta$ CoVaR is defined as the product of a conditional VaR and a quantile regression parameter (see Example 4). Most of the estimation risk comes from the quantile regression. For instance, the highest point forecast of  $\Delta$ CoVaR is 9.05 for AFL, but its bootstrap standard deviation is close to 4. In an unreported simulation, we find that even if the true DGP is exactly the one assumed here, the standard deviation of the  $\Delta$ CoVaR is still on average over 40% of its value. These results are in line with those obtained in another context by Guntay and Kupiec (2015). Replacing the quantile estimate  $\gamma_\alpha$  of Example 4 with an OLS estimate significantly reduces the uncertainty, leading to buckets of sizes in between those of MES and %SRISK.

In Table 2.4, we investigate the sensitivity of the bucketing procedure to the significance level chosen. We report the total number of estimated buckets on each of the eight days, at five different significance levels. The Model Confidence Set (Hansen, Lunde, and Nason, 2011), on which our procedure is based, is commonly estimated using confidence levels upwards of 20%. We consider 30, 20, 10, 5 and 1%, for both FWE and FDR.

As a reference, the second column of Table 2.4 gives the total number of firms under consideration, providing a cap on the number of buckets possible.

As rejection occurs more frequently with higher significance levels, the number of buckets is increasing with the significance level. The FDR procedure detects more buckets than the FWE for each significance level and each risk measure. For instance, for the MES, the FDR procedure estimates up to twice as many buckets than the FWE. With the %SRISK, the FDR procedure using high confidence levels comes close to absolute rankings, with the total number of buckets only slightly lower than the number of firms. Even at very stringent levels, we get interesting rankings with buckets that do not contain more than three or four firms. Finally, significance levels of 30% still do not help with disentangling the  $\Delta\text{CoVaR}$  of different firms on these dates. This reaffirms the uncertainty in the quantile regression estimates.

## 2.7 Conclusion

This paper introduces a bootstrap-based comparison test of two risk measures, as well as an iterative procedure to produce a grouped ranking of  $N > 2$  assets or firms, given their conditional risk measures. These tests can be applied to a wide variety of conditional risk measures, while taking into account their estimation risk. Simulation results on VaR and MES forecasts suggest that the pairwise comparison test has good properties in finite samples, both in terms of size and power. Since the bucketing procedure is clearly a multiple testing problem, we propose two versions, one controlling the FWE rate, and one controlling the FDR rate. Simulations show that both set-ups do control their respective rates, and illustrate the trade-off of using either method depending on the size of the problem.

In the empirical application, we apply the pairwise comparison test to the MES estimates of sixteen U.S. G- and D-SIBs. This application points out the advantages of the comparison of *conditional* risk measures. We highlight the importance of conditional testing, as we observe great time-variation in conditional MES estimates, and from one week to the next, firms' relative ranking often changes. We find that, on most days, due to estimation uncertainty in MES, we cannot distinguish firms in terms of their riskiness. On average across all pairs, we can statistically distinguish firms on 16.4% of days.

We applied the bucketing procedure for three popular systemic risk measures, namely the MES, the  $\Delta\text{CoVaR}$  and the SRISK. In our sample, we find that for both versions of the procedure, the MES and  $\Delta\text{CoVaR}$  are estimated with too much uncertainty to reject equality often. For most of the eight dates considered in the application, the first thirty firms belong to the same bucket of riskiest firms. Consequently, ranking firms on the basis of point forecasts of MES and  $\Delta\text{CoVaR}$  may be problematic. However, when applied on %SRISK, our bucketing procedure is able to identify a meaningful ranking of buckets containing equally risky firms in each bucket. This result is mainly due to the

## 2. RISK MEASURE INFERENCE

---

differences observed in the liabilities and the market value of the financial institutions over the period 2000-2012. Since the liabilities and market values are not estimated, these differences add cross-sectional variability in the systemic risk measures, without adding additional estimation risk. Our results clearly illustrate the importance of taking into account the estimation risk when establishing a ranking of the financial institutions according to their systemic risk.

## Appendix A: Company Tickers

Depositories(29)		Insurance (32)	
BAC	Bank of America	ABK	Ambac Financial Group
BBT	BB&T	AET	Aetna
BK	Bank of New York Mellon	AFL	Aflac
C	Citigroup	AIG	American International Group
CBH	Commerce Bancorp	AIZ	Assurant
CMA	Comerica inc	ALL	Allstate Corp
HBAN	Huntington Bancshares	AOC	Aon Corp
HCBK	Hudson City Bancorp	WRB	W.R. Berkley Corp
JPM	JP Morgan Chase	BRK	Berkshire Hathaway
KEY	Keycorp	CB	Chubb Corp
MI	Marshall & Ilsley	CFC	Countrywide Financial
MTB	M&T Bank Corp	CI	CIGNA Corp
NCC	National City Corp	CINF	Cincinnati Financial Corp
NTRS	Northern Trust	CNA	CNA Financial corp
NYB	New York Community Bancorp	CVH	Coventry Health Care
PBCT	Peoples United Financial	FNF	Fidelity National Financial
PNC	PNC Financial Services	GNW	Genworth Financial
RF	Regions Financial	HIG	Hartford Financial Group
SNV	Synovus Financial	HNT	Health Net
SOV	Sovereign Bancorp	HUM	Humana
STI	Suntrust Banks	LNC	Lincoln National
STT	State Street	MBI	MBIA
UB	Unionbancal Corp	MET	Metlife
USB	US Bancorp	MMC	Marsh & McLennan
WB	Wachovia	PFG	Principal Financial Group
WFC	Wells Fargo & Co	PGR	Progressive
WM	Washington Mutual	PRU	Prudential Financial
WU	Western Union	SAF	Safeco
ZION	Zion	TMK	Torchmark
		TRV	Travelers
		UNH	Unitedhealth Group
		UNM	Unum Group
Broker-Dealers (10)		Others (23)	
AGE	A.G. Edwards	ACAS	American Capital
BSC	Bear Stearns	AMP	Ameriprise Financial
ETFC	E-Trade Financial	AMTD	TD Ameritrade
GS	Goldman Sachs	AXP	American Express
LEH	Lehman Brothers	BEN	Franklin Resources
MER	Merill Lynch	BLK	Blackrock
MS	Morgan Stanle	BOT	CBOT Holdings
NMX	Nymex Holdings	CBG	C.B. Richard Ellis Group
SCHW	Schwab Charles	CBSS	Compass Bancshares
TROW	T.Rowe Price	CIT	CIT Group
		CME	CME Group
		COF	Capital One Financial
		EV	Eaton Vance
		FITB	Fifth Third bancorp
		FNM	Fannie Mae
		FRE	Freddie Mac
		HRB	H&R Block
		ICE	Intercontinental Exchange
		JNS	Janus Capital
		LM	Legg Mason
		NYX	NYSE Euronext
		SEIC	SEI Investments Company
		SLM	SLM Corp

## 2. RISK MEASURE INFERENCE

### Appendix B: Bucket Allocation Top 10

MES			%SRISK			$\Delta\text{CoVaR}$					
Tick	FWE	FDR	Est.	Tick	FWE	FDR	Est.	Tick	FWE	FDR	Est.
30-03-2007											
LEH	1	1	3.540	MS	1	1	0.212	LEH	1	1	1.094
BSC	1	1	3.469	FRE	2	2	0.146	AGE	1	1	1.053
MS	1	1	3.446	FNM	2	2	0.134	MS	1	1	0.979
AMTD	1	1	3.427	MER	2	3	0.105	BSC	1	1	0.952
ETFC	1	1	3.260	LEH	3	3	0.104	BEN	1	1	0.904
AGE	1	1	3.221	GS	3	3	0.103	MER	1	1	0.900
JNS	1	1	3.159	BSC	4	4	0.091	GS	1	1	0.894
GS	1	1	3.158	MET	5	5	0.034	LM	1	1	0.844
BEN	1	1	3.155	HIG	5	5	0.019	C	1	1	0.831
MER	1	1	2.816	PRU	5	5	0.018	BBT	1	1	0.821
29-06-2007											
AMTD	1	1	3.009	MS	1	1	0.179	MBI	1	1	0.944
BSC	1	1	2.799	FRE	1	2	0.164	GS	1	1	0.848
MER	1	1	2.630	MER	1	2	0.155	C	1	1	0.814
MBI	1	1	2.619	BSC	2	3	0.111	LEH	1	1	0.800
SCHW	1	1	2.601	LEH	2	3	0.109	MER	1	1	0.790
GS	1	1	2.544	FNM	2	3	0.094	JPM	1	1	0.762
LEH	1	1	2.431	GS	2	4	0.092	BSC	1	1	0.754
ETFC	1	1	2.350	MET	3	5	0.033	SCHW	1	1	0.728
TROW	1	1	2.235	PRU	3	5	0.026	EV	1	1	0.712
MS	1	1	2.197	HIG	3	5	0.021	HRB	1	1	0.685
31-12-2007											
ETFC	1	1	9.406	C	1	1	0.166	MBI	1	1	3.173
MBI	1	1	9.288	MER	2	2	0.102	ABK	1	1	2.377
ABK	1	1	7.528	MS	3	3	0.092	ETFC	1	1	2.288
FRE	1	1	6.960	FRE	3	4	0.084	SLM	1	1	2.123
CFC	1	1	6.036	FNM	3	5	0.077	NCC	1	1	1.741
WM	1	1	5.815	GS	3	5	0.076	WM	1	1	1.731
CITa	1	1	5.378	LEH	4	6	0.061	C	1	1	1.688
FNM	1	1	5.276	JPM	4	6	0.060	FRE	1	1	1.621
HBAN	1	1	4.629	BAC	4	7	0.040	FITB	1	1	1.579
MS	1	2	5.148	BSC	5	7	0.034	BBT	1	1	1.548
29-02-2008											
ABK	1	1	8.266	C	1	1	0.151	AIG	1	1	1.774
MBI	1	1	7.644	MER	2	2	0.089	CNA	1	1	1.715
LEH	1	1	5.279	MS	2	3	0.083	MI	1	1	1.674
CIT	1	1	5.279	FNM	2	4	0.078	MER	1	1	1.666
WM	1	1	5.150	JPM	2	4	0.077	C	1	1	1.575
MER	1	1	5.105	BAC	2	5	0.074	RF	1	1	1.571
FRE	1	1	4.816	FRE	3	5	0.074	LEH	1	1	1.561
CNA	1	1	4.783	GS	3	6	0.061	JPM	1	1	1.503
MI	1	1	4.613	LEH	4	7	0.045	EV	1	1	1.467
BSC	1	1	4.607	BSC	5	8	0.038	HBAN	1	1	1.402

MES			%SRISK			$\Delta\text{CoVaR}$					
Tick	FWE	FDR	Est.	Tick	FWE	FDR	Est.	Tick	FWE	FDR	Est.
29-08-2008											
FRE	1	1	13.593	C	1	1	0.134	AIG	1	1	2.972
FNM	1	1	13.399	JPM	2	2	0.097	LEH	1	1	2.495
ABK	1	1	12.942	BAC	2	2	0.096	MI	1	1	2.491
LEH	1	1	12.422	MER	3	3	0.074	FRE	1	1	2.429
MBI	1	2	9.642	FRE	3	3	0.074	MER	1	1	2.306
AIG	1	2	8.656	AIG	3	4	0.072	RF	1	1	2.304
RF	1	2	7.682	FNM	4	4	0.070	KEY	1	1	2.049
MER	2	2	8.633	MS	5	5	0.067	FNM	1	1	2.006
BAC	2	2	7.227	GS	6	6	0.060	SNV	1	1	1.967
WM	2	2	6.969	LEH	7	7	0.051	C	1	1	1.939
30-06-2010											
ABK	1	1	7.617	C	1	1	0.164	MTB	1	1	1.927
CBG	1	1	6.968	BAC	1	1	0.161	BEN	1	1	1.882
MI	1	1	6.792	JPM	2	2	0.138	TROW	1	1	1.788
JNS	1	1	6.752	AIG	3	3	0.086	EV	1	1	1.674
ETFC	1	1	6.680	MS	4	4	0.071	MI	1	1	1.655
ACAS	1	1	6.619	WFC	4	5	0.048	AFL	1	1	1.609
LNC	1	1	6.568	MET	5	5	0.046	AXP	1	1	1.575
PFG	1	1	6.319	GS	5	5	0.044	CINF	1	1	1.512
MBI	1	1	6.229	PRU	5	6	0.042	GS	1	1	1.492
AMP	1	1	6.188	HIG	6	7	0.031	SCHW	1	1	1.472

# Chapter 3

## Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity\*

### 3.1 Introduction

The availability of high-frequency data and a large variety of estimators harnessing their information, has enhanced the understanding of the covariation between financial time series. The estimation of covariance matrices is vital in many interesting and important financial, economic and statistical applications. However, most of these do not only require the estimated matrix to be accurate, but also positive semidefinite (PSD). Unfortunately, in search of the former, many proposed estimators have sacrificed the latter (e.g. Hayashi and Yoshida, 2005; Lunde, Shephard, and Sheppard, 2015; Zhang, 2011).

Estimation of realized covariance at ultra high frequencies is made difficult by two empirical phenomena which induce biases in the estimates. First, the presence of market micro-structure noise (e.g. bid-ask bounce), and second, non-synchronous trading. Until recently, lower frequency returns, for instance sampled every 20 minutes, were used to avoid these biases. Now, estimators that are robust to both problems have been proposed, and the only limit to the frequency of data one faces is due to the fact that the observations have to be synchronized.

The multivariate realized kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) uses refresh-time sampling to synchronize data. Although this synchronization technique is quite efficient in terms of observations retained, its problem is that the number of observations is always determined by the least frequently traded asset. To diminish that effect, several papers try to make more efficient use of data by splitting up estimation into subsets of the data. Hautsch, Kyj, and Oomen (2012) propose a method

---

\*This chapter is co-authored with Kris Boudt(Vrije Universiteit Brussel and Vrije Universiteit Amsterdam), Sébastien Laurent (Aix-Marseille University), Asger Lunde (Aarhus University) and Orimar Sauri (Aarhus University).

that applies the multivariate kernel to separate groups of liquid and illiquid assets, and combines these estimates into a full matrix. Aït-Sahalia, Fan, and Xiu (2010) synchronize pairs and use the so-called polarization result to estimate the covariance bivariately, using univariate estimators. Lunde, Shephard, and Sheppard (2015) use a class of Composite estimators, which estimate the variances univariately, and use bivariate sampling to obtain correlations. Fan, Li, and Yu (2012) use both the polarization and pairwise refresh-time techniques. These estimators increase efficiency by using more observations, but sacrifice positive semidefiniteness in the process.

The contribution of this paper is a method that estimates each element sequentially, whilst ensuring the final estimate to be positive semidefinite. Because our estimator of the Integrated Covariance is based on the Cholesky decomposition, we denote it CholCov. By applying an orthogonal decomposition to the covariance matrix we reduce the estimation from a  $d$ -dimensional matrix to a large number of bivariate estimations on transformed series, and obtain a well-conditioned PSD matrix. However, due to the necessary transformations, we cannot sample over just two series for each element, but have to iteratively sample over an increasing set of series. Each bivariate estimate on the transformed series can be done with any estimator that is robust to the biases induced by ultra high-frequency data.

We derive the joint asymptotic properties of the realized covariance based CholCov estimator under the assumption of no microstructure noise and constant volatility. As the different elements of the CholCov are estimated using different time-grids, this is a non-trivial and new result. We find that the asymptotic covariance of the scaled CholCov parameter estimates is proportional to that of the classical realized covariance on synchronous returns and that the proportionality factor is a function of the overlap between the different grids used to estimate the CholCov parameters.

In an extensive simulation study, we find our estimator performs at least as well as its main competitors, and it offers significant improvements when estimating vast-dimensional matrices and/or in scenarios with high heterogeneity in trading frequencies. Its performance is similar to the Composite versions without resorting to an arbitrary regularization technique to make the estimate positive semidefinite. Not only is it positive semidefinite, but simulations also show that the CholCov is on average better conditioned than its competitors, especially for larger dimensions. Finally, we conduct a simulation study which confirms that the asymptotic results also provide a good description of the finite sample distribution of the realized covariance based CholCov estimates in the absence of microstructure noise.

For the empirical application, we use our estimator to forecast portfolio Value-at-Risk and to estimate industry exposures for 52 assets. We forecast the covariance matrix using the CholCov, and compare the forecasting accuracy to those obtained using dynamic specifications on daily returns only. We find that models utilizing the CholCov in their estimation greatly improve both unconditional coverage and independence between Value-

at-Risk violations, and that they lead to the most informative sector exposure forecasts amongst our models considered.

The paper is structured as follows. In Section 3.2 we present the theoretical setup and outline the decomposition. Section 3.3 first discusses practical issues in preparing the data for estimation, and then presents the algorithm that delivers the estimate. In the next section we derive its asymptotic properties. Section 3.5 presents a summary of some extensive Monte Carlo simulations designed to highlight the improvements due to more efficient data-sampling. Sections 3.6 and 3.7 provide the empirical application on the forecasting of portfolio Value-at-Risk and the estimation of industry exposures using the CholCov. Finally, Section 3.8 concludes.

## 3.2 Theoretical setup

Our aim is to accurately estimate the Integrated Covariance (ICov) matrix of a  $d$ -dimensional Brownian semimartingale process  $Y = (Y^{(1)}, \dots, Y^{(d)})'$ . The measurement is complicated by the fact that the component processes are observed at irregular and non-synchronous time points, and that the price process of interest is observed with measurement error. The actual observed log-prices are denoted  $X = (X^{(1)}, \dots, X^{(d)})'$ . Let the set of all series be denoted  $\mathcal{D}$ , with subsets  $\mathcal{d} \subseteq \mathcal{D}$ . Each component process can be observed at different time points over the interval  $[0, T]$ . For simplicity we take  $T = 1$  in this paper. For every  $i = 1, \dots, d$ , we denote by  $T_i := \left\{ 0 = t_1^{(i)} < \dots < t_{N_i}^{(i)} \leq 1 \right\}$  the observation times of the  $i$ -th element in  $X$ .  $X$  is driven by the efficient log-price  $Y$ , a Brownian semimartingale defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq 1}, \mathbb{P})$  satisfying the usual conditions of right-continuity and completeness:

$$Y(t) = \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s), \quad (3.1)$$

where  $\mu$  is a  $d \times 1$  predictable locally bounded drift process,  $W$  is a  $d$ -dimensional vector of independent Brownian motions and  $\sigma$  a  $d \times d$  càdlàg process such that  $\Sigma(s) = \sigma(s)\sigma'(s)$  is the spot covariance matrix of  $Y$  at time  $s$ .

At very high frequencies, microstructure noise leads to a departure from the Brownian semimartingale. As a result, we do not observe  $Y$ , but instead a process  $X$  defined as:

$$X_t^{(i)} = Y_t^{(i)} + \epsilon_t^{(i)}, \quad (3.2)$$

where  $\epsilon_t^{(i)}$  is microstructure noise and  $Y_t^{(i)}$  is the  $i$ -th component of  $Y$ . In this paper,  $\epsilon_t = (\epsilon_t^{(1)}, \dots, \epsilon_t^{(d)})'$  is assumed to be covariance stationary and have zero mean. We achieve noise robustness by means of pre-averaging and refer the reader to Christensen, Kinnebrock, and Podolskij (2010) for the precise assumptions on the serial correlation of the noise and its cross-correlation with the latent price.

Our parameter of interest is the integrated covariance over the unit interval:

$$\text{ICov} = \int_0^1 \Sigma(s) ds, \quad (3.3)$$

which under (3.1) coincides with the quadratic covariation of  $Y$ , i.e.

$$[Y](1) = \text{plim}_{n \rightarrow \infty} \sum_{j=1}^n (Y_{j/n} - Y_{(j-1)/n})(Y_{j/n} - Y_{(j-1)/n})' = \text{ICov}.$$

Estimation of the off-diagonal elements of the ICov requires synchronization of the data. Two approaches are popular. One is to synchronize all observations jointly, but this has the disadvantage of letting the resulting observation frequency be determined by the least liquid asset (e.g. Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2011). The second approach is to estimate the off-diagonal elements of the ICov separately using synchronized pairs of price processes, but the result is not ensured to be positive semidefinite (e.g. Lunde, Shephard, and Sheppard, 2015). We propose a third approach that uses a factorization of the spot covariance matrix, combining the advantages of both methods.

### 3.2.1 Decomposition

For the estimation of the Integrated Covariance, it will reveal useful to decompose each spot covariance matrix into a symmetric factorization of square matrices. Such a factorization is ensured to be positive semidefinite. Additionally, we require that the parameters defining the spot covariance between  $k$  and  $l$  do not depend on the parameters defining  $m$  and  $n$  (with  $m > k$  and  $n > l$ ). The latter condition ensures that we can estimate the parameters sequentially, which allows more efficient data usage. This is the case for amongst others the Cholesky decomposition, which we will focus on.

The Cholesky decomposition is also used in the volatility context in Chiriac and Voev (2011) and Tsay (2010). Chiriac and Voev (2011) decompose the covariance matrix and use time series models to forecast the elements. This way they ensure positive semidefiniteness of covariance forecasts. Tsay (2010) uses the re-parametrization in a multivariate GARCH framework. Palandri (2009) uses an iterative Cholesky to reduce the dimension of multivariate GARCH models. The use of the Cholesky decomposition is the only similarity with our work. Their methods and goals are different from ours. We construct an ex-post covariance estimator, not a time-series volatility model.

By means of the Cholesky decomposition, the spot covariance matrix can be uniquely split into

$$\Sigma(s) = H(s)G(s)H(s)', \quad (3.4)$$

where  $H(s)$  is a lower diagonal matrix with ones on the diagonal, and  $G(s)$  a diagonal

matrix. Specifically,

$$H(s) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h_{21}(s) & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{d1}(s) & h_{d2}(s) & \cdots & 1 \end{bmatrix} \quad G(s) = \begin{bmatrix} g_{11}(s) & 0 & \cdots & 0 \\ 0 & g_{22}(s) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{dd}(s) \end{bmatrix}.$$

For instance, omitting the time-dependence, for  $d = 3$ :

$$\Sigma = HGH' = \begin{bmatrix} g_{11} & h_{21}g_{11} & h_{31}g_{11} \\ h_{21}g_{11} & h_{21}^2g_{11} + g_{22} & h_{21}h_{31}g_{11} + h_{32}g_{22} \\ h_{31}g_{11} & h_{21}h_{31}g_{11} + h_{32}g_{22} & h_{31}^2g_{11} + h_{32}^2g_{22} + g_{33} \end{bmatrix}. \quad (3.5)$$

It follows from this that the  $H$  and  $G$  matrices are linked to the elements of  $\Sigma$  as follows:

$$g_{kk} = \Sigma_{kk} - \sum_{m=1}^{k-1} h_{km}^2 g_{mm} \quad (3.6)$$

$$h_{kl} = \frac{1}{g_{ll}} \left( \Sigma_{kl} - \sum_{m=1}^{k-1} h_{km} h_{lm} g_{mm} \right), \quad (3.7)$$

for  $k > 1$  with  $g_{11} = \Sigma_{11}$ . The  $(k, l)$ -th element of the covariance matrix only depends on  $g_{mm}$  and  $h_{np}$  with  $m, n, p \leq \max(k, l)$ . The elements can therefore be estimated sequentially.

## 3.3 The Estimator

In order to apply the Cholesky decomposition in the estimation of the ICov, we first need to define how the returns are computed and, since the procedure will be sequential, how the covariance matrix is ordered. Subsection 3.3.1 reviews the concept of refresh time returns and Subsection 3.3.2 introduces the liquidity criterion that we use to structure the covariance matrix. Based on these ingredients, Subsection 3.3.3 defines the CholCov estimator.

### 3.3.1 Dealing with asynchronicity

In high-frequency data, assets are traded at irregular intervals and seldom simultaneously. Moreover, not all stocks are equally liquid, such that the number of fresh prices within a fixed time interval may substantially differ across firms. One way to synchronize the data is by means of refresh-time sampling, as proposed by Harris, McInisch, Shoesmith, and Wood (1995). It picks refresh-times at which all assets have traded at least once since the last refresh-time point. Refresh-times for  $t \in [0, 1]$  and the set of assets  $\mathcal{b} \subseteq \mathcal{D}$  are defined as follows. The first refresh-time occurs at  $\tau_1^{\mathcal{b}} = \max(t_1^{(i)}, \forall i \in \mathcal{b})$ , and subsequent

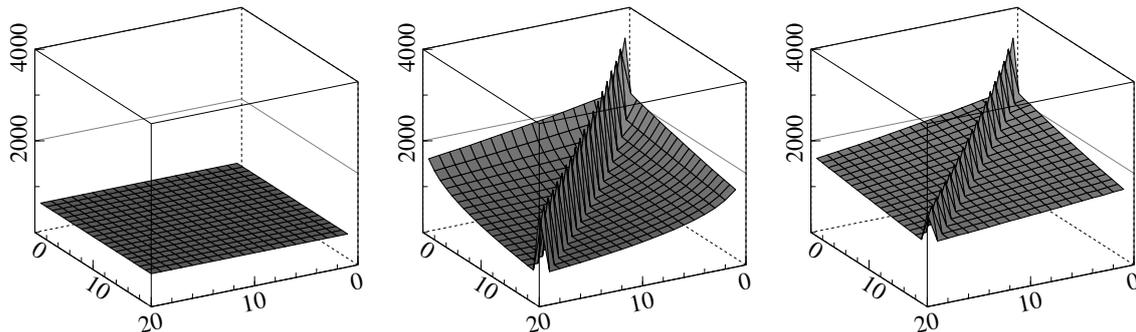
refresh-times at  $\tau_{j+1}^{\mathcal{b}} = \min(t_i^{(i)} | t_i^{(i)} > \tau_j^{(i)}, \forall i \in \mathcal{b})$ . The resulting refresh time grid is of course a function of the series over which they are sampled. In our setting, stocks are ranked on liquidity and of particular interest will be the refresh time grid obtained for the first  $k$  assets, i.e.  $\mathcal{b} = \{1, \dots, k\}$ . Denote the set of corresponding refresh times as  $\mathcal{T}_k := \{0 = \tau_0^k < \tau_1^k < \dots < \tau_{\mathcal{N}^k}^k = 1\}$ . We denote the returns of grid  $\mathcal{T}_k$  by  $r_j(\mathcal{T}_k) = X_{\tau_j^k} - X_{\tau_{j-1}^k}$ , with  $j = 1, \dots, \mathcal{N}^k$ . Finally, denote the durations as  $\Delta_j(\mathcal{T}_k) = \tau_j^k - \tau_{j-1}^k$ . Throughout the paper, for clarity of notation, the superscript may be omitted if it is clear which grid is being discussed, or in general statements concerning every grid.

As a result of the sampling scheme,  $\mathcal{N}^k \leq \min_{i \in \mathcal{b}} N_i$ , and may be a lot smaller depending on the trading pattern and number of series under consideration. Hautsch, Kyj, and Oomen (2012) illustrate that the percentage data loss can exceed 90% when the number of assets becomes large and arrival rates are unequal. The sample size is largely determined by the least liquid - in the sense of frequent trades - stocks. Including a single illiquid stock may therefore drastically reduce the estimation efficiency of all elements, including those for which a lot of data is available.

This problem is circumvented by the composite estimation technique used in Lunde, Shephard, and Sheppard (2015) and Fan, Li, and Yu (2012). The data loss is reduced, but at the expense of positive semidefiniteness, one of the defining properties of a covariance matrix. Moreover, many applications, such as principal components analysis and portfolio optimization, critically rely on it. Therefore the question arises how to project the symmetric matrix onto the space of PSD matrices.

There exist many ways to transform the matrix to a PSD alternative. A general method called shrinkage is proposed in Ledoit and Wolf (2003). In the Realized Covariance setting Barndorff-Nielsen and Shephard (2004b) and Fan, Li, and Yu (2012) set all negative eigenvalues in the spectral decomposition to zero. Hautsch, Kyj, and Oomen (2012) impose more structure by employing eigenvalue cleaning, a random matrix theory technique similar in intuition to shrinkage. While such eigenvalue cleaning may increase the efficiency of the covariance matrix estimate, it still remains that for portfolio optimization purposes the dependence of the optimized portfolio on the eigenvalue cleaning is highly undesirable. Schmelzer and Hauser (2013) discuss the severe implications of a negative eigenvalue in the context of portfolio optimization and point to the above mentioned strategies to get rid of them. They conclude, however that, from their experience, a careful analysis of the estimation process itself adds far more value to the trading strategy.

Here, we use the Cholesky decomposition to exploit the heterogeneity in trading intensities and estimate the different parameters sequentially, using as much data as possible. The decomposition holds for any PSD matrix, which is by definition true for the covariance matrix, and the recombined results will also be PSD. However, ensuring positive semidefiniteness does come at a cost. We cannot synchronize just pairs of data, but have to sample over a growing set of series. The first elements are estimated by sampling over two series, but the last elements to be estimated require a common grid on all series.

**Figure 3.1:** Number of observations available for each element in estimation of  $\Sigma$ .

*Note:* These graphs depict simulation based number of observations available for the estimation of elements in a  $20 \times 20$  covariance matrix for different sampling schemes. The observation frequencies follow independent poisson processes with parameter  $\lambda$  ranging uniformly from 10 and 15. Stocks are ranked based on  $\lambda$ . For more details, see Section 3.5. From left to right: Sampling over all series, sequential sampling and pairwise sampling.

To make optimal use of the data, it is therefore crucial to order the series in terms of decreasing liquidity.

To illustrate, Figure 3.1 plots a simulated example of the number of observations used for each element using the three different methods. We consider 20 assets, where each asset is slightly less liquid than the last, with trading activity of asset  $i$  governed by a Poisson process with intensity  $\lambda_i = 10 + i/4$ . The leftmost picture depicts a situation in which refresh-time sampling is applied to all assets at once. The rightmost picture depicts pairwise sampling, which is done for the Composite estimators of Lunde, Shephard, and Sheppard (2015). The diagonal elements are estimated sampling over just the single series, and therefore also depict the number of observations available for that asset. The off-diagonal elements are estimated using bivariate sampling. The data-loss is therefore minimal. The middle graph depicts sequential sampling, the technique used for our estimator. The diagonal elements are estimated in the same fashion as for the Composite estimator, but we require sampling over more series for the off-diagonal estimates. Specifically, for element  $(k, l)$  we synchronize series  $\{1, \dots, l, k\}$ . As such, the number of observations for all elements involving the first asset coincide with those for pairwise sampling, while the observation count for the other elements lies somewhere between the two other sampling approaches.

### 3.3.2 Ranking stocks based on liquidity

In order to best utilize the sequential estimation, we propose to sort the series in terms of liquidity. For the Regularization and Blocking ('RnB') estimator, Hautsch, Kyj, and Oomen (2012) sort series in terms of liquidity solely based on the number of observations

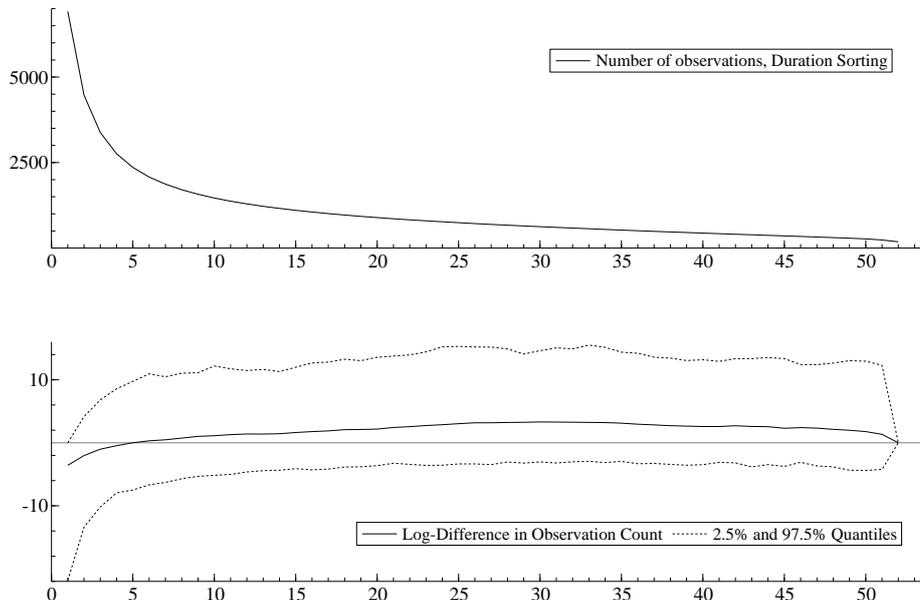
$N_i$ . However, there are many scenarios where this would not lead to a large number of observations when sampling over many series. Instead we use the squared duration as a liquidity criterion.

$$\sum_{j=0}^{N_i} \left( \Delta_j^{(i)} \right)^2, \quad (3.8)$$

where  $\Delta_j^{(i)}$  is each individual series' durations. This criterium is not just based on the number of intraday returns available, but also on the degree to which the returns are spread out over the day. We call our method duration sorting, as opposed to observation count sorting proposed in Hautsch, Kyj, and Oomen (2012).

Different sorting rules will lead to different sets being synchronized at any iteration in the sequential estimation. To illustrate the effect of duration sorting compared to observation count sorting, we preview results of our empirical application where we synchronize 52 series on 1499 days in Figure 3.2. Each day, the series are sorted according to the two methods. Next, for each of the methods, an increasing set of series is synchronized, from a single series, up to the full 52. As such, the x-axes in Figure 3.2 denote the number of series that is synchronized. For each set size, we therefore have a distribution over the different days. We plot the average amount of observations per set size for duration sorting in the top plot. In the bottom plot we display the mean and the 2.5 and 97.5% empirical quantiles of the log-difference of the number of observations between the two sampling techniques. As such, in the bottom graph the y-axis can be interpreted as percentage increases in number of observations by using duration sorting compared to observation count sorting.

The figure shows that duration sorting leads to a smoothing of available observations across the set of series. When synchronizing just the first one to six series, there are on average fewer observations with duration sorting. However, starting from the synchronization of the first seven series, up until the full set, the average number of observations using duration sorting is higher. This has large advantages. As the top graph shows, for small sets, there are plenty of observations available, while when more series are synchronized, observations are a lot less frequent. The number of observations using duration sorting is on average higher by 2 to 3%. Additionally, the distribution is skewed towards higher gains than losses, with the duration sorting offering observation increases of 10 to 15% in the 2.5% best cases, compared to losses of around 5% in the worst 2.5% cases. Finally, once all 52 series are synchronized there is no difference between the sorting rules, as the sets are the same for both.

**Figure 3.2:** Duration sorting versus observation count sorting.

*Note:* The top graph shows the number of observations available after synchronization over an increasing number of series from 1 to 52. The bottom graph shows the percentage gains in number of observations from using duration sorting compared to observation count sorting.

### 3.3.3 The CholCov

This section presents our estimator of the Integrated Covariance, which achieves positive semidefiniteness by means of the Cholesky decomposition. Instead of directly estimating the elements in  $\Sigma$ , a sequential regression algorithm is proposed to estimate the elements of the  $G$  and  $H$  matrices, which are then combined into a final estimate of  $\Sigma$ , denoted as the CholCov:

$$\text{CholCov} = \hat{H}\hat{G}\hat{H}', \quad (3.9)$$

with  $\hat{G}$  and  $\hat{H}$  the corresponding estimates for  $G$  and  $H$ . The computational cost of the algorithm is compensated by the expected gains in efficiency in the presence of asynchronicity.

The inspiration for the algorithm comes from rewriting the returns in a block of locally constant volatility,  $\Sigma_0 = H_0 G_0 H_0'$ , as a factor regression model. For simplicity, we assume the block corresponds to the  $[0,1]$  interval.<sup>1</sup> To see this, consider the vector of latent noise-free returns  $\tilde{r}_j(T) = Y_{t_j} - Y_{t_{j-1}}$ , with  $T = \{0 = t_1 < \dots < t_N = 1\}$ , which, under the previously stated assumptions and provided returns are sampled sufficiently

<sup>1</sup>The one day block simplifies the exposition and is also the one used in the implementation. Of course, the  $[0,1]$  interval can be subdivided in infinitely many blocks and the resulting CholCov estimate would be the sum over all these local CholCov estimates.

frequently such that the drift can be ignored, is normally distributed around zero with covariance matrix  $\Delta_j \Sigma_0$ . We can further define, omitting the grid-dependence,  $\tilde{f}_j = H_0^{-1} \tilde{r}_j \sim N(0, \Delta_j G_0)$ . Since  $H_0$  is triangular, we obtain that each component in the vector of returns  $\tilde{r}_j = H_0 \tilde{f}_j$ , is an explicit function of the components with a lower index. More precisely,  $\tilde{f}_j^{(1)}(T) = \tilde{r}_j^{(1)} \sim N(0, \Delta_j g_{11}^0)$ , and for  $k = 2, \dots, d$ ,

$$\begin{aligned} \tilde{r}_j^{(k)} &= h_{k1}^0 \tilde{f}_j^{(1)} + \dots + h_{k(k-1)}^0 \tilde{f}_j^{(k-1)} + \tilde{f}_j^{(k)} \\ \tilde{f}_j^{(k)} &\sim N(0, \Delta_j g_{kk}^0). \end{aligned} \quad (3.10)$$

The  $g_{kk}^0$  and  $h_{kl}^0$  elements are therefore simply the residual variances of the factors and the beta coefficients of a regression of  $\tilde{r}_j^{(k)}$  on  $\tilde{f}_j^{(k)}$ :

$$g_{kk}^0 = [\tilde{f}^{(k)}] \quad (3.11)$$

$$h_{kl}^0 = [\tilde{r}^{(k)}, \tilde{f}^{(l)}] / [\tilde{f}^{(l)}], \quad (3.12)$$

for  $l < k$ , where the  $[\cdot, \cdot]$  notation denotes the quadratic (co)variation of its element(s).

This suggests then to estimate the CholCov parameters iteratively on an expanding set of synchronized price series. Let  $\mathcal{T}_k$  be the set of refresh-times for the first  $k$ -assets and assume that  $\mathcal{T}_k \subseteq \mathcal{T}_j$  for  $j = 1, \dots, k$ .

### CholCov Estimation Algorithm

1. First sort the series in terms of decreasing liquidity according to (3.8), such that series 1 is the most liquid, and series  $d$  the least.
2. For  $k = 1, \dots, d$ :

Apply refresh-time on  $\mathcal{K} = \{1, \dots, k\}$  to obtain the grid  $\mathcal{T}_k$ .

Set  $\hat{f}_j^{(1)}(\mathcal{T}_k) = r_j^{(1)}(\mathcal{T}_k)$ .

For  $u = 1, \dots, k$  and  $v = 1, \dots, u - 1$ , compute:

$$\bar{g}_{uu}(\mathcal{T}_k) = \sum_{j=1}^{\mathcal{N}^k} \left[ \hat{f}_j^{(u)}(\mathcal{T}_k) \right]^2, \quad (3.13)$$

$$\bar{h}_{uv}(\mathcal{T}_k) = \frac{\sum_{j=1}^{\mathcal{N}^k} r_j^{(u)}(\mathcal{T}_k) \hat{f}_j^{(v)}(\mathcal{T}_k)}{\sum_{j=1}^{\mathcal{N}^k} \left[ \hat{f}_j^{(v)}(\mathcal{T}_k) \right]^2}, \quad (3.14)$$

where

$$\hat{f}_j^{(v)}(\mathcal{T}_k) := r_j^{(v)}(\mathcal{T}_k) - \sum_{w=1}^{v-1} \bar{h}_{vw}(\mathcal{T}_k) \hat{f}_j^{(w)}(\mathcal{T}_k). \quad (3.15)$$

3. Set the diagonal elements of  $\hat{G}$  and lower diagonal elements of  $\hat{H}$  as

$$\hat{g}_{kk} = \bar{g}_{kk}(\mathcal{T}_k) \quad (3.16)$$

$$\hat{h}_{kl} = \bar{h}_{kl}(\mathcal{T}_k), \quad (3.17)$$

for  $k = 1, \dots, d$  and  $l = 1, \dots, k - 1$ , and compute the (realized covariance based) CholCov as  $\hat{H}\hat{G}\hat{H}$ .

#### 3.3.4 Using CholCov in practise

The CholCov estimator given by (3.16)-(3.17) uses realized (co)variance estimates for the  $g_{kk}$  and  $h_{kl}$  parameters. This choice of estimator is no longer adequate in the presence of microstructure noise, where we only observe a noisy counterpart of  $\tilde{r}_j^{(k)}$  and  $\tilde{f}_j^{(k)}$ , namely the refresh return

$$r_j^{(k)} = \tilde{r}_j^{(k)} + \epsilon_j^{(k)} - \epsilon_{j-1}^{(k)}, \quad (3.18)$$

and the corresponding noise-contaminated factor

$$f_j^{(l)} = \tilde{f}_j^{(l)} + (\epsilon_j^{(l)} - \epsilon_{j-1}^{(l)}) - \sum_{a=1}^{l-1} h_{la}(\epsilon_j^{(a)} - \epsilon_{j-1}^{(a)}), \quad (3.19)$$

for  $l > 1$ . The CholCov estimates of  $g_{kk}$  and  $h_{kl}$  that we recommend to be used in practice therefore result from estimating the quadratic (co)variation in (3.11)-(3.12) with noise-robust integrated variance (IV) and beta estimators implemented on synchronized price series.<sup>2</sup> In the simulation study and empirical application we will use the Modulated Realized Covariance (MRC) pre-averaging estimator of Christensen, Kinnebrock, and Podolskij (2010) because of its robustness to microstructure noise and its ease of implementation and flexibility. We denote the corresponding estimate as the *mrcCholCov* or *CholCov* when there is no confusion possible.

Other approaches that are robust to microstructure noise include the two time scale estimator of Zhang (2011), the kernel estimator in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), the likelihood based methods in Ait-Sahalia, Fan, and Xiu (2010), the local spectral estimator in Bibinger and Reiß (2014), the local method of moments estimator in Bibinger, Hautsch, Malec, and Reiß (2014) as well as advances by Peluso, Corsi, and Mira (2014) and Shephard and Xiu (2014), among others.

For the estimation of  $h_{kl}$ , the CholCov estimator proposed above requires to synchronize the prices of assets  $1, \dots, k$ . When the factors are orthogonal, a consistent estimator for  $h_{kl}$  could also readily be obtained by synchronizing only the series  $1, \dots, l$  needed to compute the factor  $f^l$ , together with the observed prices for asset  $k$ . In finite samples, the factors are no longer orthogonal when using these grids on a smaller number of series.

---

<sup>2</sup>The realized beta is computed by estimating a  $2 \times 2$  realized covariance matrix  $\hat{\Sigma}$ . The realized beta is then defined as  $\hat{h}_{kl} = \hat{\Sigma}_{kl} / \hat{\Sigma}_{ll}$  (Andersen, Bollerslev, Diebold, and Wu, 2006).

Our simulation analysis shows that substantial efficiency gains are obtained by estimating  $h_{kl}$  on the refresh time grid for  $\{1, \dots, l, k\}$  versus the estimation using the grid obtained for  $\{1, \dots, k\}$ . The corresponding CholCov\* estimation algorithm is then as follows:

### CholCov\* Estimation Algorithm

1. First sort the series in terms of decreasing liquidity according to (3.8), such that series 1 is the most liquid, and series  $d$  the least.
2. *Step 1:*  
 Set  $f_j^{(1)}(T^1) = r_j^{(1)}(T^1)$ .  
 Estimate  $\hat{g}_{11}$  using an IV estimator on  $f_j^{(1)}(T^1)$ .
3. *Step 2:*  
 Apply refresh-time on  $b = \{1, 2\}$  and denote the corresponding grid by  $\mathcal{T}_b$ .  
 Estimate  $\hat{h}_{21}$  as the realized beta between  $f_j^{(1)}(\mathcal{T}_b)$  and  $r_j^{(2)}(\mathcal{T}_b)$ .  
 Estimate  $\hat{g}_{22}$  using an IV estimator on  $f_j^{(2)}(\mathcal{T}_b) = r_j^{(2)}(\mathcal{T}_b) - \hat{h}_{21}f_j^{(1)}(\mathcal{T}_b)$ .
4. *Step 3:*  
 Apply refresh-time on  $c = \{1, 3\}$  to obtain the grid  $\mathcal{T}_c$ .  
 Estimate  $\hat{h}_{31}$  as the realized beta between  $f_j^{(1)}(\mathcal{T}_c)$  and  $r_j^{(3)}(\mathcal{T}_c)$ .  
 Apply refresh-time on  $d = \{1, 2, 3\}$  to obtain the grid  $\mathcal{T}_d$ .  
 Compute  $f_j^{(2)}(\mathcal{T}_d) = r_j^{(2)}(\mathcal{T}_d) - \hat{h}_{21}f_j^{(1)}(\mathcal{T}_d)$   
 Estimate  $\hat{h}_{32}$  as the realized beta between  $f_j^{(2)}(\mathcal{T}_d)$  and  $r_j^{(3)}(\mathcal{T}_d)$ .  
 Estimate  $\hat{g}_{33}$  using an IV estimator on  $f_j^{(3)}(\mathcal{T}_d) = r_j^{(3)}(\mathcal{T}_d) - \hat{h}_{32}f_j^{(2)}(\mathcal{T}_d) - \hat{h}_{31}f_j^{(1)}(\mathcal{T}_d)$ .
5. *Step  $k=4$  to  $d$ :*  
 For  $l = 1, \dots, (k-1)$ :  
 Create the grid  $\tau^z$  by applying refresh-time on  $z = \{1, \dots, l, k\}$  to obtain the grid  $\mathcal{T}_z$ .  
 Extract the factors  $f_j^{(m)}(\mathcal{T}_z) = r_j^{(m)}(\mathcal{T}_z) - \sum_{n=1}^{m-1} h_{mn}f_j^{(n)}(\mathcal{T}_z)$ , for  $m = 1, \dots, l$  at the new grid.  
 Estimate  $\hat{h}_{kl}$  as the realized beta between  $f_j^{(l)}(\mathcal{T}_z)$  and  $r_j^{(k)}(\mathcal{T}_z)$ .  
 Estimate the  $g_{kk}$  as the IV of projections based on the estimates  $\hat{h}$ .

Finally, we also recommend to use a strip and replace method such that the diagonal elements of the CholCov are estimated using all available observations. More precisely, denote by  $\hat{D}$  the diagonal matrix with the ratio between the integrated variance estimates and the corresponding diagonal elements of  $\hat{H}\hat{G}\hat{H}$ . Then the strip and replace version of the CholCov is

$$CholCov^\bullet = \hat{D}^{1/2}\hat{H}\hat{G}\hat{H}\hat{D}^{1/2}. \quad (3.20)$$

In the simulations and application we will use the MRC pre-averaging estimator of Christensen, Kinnebrock, and Podolskij (2010) for the estimation of  $g_{kk}$  and  $h_{kl}$  and denote this by  $mrcCholCov^\bullet$  or  $CholCov$  when there is no confusion possible.

## 3.4 Asymptotic Results

In this section we derive the asymptotic distribution of the  $CholCov$  estimator in (3.16)-(3.17). Throughout this section, we assume that the refresh times  $\mathcal{T}_k$  satisfy  $\mathcal{T}_k \subset T_i$ , for  $i = 1, \dots, k$ , and  $\mathcal{T}_{k+1} \subset \mathcal{T}_k$ , such that the returns on the refresh time grid  $\mathcal{T}_{k+1}$  can be written as a linear combination of the returns on the grid  $\mathcal{T}_k$ . We further simplify the model by assuming that the underlying diffusion process has constant volatility, no drift and that the prices are observed without microstructure noise, i.e.

$$dY_t = \sigma_0 dW_t. \quad (3.21)$$

We show that even in this case, obtaining the asymptotic distribution of the parameters estimated sequentially on different time grids is non-trivial.

The main idea of our proof consists in showing that *every row* of the  $CholCov$  estimator is asymptotically equivalent to the row of the realized covariance of  $Y$  computed for the mesh  $\mathcal{T}_k$ . This property allows to provide necessary and sufficient condition for the asymptotic normality not of the whole  $CholCov$  estimator. This is done by applying the results in Nualart and Peccati (2005) and Peccati and Tudor (2005).

In order to derive the joint asymptotic distribution of  $CholCov$ , we proceed in four steps:

1. First, we show that in the synchronized case (equally spaced observations), the pair  $(\widehat{H}, \widehat{G})$  given by (3.16)-(3.17) is the Cholesky decomposition of the MLE of  $\Sigma_0 := \sigma_0 \sigma_0'$ . This implies that in the synchronized case,  $CholCov$  is just the MLE estimator of  $\Sigma_0$ . This result is used in Proposition 1, where consistency and asymptotic normality of the pair  $(\widehat{H}, \widehat{G})$  is established.
2. Second, for non-synchronous data with  $\mathcal{T}_{k+1} \subset \mathcal{T}_k$ , we show that the  $k$ th row of  $CholCov$  corresponds to the  $k$ th row of the MLE of  $\Sigma_0$  based on  $\{Y_t : t \in \mathcal{T}_k\}$ .
3. Third, let  $(A)_k$  denote the  $k$ th row of a matrix  $A$ , and  $RCov$  the realized covariance of the process  $Y$  in the mesh  $\mathcal{T}_k$ . We introduce Assumption 1 to guarantee that for  $k = 1, \dots, d$

$$(CholCov)_k = (RCov)_k + o_{\mathbb{P}}\left(\mathcal{N}_k^{1/2}\right).$$

Since the mapping  $(H, G) \mapsto HGH'$  is bijective and continuously differentiable, it follows that under Assumption 1,  $(RCov)_k$  is consistent and asymptotically normal. These results are presented in Proposition 2.

4. Finally, the joint asymptotic distribution of *CholCov* (based on the limit distribution of *RCov*) is derived in Theorem 1 using the results of Nualart and Peccati (2005) and Peccati and Tudor (2005).

### 3.4.1 Preliminaries

In this section,  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq 1}, \mathbb{P})$  will denote a filtered probability space satisfying the usual conditions of right-continuity and completeness and where the observed prices are generated by the Brownian martingale without drift and constant volatility in (3.21). Furthermore, we will assume that  $\mathcal{F} = ((B_t)_{0 \leq t \leq 1})$ , the  $\sigma$ -algebra generated by our underlying Brownian motion.

Let  $\mathbb{M}_d^+$  be the space of symmetric positive definite matrices with real entries. We will embed  $\mathbb{M}_d^+$  with the Frobenius norm such that if  $A \in \mathbb{M}_d^+$ , then  $\|A\| = \text{tr}(A^T A)^{1/2}$ . The Cholesky decomposition is unique on  $\mathbb{M}_d^+$ . This means that the mapping  $\Sigma(H, G) = HGH'$ , creates a bijection between the convex space

$$\mathbb{M}_{d, \text{chol}}^+ = \{(H, G) : G \text{ diagonal with } g_{kk} \geq 0, \text{ and } H \text{ lower diagonal with ones in its diagonal}\},$$

and  $\mathbb{M}_d^+$ . Let  $I_k$  be the  $k \times k$  identity matrix and  $\Pi_k$  the projection matrix, which, for a  $d$ -dimensional vector  $x = (x_1, \dots, x_d)'$  on  $x(k) := (x_1, \dots, x_k)'$ . Furthermore, if  $A$  is  $d \times d$  matrix and  $1 \leq k \leq d$ ,  $A(k)$  will denote the projection of  $A$  into its first submatrix of dimension  $k \times k$ , i.e.  $A(k) = (A_{i,j})_{i,j=1,\dots,k}$ . In particular, if  $A \in \mathbb{M}_d^+$ ,

$$A(k) = \Pi_k A \Pi_k'. \quad (3.22)$$

We will use a row version of the vech operator that stacks the rows of the lower-triangular part of a  $d$ -dimensional matrix  $A$  into a vector, as follows:

$$\text{rvech}(A) = (a_{11}, a_{21}, a_{22}, \dots, a_{k1}, a_{k2}, \dots, a_{kk-1}, a_{kk}, \dots, a_{dd-1}, a_{dd})'.$$

Note that

$$\text{rvech}(A(k)) = L(\Pi_k \otimes \Pi_k) M B^{-1} \text{rvech}(A), \quad (3.23)$$

with  $L$  and  $M$  such that  $\text{vech}(A) = L \text{vec}(A)$  and  $\text{vec}(A) = M \text{vech}(A)$  and  $B$  satisfying  $\text{rvech}(A) = B \text{vech}(A)$ . Obviously  $\text{rvech}$  is an injective continuously differentiable linear mapping from  $\mathbb{M}_d^+$  into a convex subset of  $\mathbb{R}^{d(d+1)/2}$ .

Throughout the following, we will use versions of  $[(H - I_d) + G]$ , where all elements on a single row are estimated on the same grid. The parameters on a lower row are estimated using refresh-times of a larger number of assets than those belonging to an upper row. The  $\text{rvech}$  of this matrix will therefore stack the elements in such a way that elements on the same grid are adjacent.

In order to derive our main results, we first recall the usual maximum likelihood

estimation results under synchronous observations and show that the CholCov estimated on non-synchronous series is consistent and marginally normal. Then, using the results in Nourdin, Peccati, and Réveillac (2010); Nualart and Peccati (2005); Peccati and Tudor (2005), we derive the joint distribution and show that the asymptotic covariance of the scaled CholCov parameter estimates is proportional to that of the Realized Covariance (Barndorff-Nielsen and Shephard, 2004a) with a factor that is a function of the overlap of the different grids. All proofs are provided in Appendix B.

### 3.4.2 The case of synchronous observations

Here we assume we dispose of a synchronous sample, i.e  $T := \mathcal{T}_1 = \mathcal{T}_2 = \dots = \mathcal{T}_d$ , of  $n$  equispaced returns generated by (3.21). The corresponding log-likelihood function is:

$$\bar{Q}_n(\Sigma_0) := -\frac{1}{2} \left\{ \ln(|\Sigma_0|) + \text{tr} [\Sigma_0^{-1} \bar{\Sigma}_n] \right\}, \quad n \in \mathbb{N}, \quad (3.24)$$

where  $\bar{\Sigma}_n$  is the realized covariance estimator using all available prices, i.e,

$$\bar{\Sigma}_n = \bar{\Sigma} := \sum_{j=1}^n (Y_{t_j^n} - Y_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n})'. \quad (3.25)$$

It is well known that  $\bar{\Sigma}_n$  maximizes  $\bar{Q}_n$  on  $\mathbb{M}_d^+$  and, under  $\mathbb{P}$ , it is strongly consistent for  $\Sigma_0$ , that is  $\bar{\Sigma}_n \xrightarrow{a.s.} \Sigma_0$ , as  $n \rightarrow \infty$ . On the other hand, since  $\bar{\Sigma}_n \in \mathbb{M}_d^+$  almost surely, there exists a unique pair of matrices,  $\bar{G}_n$  and  $\bar{H}_n$ , diagonal and lower diagonal with ones in its diagonal, respectively, such that

$$\bar{\Sigma}_n = \bar{H}_n \bar{G}_n \bar{H}_n', \quad \text{a.s.}$$

It follows from this that,  $\bar{g}_{11} = \bar{\Sigma}_{11}$ ,  $\bar{h}_{i1} = \frac{\bar{\Sigma}_{i1}}{\bar{g}_{11}}$  for  $i = 2, \dots, d$ , and for  $k \geq 2$  and  $k > l$

$$\bar{g}_{kk} = \bar{\Sigma}_{kk} - \sum_{m=1}^{k-1} \bar{h}_{km}^2 \bar{g}_{mm}, \quad (3.26)$$

$$\bar{h}_{kl} = \frac{1}{\bar{g}_{ll}} \left( \bar{\Sigma}_{kl} - \sum_{m=1}^{l-1} \bar{h}_{kl} \bar{h}_{ml} \bar{g}_{mm} \right). \quad (3.27)$$

**Proposition 1** *Suppose that  $T_1 = T_2 = \dots = T_d$  and  $\Delta_i(T) = 1/n$ . Let  $(\hat{H}, \hat{G}) \in \mathbb{M}_{d, \text{chol}}^+$  be the matrices associated to (3.16) and (3.17). Then  $(\hat{H}, \hat{G})$  is the Cholesky decomposition of  $\bar{\Sigma}$  (the MLE estimator of  $\Sigma_0$ ). In particular,  $(\hat{H}, \hat{G})$  is strongly consistent for  $(H_0, G_0)$  under  $\mathbb{P}$  and*

$$n^{1/2}(\hat{q} - q_0) \xrightarrow{d} N(0, \Phi), \quad \text{as } n \rightarrow \infty,$$

where  $\hat{q} := \text{rvech} \left[ \left( \hat{H} - I_d \right) + \hat{G} \right]$ ,  $q_0 := \text{rvech} \left[ \left( H_0 - I_d \right) + G_0 \right]$ ,  $\Phi \in \mathbb{M}_{d(d+1)/2}^+$  is the inverse of the Fisher's information associated to  $\bar{Q}_n$  as a function of  $q_0$ .

General formulas for the score function, the Hessian matrix, and the Fisher information for the parameters of  $q_0$  are given in Appendix C. In particular, we show that the estimation error of the MLE for the  $g_{kk}^0$  parameters is asymptotically uncorrelated to that of the MLE for the  $h_{kl}^0$  parameters, and that the estimation errors of  $h_{kl}^0$  and  $h_{mn}^0$  are only correlated if  $l = n$ . As such, the Fisher matrix and its inverse are block diagonal. An unsurprising but interesting result is that the asymptotic variance of the  $g_{kk}^0$  elements, which are computed as realized variances, is equal to  $2(g_{kk}^0)^2$ .

**Remark 1** Observe that in the previous proposition  $\left( \hat{H}, \hat{G} \right)$  is the Cholesky decomposition of the realized covariance associated to the mesh  $T$ . This property only holds when  $\Delta_i(T) = 1/n$ . For the general case (non-equally spaced mesh), if  $T = \{0 = t_0^n < t_1^n < \dots < t_n^n = 1\}$ , the MLE  $\left( \tilde{H}, \tilde{G} \right)$  is the Cholesky decomposition of

$$\tilde{\Sigma} = \sum_{j=1}^n \frac{1}{n \Delta_j(T)} (Y_{t_j^n} - Y_{t_{j-1}^n})(Y_{t_j^n} - Y_{t_{j-1}^n})', \quad (3.28)$$

where  $\Delta_j(T) = t_j^n - t_{j-1}^n$ , which of course does not coincide with the realized covariance. However, we can resolve this situation by imposing asymptotic behavior on the observation times. Indeed, let  $\hat{\Sigma}$  be the realized covariance associated to the mesh  $T$  and  $\tilde{\Sigma}$  the MLE estimator of (3.28). Trivially, we have

$$\left\| \hat{\Sigma} - \tilde{\Sigma} \right\| \leq \sup_j \left| \frac{1}{\Delta_j(T) n} - 1 \right| \left\| \hat{\Sigma} \right\|. \quad (3.29)$$

Therefore, if

$$\sup_j |\Delta_j(T) n - 1| = o(n^{-1/2}), \quad \text{as } n \rightarrow \infty, \quad (3.30)$$

we have that  $\hat{\Sigma} = \tilde{\Sigma} + o_{\mathbb{P}}(n^{1/2})$ . We conclude that if (3.30) holds, then the conclusions in Proposition 1 are true for non-equispaced data as well. Finally, let us remark that (3.30) implies that  $|T| := \sup_j (t_j^n - t_{j-1}^n) \sim 1/n$  as  $n \rightarrow \infty$ . This means that the convergence of  $\hat{\Sigma}$  to  $\Sigma_0$  is not only in probability but also almost surely. See De la Vega (1974) for more details.

### 3.4.3 The case of non-synchronous observations

In order to study the non-synchronized case, in analogy to Remark 1, we introduce an extra assumption:

**Assumption 1** For every  $k = 1, \dots, d-1$ ,  $T_{k+1} \subseteq T_k$  (or  $\mathcal{T}_{k+1} \subseteq \mathcal{T}_k$ ). Further, for any

### 3. CHOLCOV

$k, i = 1, \dots, d$ , as  $n \rightarrow \infty$ ,  $N_k \rightarrow \infty$ ,  $N_k/N_i = O(1)$ , and

$$\sup_j |\Delta_j(T) n - 1| = o(n^{-1/2}), \quad \text{as } n \rightarrow \infty, \quad (3.31)$$

**Remark 2** Observe that, besides the speed of convergence of  $\sup_j |\Delta_j(T) n - 1|$  to zero, Assumption 1 seems to be standard in the non-synchronous framework, e.g. Mykland (2010), Hayashi and Yoshida (2008) and Christensen, Podolskij, and Vetter (2013). Furthermore, as it was pointed out in Remark 1, our assumption on the speed of convergence on (3.31) is to control the asymptotic error between the CholCov estimator and the quadratic covariation of  $Y$ , which arises from the choice of a maximum likelihood based estimator.

In the presence of non-synchronous data, the estimation is done iteratively on expanding subsets. For a given refresh time grid  $\mathcal{T}_k$  on which we estimate the Cholesky parameters  $G_k^0$  and  $H_k^0$  of the projected covariance matrix  $\Sigma_0(k) = \Pi_k \Sigma_0 \Pi_k'$  using (3.13) and (3.14), the observations are synchronous and the following result follows directly from the previous results.

**Proposition 2** *Let Assumption 1 hold. Let  $(\hat{g}_{kk})_{k=1, \dots, d}$  and  $(\hat{h}_{kl})_{k=1, \dots, d, l < k}$  be as in (3.16) and (3.17). Define*

$$\hat{q}_k := A_k \text{rvech} \left[ \left( \hat{H} - I_d \right) + \hat{G} \right], \quad q_k^0 := A_k \text{rvech} \left[ \left( H_0 - I_d \right) + G_0 \right],$$

with  $A_k = \begin{bmatrix} 0 & I_k \end{bmatrix}$ , i.e. the  $k$ -th row of  $\left( \hat{H} - I_d \right) + \hat{G}$  and  $\left( H_0 - I_d \right) + G_0$ , respectively. Then, for every  $k = 1, \dots, d$ ,  $\hat{q}_k \xrightarrow{\mathbb{P}} q_k^0$  and

$$\sqrt{N_k} [\hat{q}_k - q_k^0] \xrightarrow{d} N(0, \bar{\Phi}_k), \quad \text{as } n \rightarrow \infty,$$

where  $\bar{\Phi}_k = A_k \Phi_k A_k'$ , with  $\Phi_k = \Pi_{k(k+1)/2} \Phi \Pi_{k(k+1)/2}'$  and  $\Phi$  is as in Proposition 1.

#### 3.4.4 A CLT for the CholCov

In this subsection, within the framework of Assumption 1, we derive the asymptotic distribution of our *CholCov* estimator. In Appendix A, we provide a summary of the theory used to proof our result.

Under the notation of Proposition 2, we have that  $q_k = A_k q$ . Furthermore, since the mapping  $(H, G) \mapsto \text{rvech}(HGH^T)$  is a continuously differentiable bijection between  $\mathbb{M}_{d, \text{chol}}^+$  and  $\mathbb{R}^{d(d+1)/2}$ , then, by the delta method,  $q$  is asymptotically normal if and only if the random vector  $\left( A_k \text{rvech} \left[ \tilde{\Sigma}(k) \right] \right)_{k=1, \dots, d}$  is asymptotically normal as well. Moreover, under Assumption 1,

$\text{rvech} \left[ \widehat{\Sigma}(k) \right] = \text{rvech} \left[ \widetilde{\Sigma}(k) \right] + o_{\mathbb{P}} \left( N_k^{-1/2} \right)$ , with  $\widehat{\Sigma}(k)$  the realized covariance of the process  $\pi_k(\bar{Y})$  for the mesh  $\mathcal{T}_k$ . Hence, the asymptotic properties of  $q$  are the ones of  $\left( A_k \text{rvech} \left[ \widehat{\Sigma}(k) \right] \right)_{k=1, \dots, d}$  which we derive in the following theorem.

**Theorem 1** *Let Assumption 1 hold. For every  $k = 1, \dots, d$ , let*

$$\beta_k = A_k \text{rvech} \left[ \sqrt{N_k} \left( \widehat{\Sigma}(k) - \Sigma_0(k) \right) \right]$$

*be the  $k$ -th row of  $\sqrt{N_k} \left( \widehat{\Sigma}(k) - \Sigma_0(k) \right)$ . Then as  $n \rightarrow \infty$ , the vector  $\beta = (\beta_k)_{k=1, \dots, d}$  converges in distribution to a normal random vector with covariance structure given by*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\beta_k(i) \beta_{k+l}(j)] := \Psi((k, i), (k+l, j)) \rho_{k, k+l}, \quad (3.32)$$

*where, for every  $k = 1, \dots, d$ ,  $l = 1, \dots, d-k$ ,  $i = 1, \dots, k$ , and  $j = 1, \dots, k+l$ ,  $\rho_{k, k} = 1$ ,*

$$\rho_{k, k+l} := \lim_{n \rightarrow \infty} \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb} \left( \Lambda_p^{k+l} \cap \Lambda_m^k \right)^2, \quad (3.33)$$

and

$$\Psi((k, i), (k+l, j)) := \Sigma_{k, k+l}^0 \Sigma_{i, j}^0 + \Sigma_{k, j}^0 \Sigma_{i, k+l}^0, \quad (3.34)$$

with

$$\Lambda_m^k = \left( t_{m-1}^k, t_m^k \right], \quad k = 1, \dots, d, m = 1, \dots, N_k, \quad (3.35)$$

and  $\text{Leb}((a, b]) = b - a$  for  $a < b$ , and zero otherwise.

**Remark 3** Note that the results in the previous theorem are in agreement with those derived in Hayashi and Yoshida (2008) and Christensen, Podolskij, and Vetter (2013). We would like to emphasize that the innovation in our result relies on the joint convergence of the CholCov estimator under different asymptotic rates.

**Remark 4** Within the framework of Assumption 1, the previous theorem can be generalized to the context of random sample times. Indeed, if we assume that the times on  $\mathcal{T}_k$  are random but independent of our driver Brownian motion, then  $\beta$  as in Theorem 1 is asymptotically mixed Gaussian, providing that the following limit exists and it is finite

$$\rho_{k, k+l} = \mathbb{P}\text{-} \lim_{n \rightarrow \infty} \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb} \left( \Lambda_p^{k+l} \cap \Lambda_m^k \right)^2.$$

The structure of the conditional covariance remains as in (3.32).

Applying the delta method, we obtain the following asymptotic normality result for the distribution of the CholCov estimator.

**Corollary 1** *Let  $q = (q_k)_{k=1,\dots,d}$ , with  $q_k$  as in Proposition 2. Under Assumption 1,  $q$  is asymptotically normal with covariance structure given by*

$$\lim_{n \rightarrow \infty} \mathbb{E} [\hat{q}\hat{q}'] = \nabla q_0' \Psi \nabla q_0,$$

where  $\Psi$  is as in Theorem 1 and  $\nabla q_0$  the gradient of  $q_0$  with respect to  $\text{rvech}(\Sigma_0)$ .

**Remark 5 A Feasible Estimator of the asymptotic covariance matrix.** Theorem 1 shows that the distribution of the CholCov using the RCOV as candidate estimator has the asymptotic distribution of the RCOV when data is synchronous. In case on non-synchronous price observations, the asymptotic covariance of the CholCov estimates correspond to the covariance of the RCOV estimates, multiplied by the scaling factor  $\rho_{k,k+l}$ .  $\rho_{k,k+l}$  can easily be computed using the observation times and for  $\Psi$ , the asymptotic covariance of the realized covariance estimator, we recommend to use the estimator of Barndorff-Nielsen and Shephard (2004a). We suggest simply synchronizing all series at once to estimate this quantity, and will show in the simulations that this feasible estimator works well in finite samples.

## 3.5 Monte Carlo simulation

The CholCov estimators are positive semidefinite by construction and, in the presence of noise, it uses robust sequential estimators of the underlying  $q$  parameter such as the pre-averaging MRC estimator. In this section, we first use a Monte Carlo analysis to compare the accuracy of the *mrcCholCov* using pre-averaging based estimates of  $q$  with the standard MRC estimator of Christensen, Kinnebrock, and Podolskij (2010) and its composite counterpart, denoted cMRC. The standard MRC is estimated on a fully synchronized sample with  $N$  observations. It is implemented with a window length of  $\lfloor N^{1/2+0.1} \rfloor$  at which it is guaranteed to be PSD, but for which the convergence is only at the rate  $N^{1/5}$ . For the cMRC estimator in which elements are estimated on pairwise grids, we use the  $N^{1/4}$  rate MRC estimator with window length  $\lfloor N^{1/2} \rfloor$ . We use the same efficient estimator for the *mrcCholCov*<sup>\*</sup>, which will still result in a PSD estimate. We report the average condition number of the different estimators for a range of sample sizes and show the CholCov systematically has a lower condition number than its counterparts. For the *mrcCholCov*, the asymptotic distribution is unknown. We conclude this section by studying the finite-sample accuracy of the normality approximation for the finite-sample distribution of the realized covariance based CholCov estimates (assuming prices are observed without microstructure noise), as described in Section 3.4.

### 3.5.1 Setup

As in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), we generate hypothetical prices, with  $Y^{(i)}(s)$  the associated log-price of asset  $i$ , from the log-price diffusion given by

$$\begin{aligned} dY_t^{(i)} &= \mu^{(i)} ds + dV_t^{(i)} + dF_t^{(i)} \\ dV_t^{(i)} &= \rho^{(i)} \sigma_t^{(i)} dB_t^{(i)} \\ dF_t^{(i)} &= \sqrt{1 - (\rho^{(i)})^2} \sigma_t^{(i)} dW_t, \end{aligned}$$

with  $k = 1, \dots, d$ . All  $B^{(i)}$  as well as  $W$  are independent Brownian motions.  $F^{(i)}$  denotes the common factor, scaled by  $\sqrt{1 - \rho^2}$  to determine its strength.

Each  $Y^{(i)}$  is a diffusive SV model with drift  $\mu^{(i)}$ . Their random spot volatility are given by  $\sigma^{(i)} = \exp(\beta_0^{(i)} + \beta_1^{(i)} \varrho^{(i)})$ , with  $d\varrho^{(i)} = \alpha^{(i)} \varrho^{(i)} dt + dB^{(i)}$ . The correlation between the changes in  $Y^{(i)}$  and  $Y^{(j)}$  is constant and equals  $\sqrt{1 - (\rho^{(i)})^2} \sqrt{1 - (\rho^{(j)})^2}$ .

We calibrate the parameters  $(\mu, \beta_0, \beta_1, \alpha, \rho)$  at  $(0.03, -5/16, 1/8, -1/40, -0.3)$  as in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011). The stationary distribution of  $\varrho$  is used to restart the process each day at  $\varrho(0) \sim \mathcal{N}(0, (-2(\beta)^2/\alpha)^{-1})$ . The parameter choice ensures that  $\mathbb{E} \left( \int_0^1 \sigma^{(i)2}(u) du \right) = 1$ . The fact that  $\rho$  is set equal for all  $i$  leads to an equicorrelation structure with common correlation coefficient 0.91.

Microstructure noise is added to the return log-prices as  $X^{(i)} = Y^{(i)} + \epsilon^{(i)}$  with

$$\epsilon^{(i)} \mid \sigma, X \stackrel{iid}{\sim} \mathcal{N}(0, \omega^2) \quad \text{with} \quad \omega^2 = \xi^2 \sqrt{N^{-1} \sum_{j=1}^N \sigma^{(i)4}(j/N)}.$$

Hence, the variance of the noise increases with the variance of the underlying process, in line with evidence from Bandi and Russell (2006).

Finally, independent Poisson processes are used to extract irregular, non-synchronous data from the complete high-frequency dataset. Each Poisson process is governed by a parameter  $\lambda^{(i)}$ , resulting in on average one observation every  $\lambda^{(i)}$  seconds for series  $i$ . On average, the series are observed  $23,400/\lambda^{(i)}$  times.

For each estimator, the bias and RMSE for variance and covariance elements are computed separately. The bias and RMSE, for a given element  $(k, l)$  of the matrix  $\hat{\Sigma}_{k,l}$ , are defined as

$$\text{Bias}_{k,l} = \hat{\Sigma}_{k,l} - \Sigma_{k,l} \quad \text{and} \quad \text{RMSE}_{k,l} = \sqrt{(\hat{\Sigma}_{k,l} - \Sigma_{k,l})^2}. \quad (3.36)$$

We report the averages of the above two statistics over 1.000 replications. We consider average bias and RMSE over multiple elements, making a distinction between liquid and illiquid assets.

### 3.5.2 Comparison of CholCov with the standard (composite) MRC estimators

The comparison is done for  $d = 20$  series with  $\lambda = \{5, 5, \dots, 5, 120\}$ .<sup>3</sup> The first nineteen series are observed on average every five seconds, and the last series once every two minutes. The MRC on the full dataset will generally use less than 23,400/120 observations, even though for the vast majority of the series more data is available. Our estimators will use the high observation frequency of the liquid series for all elements but those involving the last series.

We consider the MRC as well as its Composite counterpart, denoted cMRC. For the Composite estimator, as in Lunde, Shephard, and Sheppard (2015), we first estimate the variances,  $D$ , using the univariate version of the estimator, after which realized correlations,  $R$ , are estimated on pairs of data. Similar to the CholCov, these estimators not only have the advantage of better data-sampling, but also optimal bandwidth selection for each element. The resulting estimate of the covariance matrix,  $DRD$ , will not necessarily be PSD, so any possible negative eigenvalues are set to slightly over zero as in Barndorff-Nielsen and Shephard (2004b). Their performance is compared in terms of the three aforementioned criteria. For the bias and RMSE the results are split up by reporting the averages of those elements involving the illiquid stock, and those that do not separately.

Table 3.1 reports the results of the simulation. Panel A reports the fraction of PSD estimates for the various estimates. The CholCov and MRC estimators are both PSD by construction. The cMRC is not ensured to be PSD, and the simulations show that in the vast majority of cases it is not.

Panel B reports estimation results of the various individual elements. We show the average bias and RMSE of all covariance, correlation and variance elements, as well as their averages for those elements involving the illiquid asset and those that do not separately. First consider the MRC. As everything is estimated on a common grid, there is no real difference in the bias and RMSE for each of the elements between the liquid and illiquid stocks. Compared to the other estimators, even if the bias is sometimes similar, the RMSE is larger for all elements.

Both versions of the CholCov and cMRC are much more accurate. The only problem occurs in the variance estimate of the illiquid series, which is common to both estimators. The Strip-and-Replace estimation reduces RMSE, but has slightly larger bias compared to the mrcCholCov. The bias in variances transfers to a bias in the covariances involving the illiquid assets, but the correlation is estimated accurately. On average, the mrcCholCov<sup>•</sup> estimates the correlations with the least bias and the lowest RMSE, amongst the estimators considered.

Finally, we report more results on the quality of the estimates by reporting condition numbers for various dimensions. The condition number is defined as the ratio of

---

<sup>3</sup>Unreported simulations show similar, if less pronounced, conclusions with  $\lambda = 5$  for all assets and  $\lambda = \{2, 4, \dots, 38, 40\}$ .

**Table 3.1:** Results Simulation

Panel A	mrcCholCov		mrcCholCov <sup>•</sup>		MRC		cMRC		PSD cMRC	
<i>Fraction PSD</i>										
$\xi^2 = 0.000$	1.000		1.000		1.000		0.377		1.000	
$\xi^2 = 0.001$	1.000		1.000		1.000		0.199		1.000	
$\xi^2 = 0.010$	1.000		1.000		1.000		0.000		1.000	
Panel B	mrcCholCov		mrcCholCov <sup>•</sup>		MRC		cMRC		PSD cMRC	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
<b>Covariances</b>										
<i>All elements</i>										
$\xi^2 = 0.000$	-0.005	0.093	-0.003	0.091	0.025	0.230	-0.004	0.097	-0.004	0.097
$\xi^2 = 0.001$	-0.006	0.094	-0.005	0.093	0.030	0.233	-0.004	0.098	-0.004	0.098
$\xi^2 = 0.010$	-0.005	0.100	-0.007	0.096	0.035	0.246	-0.006	0.105	-0.006	0.105
<i>Liquid elements</i>										
$\xi^2 = 0.000$	-0.002	0.086	-0.001	0.084	0.026	0.230	-0.001	0.090	-0.001	0.090
$\xi^2 = 0.001$	-0.004	0.088	-0.001	0.087	0.029	0.233	-0.001	0.091	-0.001	0.091
$\xi^2 = 0.010$	-0.002	0.093	-0.003	0.090	0.035	0.247	-0.003	0.098	-0.003	0.098
<i>Illiquid elements</i>										
$\xi^2 = 0.000$	-0.035	0.143	-0.034	0.142	0.025	0.226	-0.035	0.143	-0.035	0.143
$\xi^2 = 0.001$	-0.028	0.139	-0.036	0.135	0.029	0.230	-0.034	0.145	-0.034	0.144
$\xi^2 = 0.010$	-0.036	0.146	-0.041	0.137	0.034	0.241	-0.034	0.157	-0.034	0.154
<b>Correlations</b>										
<i>All elements</i>										
$\xi^2 = 0.000$	-0.002	0.019	-0.002	0.019	-0.004	0.037	0.000	0.020	-0.002	0.019
$\xi^2 = 0.001$	-0.002	0.019	-0.001	0.018	-0.004	0.039	-0.001	0.021	-0.002	0.020
$\xi^2 = 0.010$	0.001	0.026	-0.001	0.023	-0.008	0.060	-0.003	0.030	-0.010	0.030
<i>Liquid elements</i>										
$\xi^2 = 0.000$	-0.001	0.017	-0.000	0.017	-0.004	0.037	0.000	0.017	-0.001	0.017
$\xi^2 = 0.001$	-0.002	0.017	-0.001	0.016	-0.004	0.039	-0.001	0.018	-0.001	0.018
$\xi^2 = 0.010$	0.001	0.025	-0.001	0.022	-0.008	0.060	-0.002	0.026	-0.007	0.026
<i>Illiquid elements</i>										
$\xi^2 = 0.000$	-0.012	0.032	-0.017	0.032	-0.004	0.035	-0.004	0.035	-0.006	0.034
$\xi^2 = 0.001$	-0.003	0.029	-0.002	0.028	-0.004	0.036	-0.004	0.036	-0.008	0.035
$\xi^2 = 0.010$	-0.001	0.035	0.001	0.031	-0.007	0.056	-0.007	0.056	-0.032	0.053
<b>Variances</b>										
<i>All elements</i>										
$\xi^2 = 0.000$	-0.004	0.111	-0.004	0.104	0.029	0.248	-0.004	0.104	-0.004	0.104
$\xi^2 = 0.001$	-0.004	0.114	-0.004	0.106	0.033	0.252	-0.004	0.106	-0.003	0.106
$\xi^2 = 0.010$	-0.004	0.129	-0.004	0.117	0.046	0.274	-0.004	0.117	-0.004	0.116
<i>Liquid elements</i>										
$\xi^2 = 0.000$	-0.002	0.103	-0.001	0.097	0.029	0.249	-0.001	0.097	-0.001	0.097
$\xi^2 = 0.001$	-0.002	0.106	-0.001	0.098	0.033	0.253	-0.001	0.098	-0.001	0.098
$\xi^2 = 0.010$	-0.003	0.121	-0.002	0.109	0.046	0.275	-0.002	0.109	0.004	0.109
<i>Illiquid elements</i>										
$\xi^2 = 0.000$	-0.048	0.209	-0.058	0.199	0.026	0.227	-0.058	0.199	-0.055	0.198
$\xi^2 = 0.001$	-0.035	0.211	-0.055	0.204	0.033	0.232	-0.055	0.204	-0.049	0.202
$\xi^2 = 0.010$	-0.031	0.231	-0.048	0.216	0.042	0.245	-0.048	0.216	-0.009	0.207

*Note:* Simulation results of the multivariate factor diffusion with  $d = 20$  with  $\lambda = 5, 5, \dots, 5, 120$ . Panel A reports the fraction of PSD estimates. Panel B reports the average bias and RMSE of the covariance, correlation and variances, displayed separately for those elements involving the illiquid asset and those that do not.

**Table 3.2:** Results Simulation II: Condition numbers

d	ICOV	mrcCholCov	mrcCholCov <sup>•</sup>	MRC	cMRC
2	21.22	23.78	24.03	22.08	22.53
5	51.55	93.01	91.58	134.65	99.30
10	102.11	321.16	305.51	822.71	400.43
20	203.22	1892.58	1714.24	13878.14	2259.41
50	506.55	9978.12	9663.10	46336.40	10831.52

*Note:* Simulation results of the multivariate factor diffusion with  $d = 2, 5, 10, 20, 50$  with  $\lambda = 5, 5, \dots, 5, 120$  and  $\xi^2 = 0.001$ .

the maximum over the minimum eigenvalue. A problem is said to be ill-conditioned if this ratio is large. In practise, this is mainly due to extremely low minimum eigenvalues. A well-conditioned matrix ensures that operations like inverting the matrix do not amplify estimation error, which is critical in many applications. As before, we investigate this within our simulations, where prices are simulated using poisson processes with parameters  $\lambda = \{5, 5, \dots, 5, 120\}$ , and we set the noise variance to  $\xi^2 = 0.001$ .

Table 3.2 reports the average condition number over 1000 simulations of the true matrix ICOV, and the estimated matrices. It becomes clear that the CholCov leads to slightly better conditioned matrices than the cMRC and vastly better conditioned matrices than the MRC. This is especially true for larger dimensional problems.

The simulation demonstrates that in a large dimensional setting, the sequential estimation of the integrated correlation utilizing the Cholesky decomposition offers vast improvements over full-dimensional estimation, and slight improvements in the Frobenius Distance over pairwise estimation. However, the CholCov is PSD by construction, whereas the Composite estimator needs a regularization to become PSD.

### 3.5.3 Normality of the CholCov estimator in finite samples with stochastic volatility.

The final simulations aim to validate the asymptotic theory for the realized covariance based CholCov estimator using the expanding refresh time grids  $\mathcal{T}_k$ . We assess the robustness of the results to deviations from the assumptions on the underlying process. To do so, we simulate a five-dimensional process according to the DGP described in Subsection 3.5.1. We consider different assumptions on the DGP. First, we distinguish between constant and time-varying volatility. Second, we consider different types of asynchronicity. Type I has no asynchronicity, and all observations are equispaced. Type II has asynchronicity, where the first series is observed every second, 5% of data is randomly discarded for the second series and an additional 5% of data is randomly discarded for the each additional series. The series are asynchronous, but  $T_{k+1} \subset T_k$ . Type III considers poisson sampling as in the previous section, with  $\lambda = 5$  for all series. We present the rejection frequencies of the hypothesis of multivariate normality CholCov estimates using

**Table 3.3:** Rejection rates of the normality assumption for the realized covariances based CholCov estimates

Asynchronicity Type	Constant Volatility	Noise	Infeasible		Feasible	
			80%	95%	80%	95%
I	Yes	No	19.8	4.9	19.1	4.3
II	Yes	No	18.7	5.2	17.0	3.9
I	No	No	20.8	4.1	18.5	5.8
II	No	No	19.5	4.2	18.9	5.7

*Note:* Type I has no asynchronicity, and all observations are equispaced. Type II has asynchronicity, where for the first series each second is observed, and for each next series, an additional 5% of observed prices are removed. The series are asynchronous, but  $T_{k+1} \subset T_k$ . Type III considers poisson sampling as in the previous section, with  $\lambda = 5$  for all series.

the Mahalanobis Distance (MD)

$$(\text{vech } CholCov - \text{vech } ICOV)' \varphi^{-1} (\text{vech } CholCov - \text{vech } ICOV) \sim \chi_{d(d+1)/2}^2 \quad (3.37)$$

where  $\varphi$  is as in (3.32). We use infeasible and feasible estimates of  $\varphi$ , and consider 80% and 95% significance tests. For the infeasible version of  $\varphi$  we simply compute the population quantity of  $\varphi$  using the simulated path, while for the feasible version we use the estimator given in Remark 2.

The results reported in Table 3.3 confirm the theory of the previous section, and show that relaxation of the constant volatility assumption does not greatly affect the CholCov's distribution. The rejection frequencies confirm that when there is no asynchronicity, i.e. Type I, the RCOV asymptotic distribution is valid. Second, if there is asynchronicity of Type II, the elements converge at different rates based on their number of observations.

## 3.6 Empirical Illustration I: Value-at-Risk forecasting

We expect our estimator to be especially useful in realistic large-scale portfolio applications that often require the estimator to be positive semidefinite and invertible. In our application we consider the forecasting of portfolio Value-at-Risk (VaR). When computing a portfolio VaR one has the option to either model the portfolio univariately or multivariately. For the univariate approach one uses the weights to compute portfolio returns and estimate its VaR based on the single series. Alternatively, one could estimate and model the full covariance matrix, and determine the portfolio VaR based on the multivariate setting. This has advantages for several reasons. First one can immediately calculate risk

estimates for many different portfolios. Additionally, it has the advantage that it can be used for dynamic portfolio allocation, such as for instance a minimum variance portfolio. Finally, and most importantly, the dynamics of each of the volatility and correlation components are modeled separately. Santos, Nogales, and Ruiz (2013) argue that for large dimensions, the information due to the multivariate modeling outweighs the additional uncertainty of estimating many parameters, and leads to better forecasts.

We contribute to the multivariate approach to portfolio VaR analysis by studying the efficiency gains of using intraday data for VaR forecast accuracy. We compare forecasts from models estimated on daily returns with estimates based on intradaily techniques, i.e. dynamic models applied to the *mrcCholCov*<sup>•</sup> estimates. We also consider the RCOV estimated on 5 minute previous-tick returns as a base case for using high-frequency estimators. Giot and Laurent (2004) and Brownlees and Gallo (2010) make the comparison between daily and high-frequency measures in a univariate setting. We are unaware of any paper comparing daily and intradaily models in the multivariate Value-at-Risk setting. For simplicity, in this application, we consider just two types of portfolios, equally- and value-weighted.

#### 3.6.1 Data

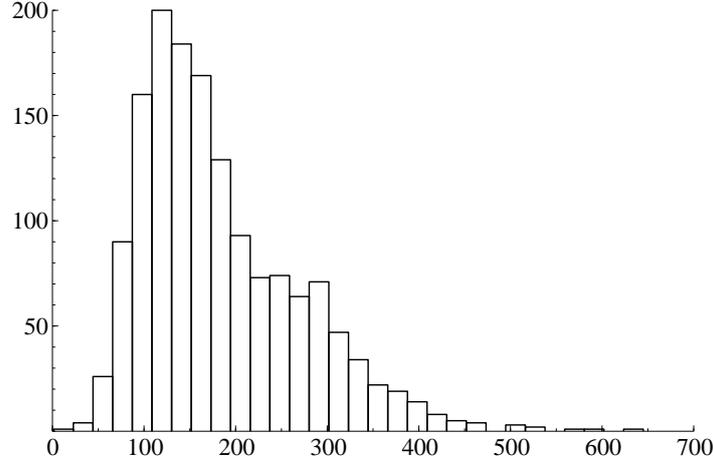
We analyze the portfolio risk for a total of 52 of the largest U.S. financial institutions.<sup>4</sup> We obtain trade data from January 2007 till December 2012 for a total of 1499 observation days. We clean the data using the step-by-step cleaning procedure of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). Throughout the analysis we use open-to-close returns obtained from the TAQ data. We use open-to-close as our ICov estimators do not take into account the overnight return. If we were to include overnight returns, gains in estimation accuracy would be more difficult to distinguish, as the overnight return is relatively dominant. The weights for the value-weighted portfolio are proportional to firms' market capitalization, determined by the shares outstanding (from CRSP) times the closing price of the stock.

The estimation problem is moderately large with 52 firms and synchronization of the data will greatly reduce the total number of observations. In Figure 3.3 we plot the frequencies of number of observations after refresh-time synchronization of all the series. The least amount of observations is 6 while the most is 624. The median is 161. There are 11 days where the number of observations is smaller than the dimension of the problem. This invalidates the use of the traditional full-dimension estimators like the MRC and MKernel, which are no longer invertible. Second, our application, like many applications involving covariance estimates, requires the estimate to be positive

---

<sup>4</sup>The Tickers are: ACAS, AET, AFL, AIG, AIZ, ALL, AMP, AXP, BAC, BBT, BEN, BK, BLK, BRKB, CB, CBG, CINF, CMA, COF, CVH, EV, FITB, FNF, GNW, GS, HBAN, HIG, HNT, ICE, JNS, KEY, MET, MTB, NTRS, NYX, PFG, PGR, PNC, PRU, RF, SEIC, SNV, STI, STT, TMK, TROW, UNH, UNM, USB, WFC, WU, ZION.

**Figure 3.3:** Daily Number of Observations



*Note:* This figure plots the daily number of observations after refresh-time synchronization of the 52 data series.

semidefinite, invalidating the composite estimation technique. As such, the only estimator that is guaranteed to be positive semidefinite - without an arbitrary correction - and uses efficient sampling is ours.

### 3.6.2 Methodology

Our aim is to forecast portfolio Value-at-Risk. For a given  $d$ -dimensional vector of weights  $w_t$ , we compute the portfolio  $VaR$  over the open-to-close horizon on day  $t$  as follows:

$$VaR_t^c = w_t' \mu_{t|t-1} + z_c \sqrt{w_t' H_{t|t-1} w_t}, \quad (3.38)$$

where  $\mu_{t|t-1}$  is the vector conditional means and  $H_{t|t-1}$  the conditional covariance matrix. Under the assumption of conditional normality,  $z_c$  is the  $c$  quantile of the standard normal distribution. The conditional normality assumption is hard to justify for single stocks, but the dynamic quantile test that we will perform does not reject it in our portfolio setting.

We consider the Value-at-Risk of both long and short positions, setting  $c = \{0.01, 0.025, 0.05, 0.95, 0.975, 0.99\}$ . For long positions, the risk comes from a drop in the price and therefore an observation in the left tail of the distribution, whereas for short positions the right tail is important. The Value-at-Risk is for long positions when  $c < 0.5$  and for short positions when  $c > 0.5$ .

The conditional mean is forecasted using AR(p) models, where the optimal lag order is individually determined by means of the Schwarz Information Criterion.

The CholCov and RCOV are ex-post measures, while for estimating the  $VaR$  a covari-

ance forecast and therefore a dynamic model is needed. We consider two types of models. In the first we impose the same dynamics on all elements, leading to a Scalar-BEKK type specification, like the HEAVY model of Noureldin, Shephard, and Sheppard (2012). In the second model that we consider, we allow for separate dynamics for the individual volatilities and correlations, like in the cRDCC of Bauwens, Storti, and Violante (2012). The performance of these two modeling strategies is compared to their counterparts using only daily returns, i.e. the Scalar-BEKK (Engle and Kroner, 1995) and cDCC (Aielli, 2013), respectively.

Before presenting the results, let us first detail the different estimation methods. The Scalar-BEKK / HEAVY models with dependence parameters  $\alpha$  and  $\beta$  take the form of

$$H_{t|t-1} = (1 - \alpha - \beta)\Omega + \alpha V_{t-1} + \beta H_{t-1|t-2}, \quad (3.39)$$

where for the Scalar-BEKK  $V_t = \varepsilon_t \varepsilon_t'$ , with  $\varepsilon_t$  corresponding to the vector of demeaned returns, and for the HEAVY model  $V_t = \hat{\Sigma}_t$ , the covariance estimate based on intraday data. To reduce the number of parameters to be estimated, we apply covariance targeting, where  $\Omega$  is the unconditional variance-covariance matrix of daily returns for the Scalar-BEKK and the average CholCov/RCOV for the HEAVY model. The HEAVY model additionally has a correction term, to match the unconditional variance of the model to that of daily returns. We implement the version of Equation (11) in Noureldin, Shephard, and Sheppard (2012).

The cDCC models take the following form:

$$\begin{aligned} H_{t|t-1} &= D_{t|t-1} R_{t|t-1} D_{t|t-1} \\ R_{t|t-1} &= \text{diag}(Q_{t|t-1})^{-1/2} Q_{t|t-1} \text{diag}(Q_{t|t-1})^{-1/2} \\ Q_{t|t-1} &= (1 - \alpha - \beta)\bar{Q} + \alpha P_{t|t-1}^* + \beta Q_{t-1|t-2}, \end{aligned} \quad (3.40)$$

where  $P_t^* = \text{diag}(Q_{t|t-1})^{1/2} D_{t|t-1}^{-1} V_t D_{t|t-1}^{-1} \text{diag}(Q_{t|t-1})^{1/2}$ ,  $V_t$  is defined as above, i.e.  $V_t = \varepsilon_t \varepsilon_t'$  for the cDCC and  $V_t = \hat{\Sigma}_t$  for the cRDCC. Note that to reduce the number of parameters to be estimated, we do correlation targeting, by replacing  $\bar{Q}$  by the mean of  $P_t^*$ . Both cDCC models can be estimated in two steps, where first univariate models are fitted to estimate the volatilities  $D_t$ , which are then used to estimate the conditional correlation  $R_t$ . For the univariate models we use ARFIMA(1,  $d$ , 0) models on the natural logarithm of the estimated variances, as there is ample evidence (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003) for the presence of long memory in realized variances. By modeling the natural logarithm of the variances, we ensure positive out-of-sample forecasts.<sup>5</sup> For a fair comparison we model the volatilities in the cDCC on daily returns using a long memory model as well, i.e. a FIGARCH(1,  $d$ , 1).

The Scalar-BEKK and cDCC are estimated using a Composite Gaussian Likelihood, while the HEAVY and cRDCC models are estimated using a Composite Wishart Like-

---

<sup>5</sup>We adjust for the bias caused by the log-transformation as in Giot and Laurent (2004).

likelihood, all on contiguous pairs. Composite likelihood techniques for large-dimensional ARCH-type models were developed in Engle, Shephard, and Sheppard (2008). They facilitate estimation, and reduce bias in parameter estimates present in large-dimensional problems. We assume Gaussian innovations for the computation of the  $VaR$  for all models.

To obtain the forecasts we estimate all the models on an increasing window of observations, making one-step-ahead forecasts for the 1.000 last days, re-estimating the parameters daily.

We test the out-of-sample performance of the  $VaR$  estimates in two ways. The first method is the dynamic quantile test of Engle and Manganelli (2004). They define a  $Hit$  variable associated with the ex-post observation of a  $VaR$  violation at time  $t$ :

$$Hit_t(c) = \begin{cases} 1 - c & \text{if } w'r_t < VaR_{t|t-1}^c \\ -c & \text{otherwise .} \end{cases} \quad (3.41)$$

Similarly,  $Hit_t(1 - c) = 1 - c$  if  $w'r_t > VaR_{t|t-1}^{1-c}$ . We run the regressions  $Hit_t(c) = \delta_0 + \sum_{k=1}^K \delta_k Hit_{t-k}(c) + \varepsilon_t$  and test the joint hypothesis  $H_0 : \delta_0 = \delta_1 = \dots = \delta_K = 0$ ,  $\forall k = 1, \dots, K$ .  $VaR$  violations are uncorrelated over time if the  $\delta_i$  with  $i > 0$  are 0, whereas the unconditional coverage is correct if  $\delta_0 = 0$ . Denote by  $\theta = (\delta_0, \delta_1, \dots, \delta_K)'$  the vector of parameters of the model and by  $X$  the matrix of explanatory variables of the regression. The test statistic is

$$\frac{\hat{\theta}' X' X \hat{\theta}}{c(1 - c)}, \quad (3.42)$$

and follows a  $\chi^2$ -distribution with  $K + 1$  degrees of freedom under the null of correct specification.

Second, as in Chen and Gerlach (2013), we measure performance by means of a loss function and the Model Confidence Set of Hansen, Lunde, and Nason (2011). They suggest the following loss function applicable to quantile forecasts, which is the criterion function minimized in quantile regressions:

$$LF_t^c = (VaR_{t|t-1}^c - w'r_t) Hit_t(c). \quad (3.43)$$

The Model Confidence Set (MCS) allows statistical comparison of the competing models by comparing their respective loss series. It uses a block-bootstrap to obtain the distribution of the mean loss and eliminates models based on the maximum of pairwise t-statistics, automatically controlling the familywise error. We set the level to 90%, such that models are removed from the set until the final model set includes the best models with 90% confidence.<sup>6</sup>

---

<sup>6</sup>The MCS is carried out using the MulCom 3.0 package for Ox (Hansen and Lunde, 2014b), using 10,000 resamples and a bootstrap block-length of 2.

#### 3.6.3 Results

Table 3.4 reports the results of the Dynamic Quantile Test. The first three columns depict the results for the long-positions, and the last three columns show the results for the short positions. The top panel reports the p-values for the Equal Weighted (EW) portfolio and the bottom panel shows the results for the Value Weighted (VW) portfolio. Each panel contains successively the results for the two models using daily returns, Scalar-BEKK and cDCC, and the two models utilizing the CholCov, HEAVY and cRDCC. We give the results for  $K = 1$  and  $K = 2$ , but they are qualitatively similar for larger  $K$ . The results are available upon request.

First consider the models on daily returns, the Scalar-BEKK and cDCC. The empirical results given in both panels tell the same story. The models using just daily returns are not flexible enough to accurately forecast the VaR. The Scalar-BEKK's unconditional coverage is rejected in many cases, and the rejection of the dynamic quantile test for all  $c$  and  $K$  shows that the violations are also dependent. The cDCC does perform a lot better, taking into account possible long-memory in volatility, but fails to model the left tail of the distribution adequately, with the p-values for the null hypothesis of the different versions of the test often smaller than 0.05, when  $c = 1\%$  or  $c = 2.5\%$ . We have considered alternative specifications, not reported for brevity, which included leverage effects, but these were also rejected. It is unclear whether the rejection is due to model misspecification or non-normal returns. As such we also estimated the model using the more flexible multivariate Student distribution. Again, this did not lead to significant improvements.

By increasing the information set to include intraday data, we can estimate the models on RCOV and CholCov. The HEAVY model applied to RCOV offers some slight improvements over the daily Scalar-BEKK. However, the model is still almost uniformly rejected. Applying the HEAVY model to CholCov offers more substantial improvements. However, the short memory HEAVY model is still rejected by the data. For the CholCov, conditional gaussianity is rejected for lag-length  $K = 2$  but not for  $K = 1$ . This suggests more lags are needed, and as such we have considered a HEAVY(2,2) model, which is also rejected by the data.

The cRDCC with ARFIMA dynamics on the variances takes into account the long-memory properties of Realized Variance. The cRDCC model utilizing RCOV has similar performance to the cDCC model using daily returns, with rejections in the left tail. However, the flexibility of the model, combined with the accurate CholCov estimates based on liquidity sorted assets, allows it to capture both the unconditional coverage and pass the test for independence of violations, with only a single rejection at the 5 percent level for the 1% VaR for  $K = 2$ .

To see the economic significance of using liquidity sorting instead of observation count sorting, Table 3.4 also reports the results using the CholCov based on observation count sorting along with the empirically best model, the ARFIMA cRDCC. The resulting VaR

### 3.6 Empirical Illustration I: Value-at-Risk forecasting

**Table 3.4:** Dynamic Quantile Test P-Values

	K\c	1.0%	2.5%	5.0%	95.0%	97.5%	99.0%
Equal Weighted							
Scalar-BEKK	1	0.018	0.000	0.011	0.001	0.000	0.000
	2	0.000	0.000	0.000	0.000	0.000	0.000
cDCC	1	0.029	0.015	0.335	0.499	0.754	0.950
	2	0.022	0.034	0.227	0.566	0.779	0.976
HEAVY-RCOV	1	0.000	0.001	0.102	0.198	0.016	0.000
	2	0.000	0.003	0.152	0.000	0.000	0.000
cRDCC-RCOV	1	0.028	0.270	0.764	0.902	0.591	0.042
	2	0.022	0.219	0.548	0.012	0.002	0.078
HEAVY-CholCov	1	0.385	0.204	0.557	0.262	0.800	0.042
	2	0.000	0.007	0.015	0.000	0.000	0.011
cRDCC-CholCov	1	0.010	0.008	0.272	0.777	0.899	0.746
<i>Observation Count Sorting</i>	2	0.002	0.014	0.210	0.213	0.944	0.856
cRDCC-CholCov	1	0.746	0.719	0.531	0.688	0.413	0.797
<i>Liquidity Sorting</i>	2	0.078	0.831	0.240	0.414	0.547	0.918
Value Weighted							
Scalar-BEKK	1	0.000	0.000	0.002	0.011	0.006	0.000
	2	0.000	0.000	0.000	0.000	0.000	0.000
cDCC	1	0.130	0.030	0.133	0.506	0.667	0.236
	2	0.055	0.057	0.246	0.712	0.735	0.354
HEAVY-RCOV	1	0.887	0.014	0.002	0.000	0.000	0.000
	2	0.060	0.028	0.004	0.000	0.000	0.000
cRDCC-RCOV	1	0.065	0.041	0.914	0.674	0.474	0.036
	2	0.000	0.076	0.954	0.174	0.003	0.068
HEAVY-CholCov	1	0.565	0.021	0.283	0.380	0.838	0.385
	2	0.000	0.003	0.189	0.000	0.000	0.085
cRDCC-CholCov	1	0.385	0.193	0.135	0.788	0.719	0.887
<i>Observation Count Sorting</i>	2	0.000	0.214	0.253	0.253	0.831	0.944
cRDCC-CholCov	1	0.950	0.675	0.900	0.946	0.667	0.625
<i>Liquidity Sorting</i>	2	0.039	0.315	0.920	0.387	0.735	0.809

*Note:* P-Values of the Dynamic Quantile Test for the Equal Weighted and Value Weighted portfolio Value-at-Risks. The Scalar-BEKK and cDCC are estimated on daily returns, while the last two models are estimated using the proposed  $\text{mrcCholCov}^\bullet$  covariance estimate.

### 3. CHOLCOV

**Table 3.5:** Loss function MCS results

c	1.0%	2.5%	5.0%	95.0%	97.5%	99%
Equal Weighted						
Scalar-BEKK						
cDCC				✓	✓	✓
HEAVY-RCOV						
cRDCC-RCOV		✓	✓	✓		
HEAVY-CholCov					✓	✓
cRDCC-CholCov <i>Observation Count Sorting</i>				✓	✓	✓
cRDCC-CholCov <i>Liquidity Sorting</i>	✓	✓	✓	✓	✓	✓
Value Weighted						
Scalar-BEKK						
cDCC				✓	✓	
HEAVY-RCOV						
cRDCC-RCOV			✓	✓		
HEAVY-CholCov			✓			
cRDCC-CholCov <i>Observation Count Sorting</i>					✓	✓
cRDCC-CholCov <i>Liquidity Sorting</i>	✓	✓	✓	✓	✓	✓

*Note:* The table shows the models included in the 90% Model Confidence set for the different VaR forecast quantiles and the two portfolios, based on the loss function (3.43).

forecasts are inferior. For the Value Weighted portfolio, rejections are similar to the liquidity sorting based forecasts, but for the Equal Weighted portfolio there are again problems in the left tail, with rejections for the two most extreme quantiles.

Second we consider the evaluation of the loss function on VaR forecasts. Table 3.5 shows which models are part of the 90% Model Confidence Set for each VaR quantile and for both portfolios. The cRDCC model on CholCov is always in the MCS for both portfolios. Nearer the centre of the distribution, the cDCC on daily returns and cRDCC on RCOV are also included in the MCS. For the right tail of the return distribution, the CholCov using observation count sorting are part of the MCS. In general, the Scalar-BEKK/HEAVY-type models are rejected here as well.

The results suggest that, for diversified portfolios like the ones considered here, accurate VaR forecasts can be obtained under the assumption of conditional normality, by utilizing the Cholcov estimator to make efficient use of the information content of high-frequency data and an appropriate dynamic model.

### 3.7 Empirical Illustration II: Forecasting betas

A further application in which we expect the CholCov to be of practical relevance is the forecasting of dynamic exposures to observed risk factors, which we will refer to as betas. As in the previous section, this requires an accurate and positive semidefinite estimator of the covariance matrix. We estimate the exposures of the 52 financial institutions in our

previous exercise to ten sector Exchange Traded Funds (ETFs), being SPY, XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, XLY. Our sample for the ETFs only runs until April 30th 2012, for a total of 1336 observations.

We compare the betas produced by one-step-ahead forecasts of the cDCC-type models for CholCov, RCOV and daily returns. As our sample is slightly reduced, we now consider the last 800 observations of our out-of-sample window. The betas are obtained by forecasting the 11-dimensional covariance matrix, of the 10 ETFs and the individual financial institution. We then obtain the 10-dimensional beta vector as

$$\hat{\beta}_{k,t|t-1} = \hat{H}_{kE,t|t-1} \hat{H}_{EE,t|t-1}^{-1},$$

where the covariance forecast  $H_{t|t-1}$  is partitioned into the firm  $k$  and the ten ETFs  $E$ :

$$H_{t|t-1} = \begin{bmatrix} H_{kk,t|t-1} & H_{Ek,t|t-1} \\ H_{kE,t|t-1} & H_{EE,t|t-1} \end{bmatrix}.$$

We use a regression-based comparison that was proposed by Engle (2015). To do so, define the variables

$$Z_{k,t|t-1} = \hat{\beta}'_{k,t|t-1} ETF_t,$$

where  $ETF_t$  is the vector of realized sector returns. We consider three different estimators of  $\hat{\beta}_k$ , based on the CholCov, RCOV and daily return, leading to three different auxiliary variables, denoted  $Z_{k,t|t-1}^{CholCov}$ ,  $Z_{k,t|t-1}^{RCOV}$ , and  $Z_{k,t|t-1}^{DCC}$ . We then estimate the following model by OLS:

$$r_{k,t} = \alpha + \delta_k^{CholCov} Z_{k,t|t-1}^{CholCov} + \delta_k^{RCOV} Z_{k,t|t-1}^{RCOV} + \delta_k^{DCC} Z_{k,t|t-1}^{DCC} + \eta_{k,t}, \quad (3.44)$$

for each  $k = 1, \dots, 52$ . A perfect specification for the  $\beta$ -vector would lead to a  $\delta$  coefficient equal to one, while the other coefficients would be zero. Engle (2015) does an in-sample comparison of a static beta with a DCC beta, and Hansen, Lunde, and Voev (2014) add the Realized Beta GARCH to the analysis and perform an out-of-sample analysis.

It follows that we can test which model is ‘superior’ by testing the following three null hypotheses:

$$\begin{aligned} H_{CholCov} &: \delta_k^{CholCov} = 1, & \delta_k^{RCOV} = \delta_k^{DCC} = 0 \\ H_{RCOV} &: \delta_k^{RCOV} = 1, & \delta_k^{CholCov} = \delta_k^{DCC} = 0 \\ H_{DCC} &: \delta_k^{DCC} = 1, & \delta_k^{CholCov} = \delta_k^{RCOV} = 0, \end{aligned}$$

taking into account heteroskedasticity by using White (1980) standard errors. Additionally, we wish to identify the benefits of using the CholCov over the RCOV. To test this, we consider the regression (3.44) but without  $Z_{k,t|t-1}^{DCC}$ , and test the hypotheses:

$$\begin{aligned} H'_{CholCov} &: \delta_k^{CholCov} = 1, & \delta_k^{RCOV} = 0 \\ H'_{RCOV} &: \delta_k^{RCOV} = 1, & \delta_k^{CholCov} = 0. \end{aligned}$$

### 3. CHOLCOV

**Table 3.6:** Beta Comparisons

<i>Panel A:</i> Summary of parameter estimates									
	Full Model			Excluding DCC		Single series			
	$\delta_i^{CholCov}$	$\delta_i^{RCOV}$	$\delta_i^{DCC}$	$\delta_i^{CholCov}$	$\delta_i^{RCOV}$	$\delta_i^{CholCov}$	$\delta_i^{RCOV}$	$\delta_i^{DCC}$	
Mean	0.724	0.064	0.258	0.922	0.122	0.976	0.995	1.017	
Stdev.	0.199	0.142	0.230	0.131	0.151	0.103	0.325	0.086	
5%	0.461	-0.178	-0.083	0.702	-0.102	0.841	0.394	0.880	
Median	0.713	0.077	0.268	0.912	0.138	0.957	0.997	1.018	
95%	1.061	0.293	0.637	1.115	0.368	1.117	1.457	1.137	
<i>Panel B:</i> Rejection Frequencies									
$H_{CholCov} : 0.577$ $H_{RCOV} : 1.000$ $H_{DCC} : 0.962$ $H'_{CholCov} : 0.404$ $H'_{RCOV} : 1.000$									
<i>Panel C:</i> Frequency in Model Confidence Set									
CholCov: 1.000			RCOV: 0.038			DCC: 0.442			

*Note:* Comparison of the betas obtained from CholCov and RCOV cRDCC forecasts, and cDCC forecasts based on daily returns. Panel A presents descriptive statistics of the estimated coefficients of regression (3.44). Panel B shows the rejection frequencies for the hypothesis outlined in the text and Panel C reports the frequency that each of the models are included in the MCS in the beta hedging exercise.

Second, we compare the betas in terms of a beta hedging tracking exercise. We compute the time series  $(r_{k,t} - Z_{k,t|t-1}^{CholCov})$ ,  $(r_{k,t} - Z_{k,t|t-1}^{RCOV})$  and  $(r_{k,t} - Z_{k,t|t-1}^{DCC})$  for each asset  $k$ , and seek the series with the smallest sample variance by means of the Model Confidence Set of Hansen, Lunde, and Nason (2011).

The results are reported in Table 3.6. The first two panels show the results from the regression based test, whereas Panel C shows the beta hedging exercise. Panel A shows descriptive statistics of the cross-sectional variation of the estimated  $\delta$  coefficients across the 52 assets. The beta stemming from the CholCov estimates has the highest weight on average and cannot be rejected as the ‘superior’ model for over 40 percent of the assets. It almost completely encompasses the explanatory power of beta forecasts based on the RCOV, which obtains very low loadings and is always rejected. The model based on daily returns has slightly higher loadings, but is also almost always rejected as the superior model.

In the direct comparison between the CholCov and RCOV forecasts, the CholCov provides superior results, and cannot be rejected in almost 60 percent of the cases.

The beta hedging exercise provides the same conclusion. We obtain Model Confidence Sets, which contain the set of models with the smallest tracking error with a probability of no less than 90%. The CholCov tracking portfolios are never rejected, and the DCC based portfolios are in slightly over 40% of the sets.

## 3.8 Conclusions

We propose an ex-post estimator of the integrated covariance that uses the Cholesky decomposition to obtain an estimate that is ensured positive semidefinite. The elements are estimated sequentially, on an increasing set of series. As such, the estimator uses many more observations than the traditional multivariate estimators, but fewer than pairwise estimators. The Cholcov is flexible and can use any other estimator for the intermediate calculations, adopting their robustness properties. The asymptotic distribution is derived for the constant volatility, no-noise case, but with asynchronicity. Simulations confirm its properties and demonstrate that the resulting estimates are accurate and better conditioned than full-dimensional or composite estimates.

In an empirical application we use the CholCov in a portfolio setting which requires the covariance matrix estimate to be positive semidefinite. The problem is moderately large, involving over 50 stocks. Using an appropriate dynamic model, which allows for long memory in the variances, we forecast portfolio Value-at-Risk and are unable to reject the model, using standard normal quantiles. This is in contrast to models based on daily returns and dynamic models on the CholCov not allowing for long memory, which are rejected by the data.

As a second application, we forecast conditional betas with respect to ten sector ETFs. We evaluate them using a regression-based comparison and a beta hedging tracking exercise. We find that the forecasts based on the CholCov provide the best results in terms of both evaluation criteria, compared to forecasts based on RCOV and daily returns.

## Appendix A: The Pecatti and Nualart's fourth moment theorem

In this appendix, we present a version of Pecatti and Nualart's fourth moment theorem (see Nualart and Peccati, 2005 and Peccati and Tudor, 2005) which plays a key role in the proof of Theorem 1.

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space and  $(W_t)_{0 \leq t \leq 1}$  a  $d$ -dimensional Brownian motion satisfying that  $\mathcal{F} = ((W_t)_{0 \leq t \leq 1})$ , the  $\sigma$ -algebra generated by  $(W_t)_{0 \leq t \leq 1}$ . Recall that the underlying log-price process  $Y_t$  is assumed to follow a multivariate Brownian Semi-Martingale with constant covariance, i.e.,

$$dY_t = \sigma_0 dW_t.$$

Observe that since  $\sigma_0 \sigma_0'$  is assumed to be positive definite, then  $\mathcal{F} = ((Y_t)_{0 \leq t \leq 1})$ .

Let  $T = [0, 1] \times \{1, \dots, d\}$ , and  $\mu(dsdx) = ds \#(dx)$ , where  $\#(dx)$  is the counting measure on  $\{1, \dots, d\}$ . Define

$$I_1^W(\xi) := \sum_{i=1}^d \int_0^1 \xi(s, i) dW_s^i \quad (3.45)$$

$$I_2^W(\varphi) := 2 \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \int_0^r \varphi[(s, i), (r, j)] dW_s^i dW_r^j \quad (3.46)$$

with  $\xi : T \rightarrow \mathbb{R}$   $\mu$ -square integrable, and  $\varphi : T \times T \rightarrow \mathbb{R}$   $(\mu \times \mu)$ -square integrable such that

$$\varphi[(s, i), (r, j)] = \varphi[(r, j), (s, i)], \quad (r, j), (s, i) \in T. \quad (3.47)$$

We define in an analogous way  $I_1^Y$  and  $I_2^Y$ . The collection of  $(\mu \times \mu)$ -square integrable functions satisfying (3.47) will be denoted by  $\mathcal{H}^{\odot 2}$ .

The following result is an ad hoc simplification of Nualart and Peccati (2005) and Peccati and Tudor (2005):

**Theorem 2 (Pecatti and Tudor, 2005)** *Let  $m \geq 2$ . Consider a sequence of functions  $\{\varphi_n^1, \dots, \varphi_n^m\}_{n \geq 1}$  such that  $\varphi_n^i \in \mathcal{H}^{\odot 2}$  and for every  $i, j = 1, \dots, m$ , the limit*

$$\lim_{n \rightarrow \infty} \mathbb{E} [I_2^W(\varphi_n^i) I_2^W(\varphi_n^j)] = C(i, j),$$

*exists. Then the following statements are equivalent*

1. *The vector  $(I_2^W(\varphi_n^i))_{i=1}^m$  converges in distributions to a  $m$ -dimensional Gaussian vector  $N = (N_1, \dots, N_m)'$  with covariance matrix  $C = (C(i, j))_{i, j=1, \dots, m}$ ;*
2. *For every  $j = 1, \dots, m$ ,  $I_2^W(\varphi_n^j) \xrightarrow{d} N_j$ ;*

Additionally for our proof of Theorem 1, we will need the following result.

**Proposition 3**

1. We have

$$I_1^Y (\xi^1) I_1^Y (\xi^2) = I_2^W (\sigma_0 \xi^1 \tilde{\otimes} \sigma_0 \xi^2) + \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \xi^1 (s, i) \Sigma_{i,j}^0 \xi^2 (s, j) ds, \quad (3.48)$$

where  $\sigma_0 \xi (s, i) = \sum_{x=1}^d \sigma_{i,x}^0 \xi (s, x)$  and

$$\xi^1 \tilde{\otimes} \xi^2 [(s, i), (r, j)] = \frac{1}{2} [\xi^1 (s, i) \xi^2 (r, j) + \xi^1 (r, j) \xi^2 (s, i)],$$

i.e. the symmetric tensor product.

2. For every  $\varphi^1, \varphi^2 \in \mathcal{H}^{\odot 2}$

$$\mathbb{E} [I_2^W (\varphi^1) I_2^W (\varphi^2)] = 2 \sum_{i=1}^d \sum_{j=1}^d \int_0^1 \int_0^1 \varphi^1 [(r, j), (s, i)] \varphi^2 [(r, j), (s, i)] ds dr. \quad (3.49)$$

**Proof.** The first part is an application of Itô's formula and the fact that

$$dY_s^i = \sum_{j=1}^d \sigma_{ij}^0 dW_s^j.$$

A general proof for part 2 can be found in Nualart (2006). ■

## Appendix B: Proofs

**Proof of Proposition 1.** In view that the mapping  $(H, G) \mapsto HGH^T$  is a continuously differentiable bijection between  $\mathbb{M}_{d, chol}^+$  and  $\mathbb{M}_d^+$  and in this case  $\widehat{\Sigma}$  is the realized covariance, we only need to check that  $(\widehat{H}, \widehat{G}) = (\overline{H}, \overline{G})$ . We only consider the bivariate case but the general case follows by induction. Trivially, we have that  $\widehat{g}_{11} = \overline{g}_{11}$ . Now, since

$$\begin{aligned} \widehat{h}_{21} &= \frac{\sum_{j=1}^n r_j^{(2)}(T) f_j^{(1)}(T)}{\widehat{g}_{11}} \\ &= \frac{\sum_{j=1}^n r_j^{(2)}(T) r_j^{(1)}(T)}{\widehat{g}_{11}}, \end{aligned}$$

we obtain that  $\widehat{h}_{21} = \overline{h}_{21}$ , or in other words  $\widehat{H} = \overline{H}$ . Finally, in view that for  $i = 1, \dots, n$

$$\overline{f}_j^{(2)}(T) = r_j^{(2)}(T) - \frac{\overline{\Sigma}_{21}}{\overline{\Sigma}_{11}} r_j^{(1)}(T),$$

we get

$$\begin{aligned} \widehat{g}_{22} &= \sum_{j=1}^n \left[ \overline{f}_j^{(2)}(T) \right]^2 \\ &= \sum_{j=1}^n \left[ r_j^{(2)}(T) - \frac{\overline{\Sigma}_{21}}{\overline{\Sigma}_{11}} r_j^{(1)}(T) \right]^2 \\ &= \overline{\Sigma}_{22} - 2 \frac{\overline{\Sigma}_{21}}{\overline{\Sigma}_{11}} \sum_{j=1}^n r_j^{(1)}(T) r_j^{(2)}(T) + \left( \frac{\overline{\Sigma}_{21}}{\overline{\Sigma}_{11}} \right)^2 \sum_{i=1}^n \left[ r_j^{(1)}(T) \right]^2 \\ &= \overline{\Sigma}_{22} - \frac{(\overline{\Sigma}_{21})^2}{\overline{\Sigma}_{11}} = \overline{g}_{22}, \end{aligned}$$

which concludes the proof. ■

**Proof of Proposition 2.** Following the same reasoning as in Remark 1, we have that

$$\left\| \widehat{\Sigma}(k) - \widetilde{\Sigma}(k) \right\| \leq \sup \left| \frac{1}{\Delta_j(T_k) N_k} - 1 \right| \left\| \widehat{\Sigma}(k) \right\|, \quad (3.50)$$

where  $\widehat{\Sigma}(k)$  is the realized covariance of the process  $\pi_k(X_t)$ . Note that from (3.22) and (3.23), we get

$$\widehat{\Sigma}(k) - \Sigma_0(k) = \Pi_k \left( \widehat{\Sigma} - \Sigma_0 \right) \Pi_k',$$

and

$$\text{rvech} \left( \widehat{\Sigma}(k) - \Sigma_0(k) \right) = B_k \text{rvech} \left( \widehat{\Sigma} - \Sigma_0 \right),$$

for some matrix  $B_k$ . It follows that  $\left\| \widehat{\Sigma}(k) - \widetilde{\Sigma}(k) \right\| = o_{\mathbb{P}} \left( N_k(n)^{-1/2} \right)$ . The first conclu-

sion of this proposition follows trivially from this and the fact that

$$q(k) - q_0(k) = \Pi_{k(k+1)/2}(\widehat{q} - q_0),$$

with  $\widehat{q}$  as in Proposition 1. The final result follows from the following equation

$$\widehat{q}_k - q_k^0 = A_k[\widehat{q}(k) - q_0(k)], \quad k = 1, \dots, d.$$

■

**Proof of Theorem 1.** From Proposition 2,  $\beta_k$  converges to a normal random vector, for  $k = 1, \dots, d$ . Thus, for any  $k = 1, \dots, d$  and  $i = 1, \dots, k$ ,  $\beta_k(i)$  is asymptotically normal. To show the desired result, we use Theorem 2. Therefore, in order to conclude the proof, we only need to show that  $\beta_k(i) = I_2^W(\varphi)$  for some  $\varphi \in \mathcal{H}^{\odot 2}$  depending on  $(i, k)$ , and that the limit  $\lim_{n \rightarrow \infty} \mathbb{E}[\beta_k(i)\beta_{k+l}(j)]$  exists and is given by (3.32), for every  $k = 1, \dots, d$ ,  $l = 1, \dots, d - k$ ,  $i = 1, \dots, k$ , and  $j = 1, \dots, k + l$ .

Let us first verify that  $\beta_k(i) = I_2^W(\varphi)$ . Let  $\widehat{\Sigma}_{ij}^k$  be the  $ij$ th element of  $\widehat{\Sigma}(k)$ . Then

$$\begin{aligned} \widehat{\Sigma}_{ij}^k &= \sum_{l=1}^{N_k} \left( Y_{t_l^i}^i - Y_{t_{l-1}^i}^i \right) \left( Y_{t_l^j}^j - Y_{t_{l-1}^j}^j \right) \\ &= \sum_{l=1}^{N_k} I_1^Y[\xi^{i,l}(T_k)] I_1^Y[\xi^{j,l}(T_k)], \end{aligned}$$

where

$$\xi^{i,l}(T_k)(s, x) := \mathbf{1}_{A_l^k}(s) \delta_i(x), \quad i = 1, \dots, k,$$

with  $A_l^k$  as in (3.35) and  $\delta_i$  the Dirac delta on  $i$ . Using (3.48) in Proposition 3, we have that

$$\begin{aligned} I_1^Y[\xi^{i,l}(T_k)] I_1^Y[\xi^{j,l}(T_k)] &= I_2^W(\sigma_0 \xi^{i,l}(T_k) \widetilde{\otimes} \sigma_0 \xi^{j,l}(T_k)) \\ &\quad + \sum_{x=1}^d \sum_{y=1}^d \int_0^1 \xi^{i,l}(T_k)(s, x) \Sigma_{x,y}^0 \xi^{j,l}(T_k)(s, y) ds, \end{aligned}$$

where  $\Sigma_{ij}^0$  is the  $ij$ th element of  $\Sigma_0$ . In view that

$$\sum_{x=1}^d \sum_{y=1}^d \int_0^1 \xi^{i,l}(T_k)(s, x) \Sigma_{x,y}^0 \xi^{j,l}(T_k)(s, y) ds = \Delta_l(T_k) \Sigma_{ij}^0,$$

### 3. CHOLCOV

we deduce that

$$\begin{aligned}\beta_k(i) &= \sqrt{N_k} \left( \widehat{\Sigma}_{ki}^k - \Sigma_{ki}^0 \right) \\ &= \sqrt{N_k} \sum_{l=1}^{N_k} I_2^W \left( \sigma_0 \xi^{k,l}(T_k) \widetilde{\otimes} \sigma_0 \xi^{i,l}(T_k) \right).\end{aligned}\quad (3.51)$$

We conclude that  $\beta_k(i) = I_2^W(\varphi)$ , with  $\varphi = \sqrt{N_k} \sum_{l=1}^{N_k} (\sigma_0 \xi^{k,l}(T_k) \widetilde{\otimes} \sigma_0 \xi^{i,l}(T_k)) \in \mathcal{H}^{\odot 2}$ .

Finally, we need to show that the limit  $\lim_{n \rightarrow \infty} \mathbb{E}[\beta_k(i) \beta_{k+l}(j)]$  exists. From (3.51), for every  $k = 1, \dots, d$ ,  $i = 1, \dots, k$ ,  $l = 1, \dots, d - k$ ,  $j = 1, \dots, k + l$ .

$$\mathbb{E}[\beta_k(i) \beta_{k+l}(j)] = \sqrt{N_k N_{k+l}} \mathbb{E} \left\{ \left[ \sum_{m=1}^{N_k} I_2^W \left( \sigma_0 \xi^{k,m}(T_k) \widetilde{\otimes} \sigma_0 \xi^{i,m}(T_k) \right) \right] \times \left[ \sum_{p=1}^{N_{k+l}} I_2^W \left( \sigma_0 \xi^{k+l,p}(T_{k+l}) \widetilde{\otimes} \sigma_0 \xi^{j,p}(T_{k+l}) \right) \right] \right\} \quad (3.52)$$

$$= \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \mathbb{E} \left[ I_2^W \left( \sigma_0 \xi^{k,m}(T_k) \widetilde{\otimes} \sigma_0 \xi^{i,m}(T_k) \right) \right] \quad (3.53)$$

$$I_2^W \left( \sigma_0 \xi^{k+l,p}(T_{k+l}) \widetilde{\otimes} \sigma_0 \xi^{j,p}(T_{k+l}) \right) \Big]. \quad (3.54)$$

Furthermore, since

$$\sigma_0 \xi^{k,m}(T_k) \widetilde{\otimes} \sigma_0 \xi^{i,m}(T_k) [(s, x), (r, y)] = \frac{1}{2} \mathbf{1}_{\Lambda_m^k}(s) \mathbf{1}_{\Lambda_m^k}(r) (\sigma_{kx}^0 \sigma_{iy}^0 + \sigma_{ky}^0 \sigma_{ix}^0),$$

we get from (3.49)

$$\begin{aligned}\mathbb{E} \left[ I_2^W \left( \sigma_0 \xi^{k,m}(T_k) \widetilde{\otimes} \sigma_0 \xi^{i,m}(T_k) \right) I_2^W \left( \sigma_0 \xi^{k+l,p}(T_{k+l}) \widetilde{\otimes} \sigma_0 \xi^{j,p}(T_{k+l}) \right) \right] \\ = \Psi((k, i), (k+l, j)) \text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k)^2,\end{aligned}$$

with  $\Psi((k, i), (k+l, j))$  as in (3.34). This implies that

$$\mathbb{E}[\beta_k(i) \beta_{k+l}(j)] = \Psi((k, i), (k+l, j)) \rho_{k,k+l}(n),$$

where

$$\rho_{k,k+l}(n) := \sqrt{N_k N_{k+l}} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k)^2.$$

If  $\Psi((k, i), (k+l, j)) = 0$ , the limit  $\lim_{n \rightarrow \infty} \mathbb{E}[\beta_k(i) \beta_{k+l}(j)]$  exists trivially, so without loss of generality we may and do assume that  $\Psi((k, i), (k+l, j)) \neq 0$ . Consequently,  $\lim_{n \rightarrow \infty} \mathbb{E}[\beta_k(i) \beta_{k+l}(j)]$  exists if and only if the limit  $\lim_{n \rightarrow \infty} \rho_{k,k+l}(n)$  exists and it is finite.

From Lemma 2

$$\begin{aligned}\rho_{k,k+l}(n) &= \tilde{\rho}_{k,k+l}(n) + \sqrt{N_k N_{k+l}} o\left[(N_k N_{k+l})^{-1}\right] \\ &= \tilde{\rho}_{k,k+l}(n) + o(1),\end{aligned}\tag{3.55}$$

where

$$\tilde{\rho}_{k,k+l}(n) := \sqrt{N_{k+l} N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} \text{Leb} \left( \tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right)^2,$$

with

$$\tilde{\Lambda}_m^k = \left( \frac{(m-1)}{N_k}, \frac{m}{N_k} \right], \quad m = 1, \dots, N_k.$$

Lemma 3 below guarantees that the limit of the sequence  $\tilde{\rho}_{k,k+l}(n)$  exists and is finite, so  $\rho_{k,k+l}(n)$ .  $\blacksquare$

**Lemma 1** *Let*

$$\tilde{F}_n^{k,k+l} := \left\{ (m, p) : \text{Leb} \left( \tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k \right) > 0 \right\}.$$

where  $\tilde{\Lambda}_m^k = \left( \frac{m-1}{N_k}, \frac{m}{N_k} \right]$ . Then

$$\#\tilde{F}_n^{k,k+l} = cN_k + O(1),$$

where  $\#$  denotes the counting measure, for some constant  $c > 0$  that does not depend on  $n$ .

**Proof.** If  $N_k = N_{k+l}$ , the result is trivial, so without loss of generality, we may and do assume that  $N_k > N_{k+l}$ . Due to Assumption 1, we may and do assume that  $N_k = \lfloor \frac{N_1}{k} \rfloor$ . In this case

$$\tilde{\Lambda}_m^k = \bigcup_{j=1}^k \tilde{\Lambda}_{k(m-1)+j}^1,\tag{3.56}$$

for every  $m = 1, \dots, N_k$ . Moreover, in this case. Note that

$$\tilde{\Lambda}_m^k \cap \tilde{\Lambda}_p^{k+l} = \bigcup_{j=1}^{k+l} \bigcup_{i=1}^k [\Lambda_{k(m-1)+i}^1 \cap \Lambda_{(k+l)(p-1)+j}^1].\tag{3.57}$$

Since  $\tilde{\Lambda}_{k(m-1)+i}^1, \tilde{\Lambda}_{(k+l)(p-1)+j}^1$  belongs to  $T_1$ , we deduce that they are disjoint unless

$$km - (k-i) = (k+l)(p-1) + j.\tag{3.58}$$

Hence, to obtain (3.57), for a given a pair  $(m, n)$ , we need to obtain all the solutions  $(i, j)$  for which (3.58) holds. Consider the following three cases:

### 3. CHOLCOV

---

**Case  $m = p$ :** If  $m = p$ , (3.58) becomes

$$(k + l)(m - 1) + j = k(m - 1) + i,$$

which implies that

$$j = i - l(m - 1). \quad (3.59)$$

For  $(i, j)$  to be a solution of (3.58), it must satisfy

$$1 \leq j \leq k + l \quad \text{and} \quad 1 \leq i \leq k. \quad (3.60)$$

Now, if  $i \leq k$ , we have trivially that  $j \leq k + l$ . On the other hand, if  $i \leq l(m - 1)$  we get that  $j \leq 0$ . Therefore, the pair in (3.59) solves (3.58) if and only if  $l(m - 1) < i \leq k$  which in turn holds if and only if  $m \leq \lfloor \frac{k}{l} \rfloor$ . It follows that

$$\Lambda_p^k \cap \Lambda_p^{k+l} = \begin{cases} \Lambda_m^k & \text{if } m = p \text{ and } m \leq \lfloor \frac{k}{l} \rfloor; \\ \emptyset & \text{otherwise.} \end{cases}$$

**Case  $m < p$ :** We now show that if  $m < p$ , there does not exist any pair  $(i, j)$  fulfilling (3.58) and (3.60). Indeed, let  $p = m + 1 + x$ , for some  $x > 1$ . Then (3.58) becomes

$$km - (k - i) = (k + l)(m + x) + j,$$

which implies that, for any  $1 \leq i \leq k$ ,

$$j = -[(k - i) + (k + l)x + lm] < 0,$$

i.e. there is no solution for (3.58) and (3.60). We conclude that in this case  $\Lambda_m^k \cap \Lambda_p^{k+l} = \emptyset$ .

**Case  $m > p$ :** Finally, assume that  $m > p$ . Put  $m = p + x$ , for  $x = 1, \dots, N_k - 1$ . Reasoning as before, (3.58) has a solution if and only if  $1 \leq j \leq k + l$  and

$$1 \leq (k + l)(p - 1) + j - k(m - 1) \leq k.$$

Therefore, if for  $x = 1, \dots, N_k - 1$  we choose  $p$  such that

$$k(x - 1) + 1 < lp \leq k(x - 1) + l,$$

then  $(i, j)$  as before, solves (3.58). We conclude that in this case

$$\Lambda_p^k \cap \Lambda_p^{k+l} = \begin{cases} \Lambda_{(p-1)(k+l)+x}^1 & \text{if } k(x - 1) + 1 < lp \leq k(x - 1) + l; \\ \emptyset & \text{otherwise.} \end{cases}$$

■

**Lemma 2** *Under Assumption 1, we have that*

1. We have that, for any

$$\text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k) = \text{Leb}(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k) + o\left[(N_k N_{k+l})^{-1/2}\right].$$

2. By denoting  $\#$  the counting measure, we have that

$$\#F_n^{k,l} = \#\tilde{F}_n^{k,l} + o(1).$$

**Proof.** Let us start by showing that 1. holds. First note that in general

$$\text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k) = \max\{t_m^k \wedge t_p^{k+l} - t_{m-1}^k \vee t_{p-1}^{k+l}, 0\},$$

where  $\wedge$  and  $\vee$  stands for the maximum and minimum of two numbers, respectively. Due to the formulae

$$\begin{aligned} a \wedge b &= \frac{1}{2}(a + b - |a - b|), \\ c \vee d &= \frac{1}{2}(c + d + |c - d|), \end{aligned}$$

we deduce that

$$\text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k) = \frac{1}{2} \max\{\Delta_m^k + \Delta_p^{k+l} - (|t_m^k - t_p^{k+l}| + |t_{m-1}^k - t_{p-1}^{k+l}|), 0\}. \quad (3.61)$$

Therefore

$$\begin{aligned} 2 \left| \text{Leb}(\Lambda_p^{k+l} \cap \Lambda_m^k) - \text{Leb}(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k) \right| &\leq \left| \Delta_m^k - \frac{1}{N_k} \right| + \left| \Delta_p^{k+l} - \frac{1}{N_{k+l}} \right| \\ &+ \left| t_m^k - \frac{m}{N_k} \right| + \left| t_p^{k+l} - \frac{p}{N_{k+l}} \right| \\ &+ \left| t_{m-1}^k - \frac{m-1}{N_k} \right| + \left| t_{p-1}^{k+l} - \frac{p-1}{N_{k+l}} \right| \end{aligned} \quad (3.62)$$

It is immediate from Assumption 1, that as  $n \rightarrow \infty$

$$\sqrt{N_k N_{k+l}} \left| \Delta_m^k - \frac{1}{N_k} \right| \rightarrow 0.$$

### 3. CHOLCOV

On the other hand, once again, applying Assumption 1, we see that

$$\begin{aligned} \sqrt{N_k N_{k+l}} \left| t_m^k - \frac{m}{N_k} \right| &\leq \left( \frac{N_{k+l}}{N_k} \right)^{1/2} \sum_{j=1}^m |\Delta_m^k - 1| \\ &\leq m \left( \frac{N_{k+l}}{N_k} \right)^{1/2} \sup_m |\Delta_m^k - 1| \rightarrow 0, \quad n \rightarrow \infty. \end{aligned} \quad (3.63)$$

Part 1 is obtained as a combination of (3.62) and (3.63).

Now we proceed to verify part 2. To do this, it is enough to show that as  $n \rightarrow \infty$

$$\# \left( F_n^{k,l} \setminus \tilde{F}_n^{k,l} \right) + \# \left( \tilde{F}_n^{k,l} \setminus F_n^{k,l} \right) \rightarrow 0. \quad (3.64)$$

In view

$$\limsup_{n \rightarrow \infty} F_n^{k,l} \setminus \tilde{F}_n^{k,l} = \limsup_{n \rightarrow \infty} \tilde{F}_n^{k,l} \setminus F_n^{k,l} = \emptyset,$$

in order to finish the proof, we only need to verify that for some  $n \geq 1$

$$\# \left( \bigcup_{j \geq n} F_j^{k,l} \setminus \tilde{F}_j^{k,l} \right) < \infty \text{ and } \# \left( \bigcup_{j \geq n} \tilde{F}_j^{k,l} \setminus F_j^{k,l} \right) < \infty,$$

because in this case

$$\begin{aligned} \limsup_{n \rightarrow \infty} \# \left( F_n^{k,l} \setminus \tilde{F}_n^{k,l} \right) &\leq \# \left[ \limsup_{n \rightarrow \infty} \left( F_n^{k,l} \setminus \tilde{F}_n^{k,l} \right) \right] = 0, \\ \limsup_{n \rightarrow \infty} \# \left( \tilde{F}_n^{k,l} \setminus F_n^{k,l} \right) &\leq \# \left[ \limsup_{n \rightarrow \infty} \left( \tilde{F}_n^{k,l} \setminus F_n^{k,l} \right) \right] = 0, \end{aligned}$$

which implies (3.64). We only verify that  $\# \left( \bigcup_{M \geq n} F_n^{k,l} \setminus \tilde{F}_n^{k,l} \right) < \infty$ , the other case is analogous. From Lemma 1

$$\begin{aligned} \bigcup_{j \geq n} F_j^{k,l} \setminus \tilde{F}_j^{k,l} &= \bigcup_{j \geq n} \left\{ (m, m) : \left\lfloor \frac{k}{l} \right\rfloor < m \leq N_k, \text{Leb} \left( \Lambda_m^{k+l} \cap \Lambda_m^k \right) > 0 \right\} \cup \\ &\quad \bigcup_{j \geq n} \bigcup_{1 \leq x \leq N_k} \left\{ (x+p, p) : k(x-1) + 1 < lp \leq k(x-1) + l, \right. \\ &\quad \left. \text{Leb} \left( \Lambda_p^{k+l} \cap \Lambda_m^k \right) > 0 \right\} \end{aligned}$$

We claim that  $\# \left( \bigcup_{j \geq n} \left\{ (m, m) : \left\lfloor \frac{k}{l} \right\rfloor < m \leq N_k, \text{Leb} \left( \Lambda_m^{k+l} \cap \Lambda_m^k \right) > 0 \right\} \right) < \infty$ . Indeed, from the first part and (3.61), if  $(m, m) \in F_n^{k,l} \setminus \tilde{F}_n^{k,l}$  for some  $n$ , then

$$0 < \text{Leb} \left( \Lambda_m^{k+l} \cap \Lambda_m^k \right) = \frac{m}{N_k} - \frac{m-1}{N_{k+l}} + o \left[ \left( N_k N_{k+l} \right)^{-1/2} \right],$$

which implies that

$$m < \frac{N_k}{N_{k+l} - N_k} + o(1).$$

Assumption 1 guarantees that  $m$  is uniformly bounded, i.e.  $m < M^*$  for any  $(m, m) \in$

$\cup_{j \geq n} \{(m, m) : \lfloor \frac{k}{l} \rfloor < m \leq N_k, Leb(\Lambda_m^{k+l} \cap \Lambda_m^k) > 0\}$ . It follows that

$$\cup_{j \geq n} \left\{ (m, m) : \left\lfloor \frac{k}{l} \right\rfloor < m \leq N_k, Leb(\Lambda_m^{k+l} \cap \Lambda_m^k) > 0 \right\} \subset \left\{ (m, m) : \left\lfloor \frac{k}{l} \right\rfloor < m < M \right\},$$

which shows our claim. Reasoning in an analogous way, we can show that

$$\begin{aligned} \#(\cup_{j \geq n} \cup_{1 \leq x \leq N_k} \{(x+p, p) : pk(x-1) + 1 < lp \leq k(x-1) + l, \\ Leb(\Lambda_p^{k+l} \cap \Lambda_m^k) > 0\}) < \infty. \end{aligned}$$

■

**Lemma 3** *Under the notation of the previous lemma, we have that*

$$\lim_{n \rightarrow \infty} \sqrt{N_{k+l} N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} Leb(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k)^2 < \infty.$$

**Proof.** Without loss of generality we may and do assume that  $N_1 = n$ ,  $n \in \mathbb{N}$ . This implies that  $N_k = \lfloor \frac{n}{k} \rfloor$ . Let  $\tilde{\rho}_{k,k+l}(n) = \sqrt{N_{k+l} N_k} \sum_{m=1}^{N_k} \sum_{p=1}^{N_{k+l}} Leb(\tilde{\Lambda}_p^{k+l} \cap \tilde{\Lambda}_m^k)^2$ . From the previous lemma, it follows that

$$\begin{aligned} \sum_{m=1}^{\lfloor n/k \rfloor} \sum_{p=1}^{\lfloor n/(k+l) \rfloor} Leb(\Lambda_p^{k+l} \cap \Lambda_m^k)^2 &= \sum_{m=1}^{\lfloor n/k \rfloor} Leb(\Lambda_m^{k+l} \cap \Lambda_m^k)^2 + \sum_{m=1}^{\lfloor n/k \rfloor} \sum_{p=1}^m Leb(\Lambda_p^{k+l} \cap \Lambda_m^k)^2 \\ &\quad + \sum_{m=1}^{\lfloor n/k \rfloor} \sum_{p=m}^{\lfloor n/(k+l) \rfloor} Leb(\Lambda_p^{k+l} \cap \Lambda_m^k)^2 \\ &= \sum_{m=1}^{\lfloor n/k \rfloor} \left[ \sum_{x=1}^{k+l} \sum_{\lfloor \frac{km+l+1-x}{k+l} \rfloor \leq p \leq \lfloor \frac{km+k+l-x}{k+l} \rfloor} Leb(\Lambda_{(p-1)(k+l)+x}^1) \right]^2 \\ &\quad + \sum_{m=1}^{\lfloor k/l \rfloor} Leb(\Lambda_m^k)^2 \\ &= \frac{1}{n^2} \sum_{m=1}^{\lfloor n/k \rfloor} \left[ \sum_{x=1}^{k+l} \left( \left\lfloor \frac{km+k+l-x}{k+l} \right\rfloor \right. \right. \\ &\quad \left. \left. - \left\lfloor \frac{km+l+1-x}{k+l} \right\rfloor \right) \right]^2 + \left\lfloor \frac{k}{l} \right\rfloor \frac{1}{n^2}. \end{aligned}$$

We conclude that for  $n \geq 2d$

$$\begin{aligned} \tilde{\rho}_{k,k+l}(n) &= \sqrt{N_k N_{k+l}} \left\lfloor \frac{k}{l} \right\rfloor \frac{1}{n^2} \\ &\quad + \frac{\sqrt{N_k N_{k+l}}}{n^2} \sum_{m=1}^{\lfloor n/k \rfloor} \left( \sum_{x=1}^{k+l} \left\lfloor \frac{km + k + l - x}{k + l} \right\rfloor - \left\lfloor \frac{km + l + 1 - x}{k + l} \right\rfloor \right)^2. \end{aligned}$$

Finally, since

$$\frac{\sqrt{N_k N_{k+l}}}{n^2} \sum_{m=1}^{\lfloor n/k \rfloor} \left( \sum_{x=1}^{k+l} \left\lfloor \frac{km + k + l - x}{k + l} \right\rfloor - \left\lfloor \frac{km + l + 1 - x}{k + l} \right\rfloor \right)^2 \approx \frac{\sqrt{N_k N_{k+l}}}{n^2} \lfloor n/k \rfloor$$

we conclude that  $\lim_{n \rightarrow \infty} \tilde{\rho}_{k,k+l}(n) < \infty$ . ■

## Appendix C: The score function and Fisher Information of the ML estimators under the assumptions of Proposition 1

This section is devoted to deriving the score function, the Hessian matrix, and the Fisher information for the ML estimator of  $q_0$  under the assumptions of Proposition 1. For notational clarity we will omit the 0 subscript and the grid dependence of the durations in the remainder of this section. We derive these quantities with respect to generic  $g_{kk}$  and  $h_{kl}$  elements.

**Remark 6** Note that  $f^{(k)} = r^{(k)} - \sum_{l=1}^{k-1} h_{kl} f^{(l)}$  when  $k > l$ . As such,  $\frac{\partial f^{(k)}}{\partial h_{lm}} = -\sum_{i=1}^{k-1} h_{ki} \frac{\partial f^{(i)}}{\partial h_{lm}}$  for  $k > l > m$ , and 0 otherwise.

We first compute the score functions:

$$\frac{\partial \bar{Q}_n}{\partial g_{kk}} = -\frac{1}{2} \sum_{j=1}^N \left[ \frac{1}{\Delta_j g_{kk}} - \frac{f_j^{(k)2}}{\Delta_j g_{kk}^2} \right], \quad (3.65)$$

$$\frac{\partial \bar{Q}_n}{\partial h_{kl}} = -\frac{1}{2} \left[ \sum_{i=1}^d \sum_{j=1}^N \frac{2f_j^{(i)}}{\Delta_j g_{ii}} \frac{\partial f_j^{(i)}}{\partial h_{kl}} \right]. \quad (3.66)$$

And the Hessian is as follows:

$$\frac{\partial^2 \bar{Q}_n}{\partial g_{kk} \partial g_{kk}} = \sum_{j=1}^N \left[ \frac{1}{2\Delta_j g_{kk}^2} - \frac{f_j^{(k)2}}{\Delta_j g_{kk}^3} \right] \quad (3.67)$$

$$\frac{\partial^2 \bar{Q}_n}{\partial h_{mn} \partial g_{kk}} = \frac{\partial^2 \bar{Q}_n}{\partial g_{kk} \partial h_{mn}} = - \sum_{j=1}^N \left[ \frac{f_j^{(k)}}{\Delta_j g_{kk}^2} \frac{\partial f_j^{(k)}}{\partial h_{mn}} \right] \quad (3.68)$$

$$\frac{\partial^2 \bar{Q}_n}{\partial h_{kl} \partial h_{mn}} = - \sum_{j=1}^N \left[ \sum_{i=1}^d \frac{1}{\Delta_j g_{ii}} \left[ \frac{\partial f_j^{(i)}}{\partial h_{mn}} \frac{\partial f_j^{(i)}}{\partial h_{kl}} + f_j^{(i)} \frac{\partial^2 f_j^{(i)}}{\partial h_{kl} \partial h_{mn}} \right] \right]. \quad (3.69)$$

We compute the components of the Fisher matrix as the negative of the expected value of the Hessian.

$$-\mathbb{E} \left[ \frac{\partial^2 \bar{Q}_n}{\partial g_{kk}^2} \right] = -\mathbb{E} \left[ \sum_{j=1}^N \frac{1}{2\Delta_j g_{kk}^2} - \frac{f_{t_j}^{(k)2}}{g_{kk}^3} \right] = \frac{1}{2g_{kk}^2} \quad \forall k \quad (3.70)$$

$$-\mathbb{E} \left[ \frac{\partial^2 \bar{Q}_n}{\partial g_{kk} \partial h_{mn}} \right] = -\mathbb{E} \left[ - \sum_{j=1}^N \left[ \frac{f_{t_j}^{(k)}}{\Delta_j g_{kk}^2} \frac{\partial f_{t_j}^{(k)}}{\partial h_{mn}} \right] \right] = 0, \quad (3.71)$$

since not a single partial derivative of  $f_j^{(k)}$  contains  $f_j^{(l)}$  and  $\mathbb{E}(f^{(k)} f^{(l)}) = 0$  for  $k \neq l$ .

Next let,  $p = \max(k, m)$ .

$$-\mathbb{E} \left[ \frac{\partial^2 \bar{Q}_n}{\partial h_{kl} \partial h_{mn}} \right] = \begin{cases} 0 & \text{if } l \neq n \\ \frac{g_{pp}}{g_{ll}} \frac{\partial^2 f^{(p)}}{\partial h_{kl} \partial f^{(l)}} \frac{\partial^2 f^{(p)}}{\partial h_{mn} \partial f^{(l)}} & \text{otherwise,} \end{cases} \quad (3.72)$$

where the last two partial derivatives are in fact functionals of  $h_{kl}$ .

# Chapter 4

## Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting\*

### 4.1 Introduction

Volatility, and volatility forecasting in particular, plays a crucial role in asset pricing and risk management. Access to accurate volatility forecasts is of the utmost importance for many financial market practitioners and regulators. A long list of competing GARCH and stochastic volatility type formulations have been proposed in the literature for estimating and forecasting financial market volatility. The latent nature of volatility invariably complicates implementation of these models. The specific parametric models hitherto proposed in the literature generally also do not perform well when estimated directly with intraday data, which is now readily available for many financial assets. To help circumvent these complications and more effectively exploit the information inherent in high-frequency data, Andersen, Bollerslev, Diebold, and Labys (2003) suggested the use of reduced form time series forecasting models for the daily so-called realized volatilities constructed from the summation of the squared high-frequency intraday returns.<sup>1</sup>

Set against this background, we propose a new family of easy-to-implement volatility forecasting models. The models directly exploit the asymptotic theory for high-frequency realized volatility estimation by explicitly allowing the dynamic parameters of the models, and in turn the forecasts constructed from the models, to vary with the degree of estimation error in the realized volatility measures.

The realized volatility for most financial assets is a highly persistent process. An-

---

\*This chapter is co-authored with Tim Bollerslev (Duke University) and Andrew Patton (Duke University and New York University) and is published in the *Journal of Econometrics*.

<sup>1</sup>The use of realized volatility for accurately measuring the true latent integrated volatility was originally proposed by Andersen and Bollerslev (1998), and this approach has now become very popular for both measuring, modeling and forecasting volatility; see, e.g., the discussion and many references in the recent survey by Andersen, Bollerslev, Christoffersen, and Diebold (2013).

dersen, Bollerslev, Diebold, and Labys (2003) originally suggested the use of fractionally integrated ARFIMA models for characterizing this strong dependency. However, the simple and easy-to-estimate approximate long-memory HAR (Heterogeneous AR) model of Corsi (2009) has arguably emerged as the preferred specification for realized volatility based forecasting. Empirically, the volatility forecasts constructed from the HAR model, and other related reduced-form time series models that treat the realized volatility as directly observable, generally perform much better than the forecasts from traditional parametric GARCH and stochastic volatility models.<sup>2</sup>

Under certain conditions, realized volatility ( $RV$ ) is consistent (as the sampling frequency goes to zero) for the true latent volatility, however in any given finite sample it is, of course, subject to measurement error. As such,  $RV$  will be equal to the sum of two components: the true latent Integrated Volatility ( $IV$ ) and a measurement error. The dynamic modeling of  $RV$  for the purposes of forecasting the true latent  $IV$  therefore suffers from a classical errors-in-variables problem. In most situations this leads to what is known as an attenuation bias, with the directly observable  $RV$  process being less persistent than the latent  $IV$  process. The degree to which this occurs obviously depends on the magnitude of the measurement errors; the greater the variance of the errors, the less persistent the observed process.<sup>3</sup>

Standard approaches for dealing with errors-in-variables problems treat the variance of the measurement error as constant through time.<sup>4</sup> In contrast, we explicitly take into account the temporal variation in the errors when modeling the realized volatility, building on the asymptotic distribution theory for the realized volatility measure developed by Barndorff-Nielsen and Shephard (2002). Intuitively, on days when the variance of the measurement error is small, the daily  $RV$  provides a stronger signal for next day's volatility than on days when the variance is large (with the opposite holding when the measurement error is large). Our new family of models exploits this heteroskedasticity in the error, by allowing for time-varying autoregressive parameters that are high when the variance of the realized volatility error is low, and adjusted downward on days when the variance is high and the signal is weak. Our adjustments are straightforward to implement

---

<sup>2</sup>Andersen, Bollerslev, and Meddahi (2004) and Sizova (2011) show how minor model misspecification can adversely affect the forecasts from tightly parameterized volatility models, thus providing a theoretical explanation for this superior reduced-form forecast performance.

<sup>3</sup>Alternative realized volatility estimators have been developed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008); Jacod, Li, Mykland, Podolskij, and Vetter (2009); Zhang, Mykland, and Aït-Sahalia (2005) among others. Forecasting in the presence of microstructure “noise” has also been studied by Aït-Sahalia and Mancini (2008); Andersen, Bollerslev, and Meddahi (2011); Bandi, Russell, and Yang (2013); Ghysels and Sinko (2011). The analysis below effectively abstracts from these complications, by considering a coarse five-minute sampling frequency and using simple  $RV$ . We consider some of these alternative estimators in Section 4.4.1 below.

<sup>4</sup>General results for the estimation of autoregressive processes with measurement error are discussed in Staudenmayer and Buonaccorsi (2005). Hansen and Lunde (2014a) have also recently advocated the use of standard instrumental variable techniques for estimating the persistence of the latent  $IV$  process, with the resulting estimates being significantly more persistent than the estimates for the directly observable  $RV$  process.

## 4. EXPLOITING THE ERRORS

---

and can easily be tailored to any autoregressive specification for  $RV$ . For concreteness, however, we focus our main discussion on the adaptation to the popular HAR model, which we dub the HARQ model. But, in our empirical investigation we also consider a number of other specifications and variations of the basic HARQ model.

Our empirical analysis relies on high-frequency data from 1997-2013 and corresponding realized volatility measures for the S&P 500 index and the individual constituents of Dow Jones Industrial Average. By explicitly incorporating the time-varying variance of the measurement errors into the parameterization of the model, the estimated HARQ models exhibit more persistence in “normal times” and quicker mean reversion in “erratic times” compared to the standard HAR model with constant autoregressive parameters.<sup>5</sup> Applying the HARQ model in an extensive out-of-sample forecast comparison, we document significant improvements in the accuracy of the forecasts compared to the forecasts from a challenging set of commonly used benchmark models. Interestingly, the forecasts from the HARQ models are not just improved in times when the right-hand side  $RV$ s are very noisy, and thus contain little relevant information, but also during tranquil times, when the forecasts benefit from the higher persistence afforded by the new models. Consistent with the basic intuition, the HARQ type models also offer the largest gains over the standard models for the assets for which the temporal variation in the magnitudes of the measurement errors are the highest.

The existing literature related to the dynamic modeling of  $RV$  and  $RV$ -based forecasting has largely ignored the issue of measurement errors, and when it has been considered, the errors have typically been treated as homoskedastic. Andersen, Bollerslev, and Meddahi (2011), for instance, advocate the use of ARMA models as a simple way to account for measurement errors, while Asai, McAleer, and Medeiros (2012) estimate a series of state-space models for the observable  $RV$  and the latent  $IV$  state variable with homoskedastic innovations. The approach for estimating stochastic volatility models based on realized volatility measures developed by Dobrev and Szerszen (2010) does incorporate the variance of the realized volatility error into the estimation of the models, but the parameters of the estimated models are assumed to be constant, and as such the dynamic dependencies and the forecasts from the models are not directly affected by the temporal variation in the size of the measurement errors. The motivation for the new family of HARQ models also bears some resemblance to the GMM estimation framework recently developed by Li and Xiu (2013). The idea of the paper is also related to the work of Bandi, Russell, and Yang (2013), who advocate the use of an “optimal,” and possibly time-varying, sampling frequency when implementing  $RV$  measures, as a way to account for heteroskedasticity in the market microstructure “noise.” In a similar vein, Shephard and Xiu (2014) interpret the magnitude of the parameter estimates associated with different  $RV$  measures in a GARCH-X model as indirect signals about the quality

---

<sup>5</sup>The persistence of the estimated HARQ models at average values for the measurement errors is very similar to the unconditional estimates based on Hansen and Lunde (2014a), and as such also much higher than the persistence of the standard HAR models. We discuss this further below.

of the different measures: the lower the parameter estimate, the less smoothing, and the more accurate and informative the specific  $RV$  measure.

The rest of the paper is structured as follows. Section 4.2 provides the theoretical motivation for the new class of models, together with the results from a small scale simulation study designed to illustrate the workings of the models. Section 4.3 reports the results from an empirical application of the basic HARQ model for forecasting the volatility of the S&P 500 index and the individual constituents of the Dow Jones Industrial Average. Section 4.4 provides a series of robustness checks and extensions of the basic HARQ model. Section 4.5 concludes.

## 4.2 Realized Volatility-Based Forecasting and Measurement Errors

### 4.2.1 Realized Variance High-Frequency Distribution Theory

To convey the main idea, consider a single asset for which the price process  $P_t$  is determined by the stochastic differential equation,

$$d \log(P_t) = \mu_t dt + \sigma_t dW_t, \quad (4.1)$$

where  $\mu_t$  and  $\sigma_t$  denote the drift and the instantaneous volatility processes, respectively, and  $W_t$  is a standard Brownian motion. For simplicity and ease of notation, we do not include jumps in this discussion, but the main idea readily extends to discontinuous price processes, and we investigate this in Section 4 below. Following the vast realized volatility literature, our aim is to forecast the latent Integrated Variance ( $IV$ ) over daily and longer horizons. Specifically, normalizing the unit time interval to a day, the one-day integrated variance is formally defined by,

$$IV_t = \int_{t-1}^t \sigma_s^2 ds. \quad (4.2)$$

The integrated variance is not directly observable. However, the Realized Variance ( $RV$ ) defined by the summation of high-frequency returns,

$$RV_t \equiv \sum_{i=1}^M r_{t,i}^2, \quad (4.3)$$

where  $M = 1/\Delta$ , and the  $\Delta$ -period intraday return is defined by  $r_{t,i} \equiv \log(P_{t-1+i\Delta}) - \log(P_{t-1+(i-1)\Delta})$ , provides a consistent estimator as the number of intraday observations increases, or equivalently  $\Delta \rightarrow 0$  (see, e.g., Andersen and Bollerslev, 1998).

In practice, data limitations invariably put an upper bound on the value of  $M$ . The

## 4. EXPLOITING THE ERRORS

---

resulting estimation error in  $RV$  may be characterized by the asymptotic (for  $\Delta \rightarrow 0$ ) distribution theory of Barndorff-Nielsen and Shephard (2002),

$$RV_t = IV_t + \eta_t, \quad \eta_t \sim MN(0, 2\Delta IQ_t), \quad (4.4)$$

where  $IQ_t \equiv \int_{t-1}^t \sigma_s^4 ds$  denotes the Integrated Quarticity ( $IQ$ ), and MN stands for mixed normal; i.e., a normal distribution conditional on the realization of  $IQ_t$ . In parallel, to the integrated variance, the integrated quarticity may be consistently estimated by the Realized Quarticity ( $RQ$ ),

$$RQ_t \equiv \frac{M}{3} \sum_{i=1}^M r_{t,i}^4. \quad (4.5)$$

### 4.2.2 The ARQ model

The consistency of  $RV$  for  $IV$ , coupled with the fact that the measurement error is serially uncorrelated under general conditions, motivate the use of reduced form time series models for the observable realized volatility as a simple way to forecast the latent integrated volatility of interest.<sup>6</sup>

To illustrate, suppose that the dynamic dependencies in  $IV$  may be described by an AR(1) model,

$$IV_t = \phi_0 + \phi_1 IV_{t-1} + u_t. \quad (4.6)$$

Under the strong assumption that  $u_t$  and the measurement error  $\eta_t$  are both *i.i.d.*, with variances  $\sigma_u^2$  and  $\sigma_\eta^2$ , then it follows by standard arguments that  $RV$  follows an ARMA(1,1) model, with AR-parameter equal to  $\phi_1$  and (invertible) MA-parameter equal to,

$$\theta_1 = \frac{-\sigma_u^2 - (1+\phi_1^2)\sigma_\eta^2 + \sqrt{\sigma_u^4 + 2(1+\phi_1^2)\sigma_u^2\sigma_\eta^2 + (1-\phi_1^2)^2\sigma_\eta^4}}{2\phi_1\sigma_\eta^2}, \quad (4.7)$$

for  $\sigma_\eta^2 > 0$  and  $\phi_1 \neq 0$ . It is possible to show that  $\theta$  is increasing (in absolute value) in the variance of the measurement error  $\sigma_\eta^2$ , and that  $\theta_1 \rightarrow 0$  as  $\sigma_\eta^2 \rightarrow 0$  or  $\phi_1 \rightarrow 0$ .

Instead, suppose that the researcher estimates a simple (and incorrect) AR(1) model for  $RV$ ,

$$IV_t + \eta_t = \beta_0 + \beta_1 (IV_{t-1} + \eta_{t-1}) + u_t. \quad (4.8)$$

The measurement error on the left hand side is unpredictable, and merely results in an increase in the standard errors of the parameter estimates. The measurement error on the right hand side, however, directly affects the parameter  $\beta_1$ , and in turn propagates into the forecasts from the model. If we continue to assume that  $u_t$  and  $\eta_t$  are both *i.i.d.*, so that  $Cov(RV_t, RV_{t-1}) = \phi_1 Var(IV_t)$  and  $Var(RV_t) = Var(IV_t) + 2\Delta IQ$ , (note IQ is

---

<sup>6</sup>A formal theoretical justification for this approach is provided by Andersen, Bollerslev, Diebold, and Labys (2003). Further, as shown by Andersen, Bollerslev, and Meddahi (2004), for some of the most popular stochastic volatility models used in the literature, simple autoregressive models for  $RV$  provide close to efficient forecasts for  $IV$ .

constant by the assumption that  $\eta_t$  is *i.i.d.*) then the population value for  $\beta_1$  may be expressed as,

$$\beta_1 = \phi_1 \left( 1 + \frac{2\Delta IQ}{Var(IV_t)} \right)^{-1}. \quad (4.9)$$

The estimated autoregressive coefficient for  $RV$  will therefore be smaller than the  $\phi_1$  coefficient for  $IV$ .<sup>7</sup> This discrepancy between  $\beta_1$  and  $\phi_1$  is directly attributable to the well-known attenuation bias arising from the presence of measurement errors. The degree to which  $\beta_1$  is attenuated is a direct function of the measurement error variance: if  $2\Delta IQ = 0$ , then  $\beta_1 = \phi_1$ , but if  $2\Delta IQ$  is large, then  $\beta_1$  goes to zero and  $RV$  is effectively unpredictable.<sup>8</sup>

The standard expression for  $\beta_1$  in equation (4.9) is based on the assumption that the variance of the measurement error is constant. However, the variance pertaining to the estimation error in  $RV$  generally changes through time: there are days when  $IQ$  is low and  $RV$  provides a strong signal about the true  $IV$ , and days when  $IQ$  is high and the signal is relatively weak. The OLS-based estimate of  $\beta_1$  will effectively be attenuated by the *average* of this measurement error variance. As such, the assumption of a constant AR parameter is suboptimal from a forecasting perspective. Instead, by explicitly allowing for a time-varying autoregressive parameter, say  $\beta_{1,t}$ , this parameter should be close to  $\phi_1$  on days when there is little measurement error, while on days where the measurement error variance is high,  $\beta_{1,t}$  should be low and the model quickly mean reverting.<sup>9</sup>

The AR(1) representation for the latent integrated volatility in equation (4.6) that underlies these ideas merely serves as an illustration, and even in that situation the simple expression for the impact of measurement errors in equation (4.9) requires homoskedasticity assumptions that are not likely to be satisfied in practice. Hence, rather than relying directly on the expression in equation (4.9), in our practical implementation we use a more flexible and robust specification in which we allow the time-varying AR parameter to depend linearly on an estimate of  $IQ^{1/2}$ . We term this specification the ARQ model

---

<sup>7</sup>The  $R^2$  from the regression based on  $RV$  will similarly be downward biased compared to the  $R^2$  from the infeasible regression based on the latent  $IV$ . Andersen, Bollerslev, and Meddahi (2005) provide a simple adjustment for this unconditional bias in the  $R^2$ .

<sup>8</sup>As previously noted, Hansen and Lunde (2014a) propose the use of an instrumental variable procedure for dealing with this attenuation bias and obtain a consistent estimator of the autoregressive parameters of the true latent  $IV$  process. For forecasting purposes, it is  $\beta_1$  and not  $\phi_1$  that is the parameter of interest, as the explanatory variable is the noisy realized variance, not the true integrated variance.

<sup>9</sup>These same arguments carry over to the  $\theta_1$  parameter for the ARMA(1,1) model in equation (4.7) and the implied persistence as a function of the measurement error variance. Correspondingly, in the GARCH(1,1) model, in the usual notation of that model, the autoregressive parameter given by  $\alpha + \beta$  should be constant, while the values of  $\alpha$  and  $\beta$  should change over time so that  $\beta$  is larger (smaller) and  $\alpha$  is smaller (larger) resulting in more (less) smoothing when the variance of the squared return is high (low).

for short,<sup>10</sup>

$$RV_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q}RQ_{t-1}^{1/2})}_{\beta_{1,t}}RV_{t-1} + u_t. \quad (4.10)$$

For ease of interpretation, we demean  $RQ^{1/2}$  so that the estimate of  $\beta_1$  corresponds to the average autoregressive coefficient, directly comparable to  $\beta_1$  in equation (4.8). The simple specification above has the advantage that it can easily be estimated by standard OLS, rendering both estimation and forecasting straightforward and fast. Importantly, the value of the autoregressive  $\beta_{1,t}$  parameter will vary with the estimated measurement error variance.<sup>11</sup> In particular, assuming that  $\beta_{1Q} < 0$  it follows that uninformative days with large measurement errors will have smaller impact on the forecasts than days where  $RV$  is estimated precisely and  $\beta_{1,t}$  is larger. If  $RQ$  is constant over time, the ARQ model reduces to a standard AR(1) model. Thus, the greater the temporal variation in the measurement error variance, the greater the expected benefit of modeling and forecasting the volatility with the ARQ model, a prediction we confirm in our empirical analysis below.

### 4.2.3 The HARQ model

The AR(1) model in equation (4.8) is too simplistic to satisfactorily describe the long-run dependencies in most realized volatility series. Instead, the Heterogeneous Autoregression (HAR) model of Corsi (2009) has arguably emerged as the most popular model for daily realized volatility based forecasting,

$$RV_t = \beta_0 + \beta_1RV_{t-1} + \beta_2RV_{t-1|t-5} + \beta_3RV_{t-1|t-22} + u_t, \quad (4.11)$$

where  $RV_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h RV_{t-i}$ , with  $j \leq h$ . The choice of a daily, weekly and monthly lag on the right-hand-side conveniently captures the approximate long-memory dynamic dependencies observed in most realized volatility series.

Of course, just like the simple AR(1) model discussed in the previous section, the beta coefficients in the HAR model are affected by measurement errors in the realized volatilities. In parallel to the ARQ model, this naturally suggests the following extension of the basic HAR model that directly adjust the coefficients in proportion to the magnitude

---

<sup>10</sup> $IQ$  is notoriously difficult to estimate in finite samples, and its inverse even more so. The use of the square-root as opposed to the inverse of  $RQ$  imbues the formulation with a degree of built-in robustness. However, we also consider a variety of other estimators and transformations of  $IQ$  in the robustness section below.

<sup>11</sup>In our additional empirical investigations reported in Section 4.4.3 below, we also consider models that allow the intercept  $\beta_0$  to vary with  $RQ$ .

of the corresponding measurement errors,

$$\begin{aligned}
 RV_t = & \beta_0 + \underbrace{(\beta_1 + \beta_{1Q}RQ_{t-1}^{1/2})}_{\beta_{1,t}}RV_{t-1} + \underbrace{(\beta_2 + \beta_{2Q}RQ_{t-1|t-5}^{1/2})}_{\beta_{2,t}}RV_{t-1|t-5} \\
 & + \underbrace{(\beta_3 + \beta_{3Q}RQ_{t-1|t-22}^{1/2})}_{\beta_{3,t}}RV_{t-1|t-22} + u_t,
 \end{aligned} \tag{4.12}$$

where  $RQ_{t-1|t-k} = \frac{1}{k} \sum_{j=1}^k RQ_{t-j}$ . Of course, the magnitude of the (normalized) measurement errors in the realized volatilities will generally decrease with the horizon  $k$  as the errors are averaged out, indirectly suggesting that adjusting for the measurement errors in the daily lagged realized volatilities is likely to prove more important than the adjustments for the weekly and monthly coefficients. Intuitively, this also means that in the estimation of the standard HAR model some of the weight will be shifted away from the noisy daily lag to the “cleaner,” though older, weekly and monthly lags that are less prone to measurement errors.

To directly illustrate how the measurement errors manifest over different sampling frequencies and horizons, Figure 4.1 plots the simulated  $RV$  measurement errors based on ten, five, and one-“minute” sampling ( $M = 39, 78, 390$ ) and horizons ranging from “daily,” to “weekly,” to “monthly” ( $k = 1, 5, 22$ ); the exact setup of the simulations are discussed in more detail in Section 4.2.4 and Appendix A. To facilitate comparison across the different values of  $M$  and  $k$ , we plot the distribution of  $RV/IV - 1$ , so that a value of 0.5 may be interpreted as an estimate that is 50% higher than the true  $IV$ . Even with an observation every minute ( $M = 390$ ), the estimation error in the daily ( $k = 1$ ) simulated  $RV$  can still be quite substantial. The measurement error variance for the weekly and monthly (normalized)  $RV$  are, as expected, much smaller and approximately  $1/5$  and  $1/22$  that of the daily  $RV$ . Thus, the attenuation bias in the standard HAR model will be much less severe for the weekly and monthly coefficients.

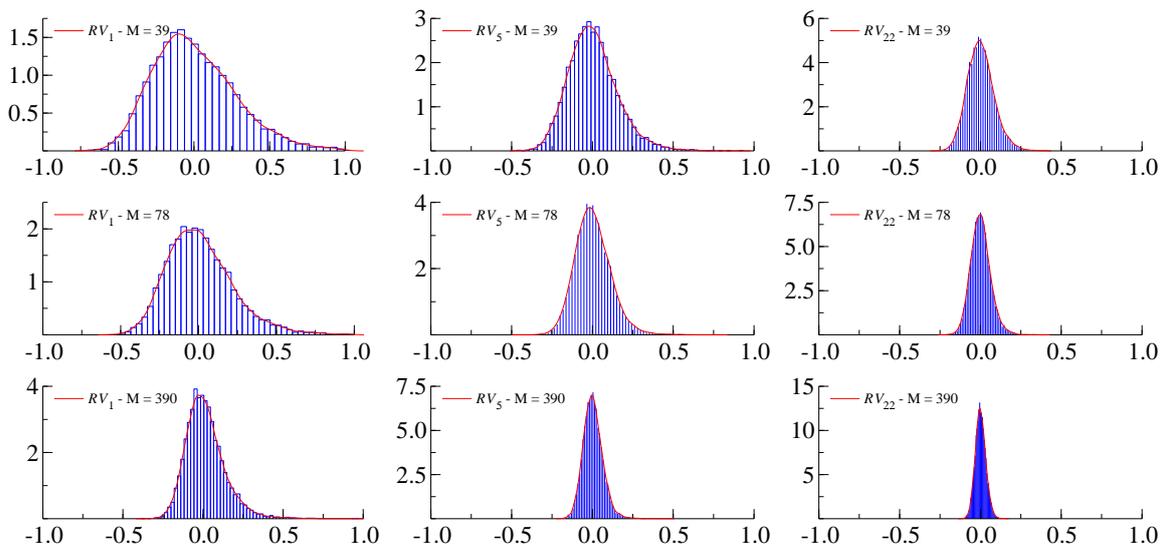
Motivated by these observations, coupled with the difficulties in precisely estimating the  $\beta_Q$  adjustment parameters, we will focus our main empirical investigations on the simplified version of the model in equation (4.12) that only allows the coefficient on the daily lagged  $RV$  to vary as a function of  $RQ^{1/2}$ ,

$$RV_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q}RQ_{t-1}^{1/2})}_{\beta_{1,t}}RV_{t-1} + \beta_2RV_{t-1|t-5} + \beta_3RV_{t-1|t-22} + u_t. \tag{4.13}$$

We will refer to this model as the HARQ model for short, and the model in equation (4.12) that allows all of the parameters to vary with an estimate of the measurement error variance as the “full HARQ” model, or HARQ-F.

To illustrate the intuition and inner workings of the HARQ model, Figure 4.2 plots the HAR and HARQ model estimates for the S&P 500 for ten consecutive trading days in October 2008; further details concerning the data are provided in the empirical section

Figure 4.1: Estimation Error of RV

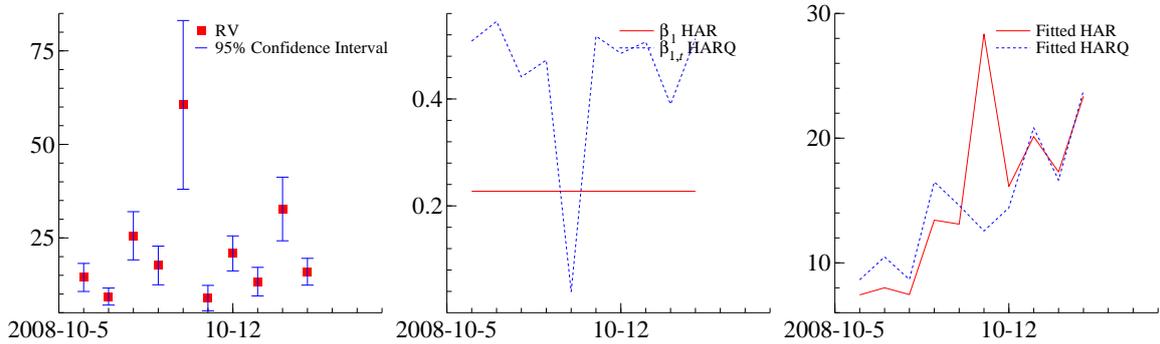


*Note:* The figure shows the simulated distribution of  $RV/IV - 1$ . The top, middle and bottom panels show the results for  $M = 39, 78$ , and  $390$ , respectively, while the left, middle and right panels show the results for daily, weekly, and monthly forecast horizons, respectively.

below. The left panel shows the estimated  $RV$  along with 95% confidence bands based in the asymptotic approximation in (4.4). One day in particular stands out: on Friday, October 10 the realized volatility was substantially higher than for any of the other ten days, and importantly, also far less precisely estimated, as evidenced by the wider confidence bands.<sup>12</sup> The middle panel shows the resulting  $\beta_1$  and  $\beta_{1,t}$  parameter estimates. The level of  $\beta_{1,t}$  from the HARQ model is around 0.5 on “normal” days, more than double that of  $\beta_1$  of just slightly above 0.2 from the standard HAR model. However, on the days when  $RV$  is estimated imprecisely,  $\beta_{1,t}$  can be much lower, as illustrated by the precipitous drop to less than 0.1 on October 10, as well as the smaller drop on October 16. The right-most panel shows the effect that this temporal variation in  $\beta_{1,t}$  has on the one-day-ahead forecasts from the HARQ model. In contrast to the HAR model, where the high  $RV$  on October 10 leads to an increase in the fitted value for the next day, the HARQ model actually forecasts a *lower* value than the day before. Compared to the standard HAR model, the HARQ model allows for higher average persistence, together with forecasts closer to the unconditional volatility when the lagged  $RV$  is less informative.

<sup>12</sup>October 10 was marked by a steep loss in the first few minutes of trading followed by a rise into positive territory and a subsequent decline, with all of the major indexes closing down just slightly for the day, including the S&P 500 which fell by 1.2%.

Figure 4.2: HAR vs. HARQ



*Note:* The figure illustrates the HARQ model for ten successive trading days. The left-panel shows the estimated  $RV$ s with 95% confidence bands based on the estimated  $RQ$ . The middle panel shows the  $\beta_{1,t}$  estimates from the HARQ model, together with the estimate of  $\beta_1$  from the standard HAR model. The right panel shows the resulting one-day-ahead  $RV$  forecasts from the HAR and HARQ models.

#### 4.2.4 Simulation Study

To further illustrate the workings of the HARQ model, this section presents the results from a small simulation study. We begin by demonstrating non-trivial improvements in the in-sample fits from the ARQ and HARQ models compared to the standard AR and HAR models. We then show how these improved in-sample fits translates into superior out-of-sample forecasts. Finally, we demonstrate how these improvements may be attributed to the increased average persistence of the estimated ARQ and HARQ models obtained by shifting the weights of the lags to more recent observations.

Our simulations are based on the two-factor stochastic volatility diffusion with noise previously analyzed by Huang and Tauchen (2005), Gonçalves and Meddahi (2009) and Patton (2011a), among others. Details about the exact specification of the model and the parameter values used in the simulation are given in Appendix A. We report the results based on  $M = 39, 78, 390$  “intraday” return observations, corresponding to ten, five, and one-“minute” sampling frequencies. We consider five different forecasting models: AR, HAR, ARQ, HARQ and HARQ-F. The AR and HAR models help gauge the magnitude of the improvements that may realistically be expected in practice. All of the models are estimated by OLS based on  $T = 1,000$  simulated “daily” observations. Consistent with the OLS estimation of the models, we rely on a standard MSE measure to assess the in-sample fits,

$$MSE(RV_t, X_t) \equiv (RV_t - X_t)^2,$$

where  $X_t$  refers to the fit from any one of the different models. We also calculate one-day-ahead out-of-sample forecasts from all of the models. For the out-of-sample comparisons

## 4. EXPLOITING THE ERRORS

**Table 4.1:** Simulation Results

	AR	HAR	ARQ	HARQ	HARQ-F
M	In-Sample MSE				
39	1.0291	1.0000	0.9980	0.9773	0.9718
78	1.0285	1.0000	0.9996	0.9791	0.9735
390	1.0277	1.0000	1.0064	0.9851	0.9793
	Out-of-Sample MSE				
39	1.0438	1.0000	1.0166	0.9878	0.9900
78	1.0425	1.0000	1.0188	0.9901	0.9920
390	1.0413	1.0000	1.0268	0.9968	0.9985
	Out-of-Sample QLIKE				
39	1.0893	1.0000	1.0258	0.9680	0.9850
78	1.0881	1.0000	1.0186	0.9644	0.9821
390	1.0841	1.0000	1.0187	0.9678	0.9859
	Persistence				
39	0.4303	0.6593	0.6552	0.8132	0.9200
78	0.4568	0.6736	0.6876	0.8328	0.9449
390	0.4739	0.6803	0.6913	0.8297	0.9621
	Mean Lag				
39		5.6598		4.2410	4.6956
78		5.4963		4.1026	4.5968
390		5.3685		4.1196	4.6530

*Note:* The table reports the MSE and QLIKE losses for the different approximate models. The average losses are standardized by the loss of the HAR model. The bottom panel reports the average estimated persistence and mean lag across the different models. The two-factor stochastic volatility model and the exact design underlying the simulations are further described in Appendix A.

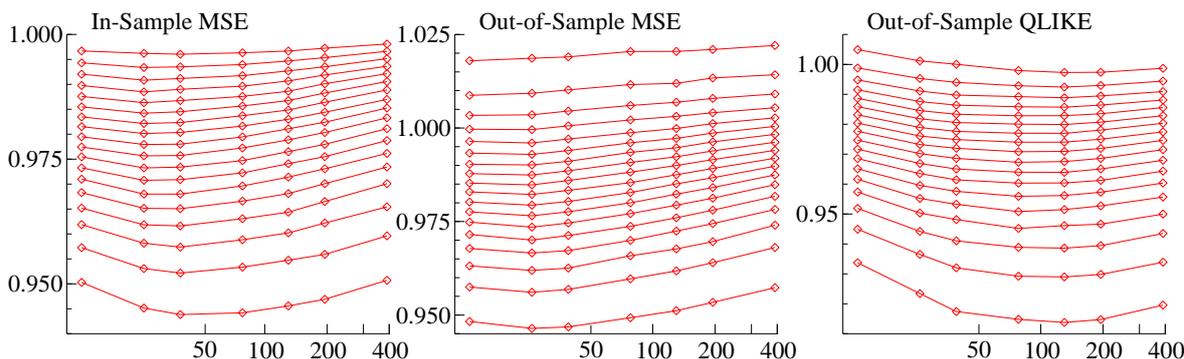
we consider both the  $MSE(RV_t, F_t)$ , and the QLIKE loss,

$$QLIKE(RV_t, F_t) \equiv \frac{RV_t}{F_t} - \log\left(\frac{RV_t}{F_t}\right) - 1,$$

where  $F_t$  refers to the one-day-ahead forecasts from the different models.<sup>13</sup> To facilitate direct comparisons of the in- and out-of-sample results, we rely on a rolling window of 1,000 observations for the one-step-ahead forecasts and use these same 1,000 forecasted observations for the in-sample estimation. All of the reported simulation results are based on 1,000 replications.

Table 4.1 summarizes the key findings. To make the relative gains stand out more clearly, we standardize the relevant loss measures in each of the separate panels by the loss of the HAR model. As expected, the ARQ model systematically improves on the AR

<sup>13</sup>Very similar out-of-sample results and rankings of the different models are obtained for the MSE and QLIKE defined relative to the true latent integrated volatility within the simulations; i.e.,  $MSE(IV_t, F_t)$  and  $QLIKE(IV_t, F_t)$ , respectively.

**Figure 4.3:** Distribution of HARQ/HAR ratio

*Note:* The figures depicts the quantiles ranging from 0.05 to 0.95 in increments of 0.05 for the simulated MSE and QLIKE loss ratios for the HARQ model relative to the standard HAR model. The horizontal axis shows the number of observations used to estimate  $RV$ , ranging from 13 to 390 per “day.”

model, and the HARQ model similarly improves on the HAR model. This holds true both in- and out-of-sample. The difficulties in accurately estimating the additional adjustment parameters for the weekly and monthly lags in the HARQ-F model manifest in this model generally not performing as well out-of-sample as the simpler HARQ model that only adjusts for the measurement error in the daily realized volatility. Also, the improvements afforded by the (H)ARQ models are decreasing in  $M$ , as more high-frequency observations help reduce the magnitude of the measurement errors, and thus reduce the gains from exploiting them.<sup>14</sup>

Figure 4.3 further highlights this point. The figure plots the simulated quantiles of the ratio distribution of the HARQ to HAR models for different values of  $M$ . Each line represents one quantile, ranging from 5% to 95% in 5% increments. For all criteria, in- and out-of-sample MSE and out-of-sample QLIKE, the loss ratio shows a U-shaped pattern, with the gains of the HARQ model relative to the standard HAR model maximized somewhere between 2- and 10-“minute” sampling. When  $M$  is very large, the measurement error decreases and the gains from using information on the magnitude of the error diminishes. When  $M$  is very small, estimating the measurement error variance by  $RQ$  becomes increasingly more difficult and the adjustments in turn less accurate. As such, the adjustments that motivate the HARQ model are likely to work best in environments where there is non-negligible measurement error in  $RV$ , and the estimation of this measurement error via  $RQ$  is at least somewhat reliable. Whether this holds in practice is an

<sup>14</sup>In order to investigate whether the apparent gains from the “Q” models might be attributable to their ability to capture some nonlinearity in the DGP for the daily  $IV$  that is missed by the linear AR and HAR specifications, we also considered simulations in which we predict the future  $IV$ s with HAR models based on the past  $IV$ s and HARQ models based on the past  $IV$ s and  $IQ$ s. The results show that parts of the gains are indeed attributable to capturing nonlinearity, but that most of the gains observed in Table 4.1 stem from the RV-based “Q” models’ ability to capture time-varying measurement errors.

empirical question, and one that we study in great detail in the next section.

The second half of Table 4.1 reports the persistence for all of the models defined by the estimates of  $\beta_1 + \beta_2 + \beta_3$ , as well as the mean lags for the HAR, HARQ and HARQ-F models. In the HAR models, the weight on the first lag equals  $b_1 = \beta_1 + \beta_2/5 + \beta_3/22$ , on the second lag  $b_2 = \beta_2/5 + \beta_3/22$ , on the sixth lag  $b_6 = \beta_3/22$ , and so forth, so that these mean lags are easily computed as  $22 \sum_{i=1}^{22} ib_i / \sum_{i=1}^{22} b_i$ . For the HARQ models, this corresponds to the mean lag at the *average* measurement error variance. The mean lag gives an indication of the location of the lag weights. The lower the mean lag, the greater the weight on more recent *RVs*.

The results confirm that at the mean measurement error variance, the HARQ model is far more persistent than the standard HAR model. As  $M$  increases, and the measurement error decreases, the gap between the models narrows. However, the persistence of the HARQ model is systematically higher, and importantly, much more stable across the different values of  $M$ . As  $M$  increases and the measurement error decreases, the loading on *RQ* diminishes, but this changes little in terms of the persistence of the underlying latent process that is being approximated by the HARQ model.<sup>15</sup> The result pertaining to the mean lags reported in the bottom panel further corroborates the idea that on average, the HARQ model assigns more weight to more recent *RVs* than the does the standard HAR model.

### 4.3 Modeling and Forecasting Volatility

#### 4.3.1 Data

We focus our empirical investigations on the S&P 500 aggregate market index. High-frequency futures prices for the index are obtained from Tick Data Inc. We complement our analysis of the aggregate market with additional results for the 27 Dow Jones Constituents as of September 20, 2013 that traded continuously from the start to the end of our sample. Data on these individual stocks comes from the TAQ database. Our sample starts on April 21, 1997, one thousand trading days (the length of our estimation window) before the final decimalization of NASDAQ on April 9, 2001. The sample for the S&P 500 ends on August 30, 2013, while the sample for the individual stocks ends on December 31, 2013, yielding a total of 3,096 observations for the S&P 500 and 3,202 observations for the DJIA constituents. The first 1,000 days are only used to estimate the models, so that the in-sample estimation results and the rolling out-of-sample forecasts are all based on the same samples.

Table 4.2 provides a standard set of summary statistics for the daily realized volatilities. Following common practice in the literature, all of the *RVs* are based on five-minute

---

<sup>15</sup>Interestingly, the HARQ-F model is even more persistent. This may be fully attributed to an increase in the monthly lag parameter, combined with a relatively high loading on the interaction of the monthly *RV* and *RQ*.

**Table 4.2:** Summary Statistics

Company	Symbol	Min	Mean	Median	Max	AR	AR-HL	ARQ
S&P 500		0.043	1.175	0.629	60.563	0.651	0.953	0.983
Microsoft	MSFT	0.166	3.087	1.824	59.164	0.718	0.952	0.889
Coca-Cola	KO	0.049	2.011	1.154	54.883	0.618	0.949	0.834
DuPont	DD	0.093	3.327	2.165	81.721	0.707	0.950	0.956
ExxonMobil	XOM	0.114	2.348	1.476	130.667	0.668	0.947	0.997
General Electric	GE	0.131	3.440	1.794	173.223	0.681	0.915	0.987
IBM	IBM	0.115	2.464	1.340	72.789	0.657	0.959	0.890
Chevron	CVX	0.105	2.286	1.483	139.984	0.653	0.966	0.954
United Technologies	UTX	0.126	2.793	1.658	92.105	0.648	0.943	0.883
Procter & Gamble	PG	0.085	2.007	1.064	80.124	0.587	0.866	0.786
Caterpillar	CAT	0.207	3.810	2.401	127.119	0.727	0.954	0.896
Boeing	BA	0.167	3.371	2.147	79.760	0.630	0.936	0.822
Pfizer	PFE	0.176	2.822	1.809	60.302	0.570	0.933	0.837
Johnson & Johnson	JNJ	0.062	1.680	0.999	58.338	0.613	0.946	0.933
3M	MMM	0.140	2.278	1.358	123.197	0.495	0.952	0.748
Merck	MRK	0.127	2.758	1.718	223.723	0.372	0.913	0.708
Walt Disney	DIS	0.135	3.641	2.030	129.661	0.629	0.916	0.772
McDonald's	MCD	0.090	2.678	1.680	130.103	0.390	0.937	0.672
JPMorgan Chase	JPM	0.114	5.420	2.552	261.459	0.716	0.832	0.940
Wal-Mart	WMT	0.148	2.761	1.443	114.639	0.611	0.925	0.810
Nike	NKE	0.192	3.431	1.980	84.338	0.581	0.943	0.785
American Express	AXP	0.088	4.603	2.184	290.338	0.602	0.948	0.949
Intel	INTC	0.208	4.654	2.674	89.735	0.731	0.949	0.968
Travelers	TRV	0.098	3.579	1.637	273.579	0.646	0.933	0.915
Verizon	VZ	0.145	2.788	1.637	99.821	0.646	0.952	0.859
The Home Depot	HD	0.171	3.798	2.161	133.855	0.633	0.946	0.992
Cisco Systems	CSCO	0.234	5.120	2.742	96.212	0.715	0.939	0.942
UnitedHealth Group	UNH	0.222	4.145	2.304	169.815	0.616	0.920	0.846

*Note:* The table provides summary statistics for the daily  $RV$ s for each of the series. The column labeled AR reports the standard first order autocorrelation coefficients, the column labeled AR-HL gives the instrumental variable estimator of Hansen and Lunde (2014a), while  $\beta_1$  refers to the corresponding estimates from the ARQ model in equation (4.10)

returns.<sup>16</sup> In addition to the usual summary measures, we also report the first order autocorrelation corresponding to  $\beta_1$  in equation (4.8), the instrumental variable estimator of Hansen and Lunde (2014a) denoted AR-HL, and the estimate of  $\beta_1$  from the ARQ model in equation (4.10) corresponding to the autoregressive parameter at the average measurement error variance. The AR-HL estimates are all much larger than the standard AR estimates, directly highlighting the importance of measurement errors. By exploiting the heteroskedasticity in the measurement errors, the ARQ model allows for far greater persistence on average than the standard AR model, bridging most of the gap between the AR and AR-HL estimates.

<sup>16</sup>Liu, Patton, and Sheppard (2015) provide a recent discussion and empirical justification for this common choice. In some of the additional results discussed below, we also consider other sampling frequencies and  $RV$  estimators. Our main empirical findings remain intact to these other choices.

## 4. EXPLOITING THE ERRORS

**Table 4.3:** In-Sample Estimation Results

	AR	HAR	AR-HL	ARQ	HARQ	HARQ-F
$\beta_0$	0.4109 (0.1045)	0.1123 (0.0615)		0.0892 (0.0666)	-0.0098 (0.0617)	-0.0187 (0.0573)
$\beta_1$	0.6508 (0.1018)	0.2273 (0.1104)	0.9529 (0.0073)	0.9830 (0.0782)	0.6021 (0.0851)	0.5725 (0.0775)
$\beta_2$		0.4903 (0.1352)			0.3586 (0.1284)	0.4368 (0.1755)
$\beta_3$		0.1864 (0.1100)			0.0976 (0.1052)	0.0509 (0.1447)
$\beta_{1Q}$				-0.5139 (0.0708)	-0.3602 (0.0637)	-0.3390 (0.0730)
$\beta_{2Q}$						-0.1406 (0.3301)
$\beta_{3Q}$						0.0856 (0.3416)
$R^2$	0.4235	0.5224	0.3323	0.5263	0.5624	0.5628
MSE	3.1049	2.5722	3.5964	2.5512	2.3570	2.3546
QLIKE	0.2111	0.1438	0.1586	0.1530	0.1358	0.1380
$\bar{R}^2$ Stocks	0.3975	0.4852	0.2935	0.4676	0.5090	0.5139
$\overline{MSE}$ Stocks	17.4559	14.9845	20.0886	15.2782	14.1702	14.0154
$\overline{QLIKE}$ Stocks	0.2095	0.1496	0.1759	0.1804	0.1470	0.1547

*Note:* The table provides in-sample parameter estimates and measures of fit for the various benchmark and (H)ARQ models. The top two panels report the actual parameter estimates for the S&P500 with robust standard errors in parentheses, together with the  $R^2$ s, MSE and QLIKE losses from the regressions. The bottom panel summarizes the in-sample losses for the different models averaged across all of the individual stocks.

### 4.3.2 In-Sample Estimation Results

We begin by considering the full in-sample results. The top panel in Table 4.3 reports the parameter estimates for the S&P 500, with robust standard errors in parentheses, for the benchmark AR and HAR models, together with the ARQ, HARQ and HARQ-F models. For comparison purposes, we also include the AR-HL estimates, even though they were never intended to be used for forecasting purposes. The second and third panel report the  $R^2$ , MSE and QLIKE for the S&P500, and the average of those three statistics across the 27 DJIA individual stocks. Further details about the model parameter estimates for the individual stocks are available in Appendix A.

As expected, all of the  $\beta_{1Q}$  coefficients are negative and strongly statistically significant. This is consistent with the simple intuition that as the measurement error and the current value of  $RQ$  increases, the informativeness of the current  $RV$  for future  $RV$ s decreases, and therefore the  $\beta_{1,t}$  coefficient on the current  $RV$  decreases towards zero. Directly comparing the AR coefficient to the autoregressive parameter in the ARQ model also reveals a marked difference in the estimated persistence of the models. By failing to take into account the time-varying nature of the informativeness of the  $RV$  measures, the estimated AR coefficients are sharply attenuated.

The findings for the HARQ model are slightly more subtle. Comparing the HAR model with the HARQ model, the HAR places greater weight on the weekly and monthly lags, which are less prone to measurement errors than the daily lag, but also further in the past. These increased weights on the weekly and monthly lags hold true for the S&P500 index, and for every single individual stock in the sample. By taking into account the time-varying nature of the measurement error in the daily  $RV$ , the HARQ model assigns a greater average weight to the daily lag, while down-weighting the daily lag when the measurement error is large. The HARQ-F model parameters differ slightly from the HARQ model parameters, as the weekly and monthly lags are now also allowed to vary. However, the estimates for  $\beta_{2Q}$  and  $\beta_{3Q}$  are not statistically significant, and the improvement in the in-sample fit compared to the HARQ model is minimal.

To further corroborate the conjecture that the superior performance of the HARQ model is directly attributable to the measurement error adjustments, we also calculated the mean lags implied by the HAR and HARQ models estimated with less accurate realized volatilities based on coarser sampled 10- and 15-minute intraday returns. Consistent with the basic intuition of the measurement errors on average pushing the weights further in the past, the mean lags are systematically lower for the models that rely on the more finely sampled  $RV$ s. For instance, the average mean lag across all of the individual stocks for the HAR models equal 5.364, 5.262 and 5.003 for 15-, 10- and 5-minute  $RV$ s, respectively. As the measurement error decreases, the shorter lags become more accurate and informative for the predictions. By comparison, the average mean lag across all of the stocks for the HARQ models equal 4.063, 3.877 and 3.543 for 15-, 10- and 5-minute  $RV$ s, respectively. Thus, on average the HARQ models always assign more weight to the more recent  $RV$ s than the standard HAR models, and generally allow for a more rapid response, except, of course, when the signal is poor.

### 4.3.3 Out-of-Sample Forecast Results

Many other extensions of the standard HAR model have been proposed in the literature. To help assess the forecasting performance of the HARQ model more broadly, in addition to the basic AR and HAR models considered above, we therefore also consider the forecasts from three alternative popular HAR type formulations.

Specifically, following Andersen, Bollerslev, and Diebold (2007) we include both the the HAR-with-Jumps (HAR-J) and the Continuous-HAR (CHAR) models in our forecast comparisons. Both of these models rely on a decomposition of the total variation into a continuous and a discontinuous (jump) part. This decomposition is most commonly implemented using the Bi-Power Variation ( $BPV$ ) measure of Barndorff-Nielsen and Shephard (2004c), which affords a consistent estimate of the continuous variation in the presence of jumps. The HAR-J model, in particular, includes a measure of the jump

#### 4. EXPLOITING THE ERRORS

---

variation as an additional explanatory variable in the standard HAR model,

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + u_t, \quad (4.14)$$

where  $J_t \equiv \max[RV_t - BPV_t, 0]$ , and the  $BPV$  measure is defined as,

$$BPV_t \equiv \mu_1^{-2} \sum_{i=1}^{M-1} |r_{t,i}| |r_{t,i+1}|, \quad (4.15)$$

with  $\mu_1 = \sqrt{2/\pi} = \mathbb{E}(|Z|)$ , and  $Z$  a standard normally distributed random variable. Empirically, the jump component has typically been found to be largely unpredictable. This motivates the alternative CHAR model, which only includes measures of the continuous variation on the right hand side,

$$RV_t = \beta_0 + \beta_1 BPV_{t-1} + \beta_2 BPV_{t-1|t-5} + \beta_3 BPV_{t-1|t-22} + u_t. \quad (4.16)$$

Several empirical studies have documented that the HAR-J and CHAR models often perform (slightly) better than the standard HAR model.

Meanwhile, Patton and Sheppard (2015) have recently argued that a Semivariance-HAR (SHAR) model sometimes performs even better than the HAR-J and CHAR models. Building on the semi-variation measures of Barndorff-Nielsen, Kinnebrock, and Shephard (2010), the SHAR model decomposes the total variation in the standard HAR model into the variation due to negative and positive intraday returns, respectively. In particular, let  $RV_t^- \equiv \sum_{i=1}^M r_{t,i}^2 \mathbb{I}_{\{r_{t,i} < 0\}}$  and  $RV_t^+ \equiv \sum_{i=1}^M r_{t,i}^2 \mathbb{I}_{\{r_{t,i} > 0\}}$ , the SHAR model is then defined as:

$$RV_t = \beta_0 + \beta_1^+ RV_{t-1}^+ + \beta_1^- RV_{t-1}^- + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t. \quad (4.17)$$

Like the HARQ models, the HAR-J, CHAR and SHAR models are all easy to estimate and implement.

We focus our discussion on the one-day-ahead forecasts for the S&P500 index starting on April 9, 2001 through the end of the sample. However, we also present summary results for the 27 individual stocks, with additional details available in Appendix A. The forecast are based on re-estimating the parameters of the different models each day with a fixed length Rolling Window ( $RW$ ) comprised of the previous 1,000 days, as well as an Increasing Window ( $IW$ ) using all of the available observations. The sample sizes for the increasing window for the S&P500 thus range from 1,000 to 3,201 days.

The average MSE and QLIKE for the S&P500 index are reported in the top panel in

Table 4.4: Out-of-Sample Forecast Losses

		AR	HAR	HAR- J	CHAR	SHAR	ARQ	HARQ	HARQ- F	
S&P 500										
MSE	RW	0.9166	1.0000	0.9176	0.9583	0.8375	<b>0.8115</b>	0.8266	0.9750	
	IW	1.2315	1.0000	0.9676	0.9707	0.9012	0.9587	<b>0.8944</b>	0.9312	
QLIKE	RW	1.4559	1.0000	1.0062	1.0124	<b>0.9375</b>	0.9570	0.9464	0.9934	
	IW	1.7216	1.0000	0.9716	0.9829	0.8718	1.1845	0.8809	<b>0.8686</b>	
Individual Stocks										
MSE	RW	Avg	1.1505	1.0000	1.0151	1.0080	1.0083	0.9659	<b>0.9349</b>	1.0149
		Med	1.1730	1.0000	1.0115	1.0158	1.0020	0.9864	<b>0.9418</b>	1.0263
	IW	Avg	1.2130	1.0000	1.0040	1.0013	0.9947	1.0371	<b>0.9525</b>	1.0071
		Med	1.2161	1.0000	1.0028	1.0010	0.9968	1.0396	<b>0.9525</b>	0.9660
QLIKE	RW	Avg	1.4204	1.0000	1.0018	0.9999	0.9902	1.1498	<b>0.9902</b>	1.1516
		Med	1.4044	1.0000	0.9976	1.0025	0.9941	1.1781	<b>0.9916</b>	1.1051
	IW	Avg	1.5803	1.0000	0.9930	1.0148	0.9829	1.2024	<b>0.9487</b>	0.9843
		Med	1.5565	1.0000	0.9959	1.0163	0.9887	1.1732	<b>0.9550</b>	0.9630

*Note:* The table reports the ratio of the losses for the different models relative to the losses of the HAR model. The top panel shows the results for the S&P500. The bottom panel reports the average and median loss ratios across all of the individual stocks. The lowest ratio in each row is highlighted in bold.

Table 4.4, with the results for the individual stocks summarized in the bottom panel.<sup>17,18</sup> The results for S&P500 index are somewhat mixed, with each of the three “Q” models performing the best for one of the loss functions/window lengths combinations, and the remaining case being won by the SHAR model. The lower panel pertaining to the individual stocks reveals a much cleaner picture: across both loss functions and both window lengths, the HARQ model systematically exhibits the lowest average and median loss. The HARQ-F model fails to improve on the HAR model, again reflecting the difficulties in accurately estimating the weekly and monthly adjustment parameters. Interestingly, and in contrast to the results for the S&P 500, the CHAR, HAR-J and SHAR models generally perform only around as well as the standard HAR model for the individual stocks.

In order to formally test whether the HARQ model significantly outperforms all of the other models, we use a modification of the Reality Check ( $RC$ ) of White (2000). The standard  $RC$  test determines whether the loss from the best model from a set of competitor models is significantly lower than a given benchmark. Instead, we want to

<sup>17</sup>Due to the estimation errors in  $RQ$ , the HARQ models may on are occasions produce implausibly large or small forecasts. Thus, to make our forecast analysis more realistic, we apply an “insanity filter” to the forecasts; see, e.g., Swanson and White (1997). If a forecast is outside the range of values of the target variable observed in the estimation period, the forecast is replaced by the unconditional mean over that period: “insanity” is replaced by “ignorance.” This same filter is applied to all of the benchmark models. In practice this trims fewer than 0.1% of the forecasts for any of the series, and none for many.

<sup>18</sup>Surprisingly, the rolling window forecasts provided by the AR model have *lower* average MSE than the HAR model. However, in that same setting the ARQ model also beats the HARQ model.

## 4. EXPLOITING THE ERRORS

---

test whether the loss of a given model (HARQ) is lower than that from the best model among a set of benchmark models. As such, we adjust the hypotheses accordingly, testing

$$H_0 : \min_{k=1, \dots, K} \mathbb{E}[L^k(RV, X) - L^0(RV, X)] \leq 0,$$

versus

$$H_A : \min_{k=1, \dots, K} \mathbb{E}[L^k(RV, X) - L^0(RV, X)] > 0,$$

where  $L^0$  denotes the loss of the HARQ model, and  $L^k$ ,  $k = 1, \dots, K$  refers to the loss from all of the other benchmark models. A rejection of the null therefore implies that the loss of the HARQ model is significantly lower than all benchmark models. As suggested by White (2000), we implement the Reality Check using the stationary bootstrap of Politis and Romano (1994) with 999 re-samplings and an average block length of five. (The results are not sensitive to this choice of block-length.)

For the S&P500 index, the null hypothesis is rejected at the 10% level for the MSE loss with a p-value of 0.063, but not for QLIKE where the p-value equals 0.871. For the individual stocks, we reject the null in favor of the HARQ model under the MSE loss for 44% (63%) of stocks at the 5% (10%) significance level, respectively, and for 30% (37%) of the stocks under QLIKE loss. On the other hand, none of the benchmark models significantly outperforms the other models for more than one of the stocks. We thus conclude that for a large fraction of the stocks, the HARQ model significantly beats a challenging set of benchmark models commonly used in the literature.

### 4.3.4 Dissecting the Superior Performance of the HARQ Model

Our argument as to why the HARQ model improves on the familiar HAR model hinges on the model's ability to place a larger weight on the lagged daily  $RV$  on days when  $RV$  is measured relatively accurately ( $RQ$  is low), and to reduce the weight on days when  $RV$  is measured relatively poorly ( $RQ$  is high). At the same time,  $RV$  is generally harder to measure when it is high, making  $RV$  and  $RQ$  positively correlated. Moreover, days when  $RV$  is high often coincide with days that contain jumps. Thus, to help alleviate concerns that the improvements afforded by the HARQ model are primarily attributable to jumps, we next provide evidence that the model works differently from any of the previously considered models that explicitly allow for distinct dynamic dependencies in the jumps. Consistent with the basic intuition underlying the model, we demonstrate that the HARQ model achieves the greatest forecast improvements in environments where the measurement error is highly heteroskedastic. In particular, in an effort to dissect the forecasting results in Table 4.4, Table 4.5 further breaks down the results in the previous table into forecasts for days when the previous day's  $RQ$  was very high (Top 5%  $RQ$ ) and the rest of the sample (Bottom 95%  $RQ$ ). As this breakdown shows, the superior performance of the HARQ model isn't merely driven by adjusting the coefficients when

Table 4.5: Stratified Out-of-Sample Forecast Losses

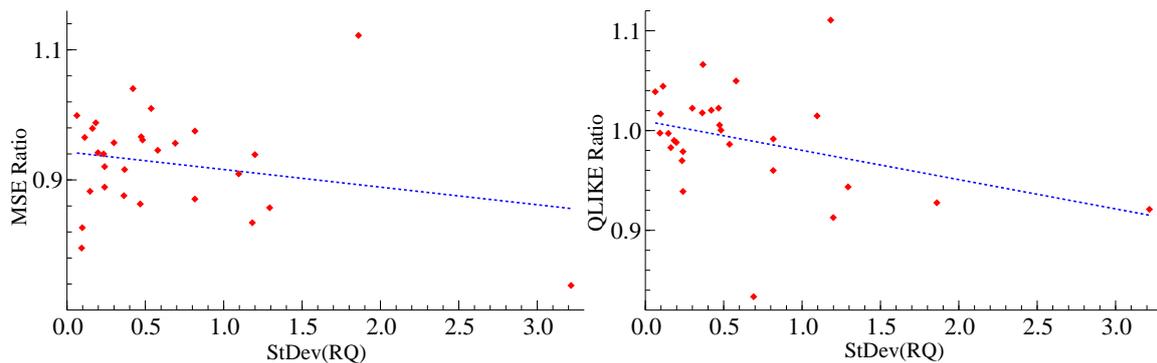
		AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ	HARQ-F	
Bottom 95% $RQ_t$										
S&P 500										
MSE	RW	1.0745	1.0000	0.9874	0.9629	0.9038	0.9202	<b>0.8997</b>	0.9419	
	IW	1.1566	1.0000	0.9777	0.9745	0.9139	0.9697	<b>0.9051</b>	0.9116	
QLIKE	RW	1.5431	1.0000	1.0040	1.0252	<b>0.9298</b>	1.1146	1.0145	1.2032	
	IW	1.7730	1.0000	0.9718	0.9849	0.8620	1.2025	0.8730	<b>0.8575</b>	
Individual Stocks										
MSE	RW	Avg	1.2081	1.0000	0.9939	1.0044	0.9869	1.0324	<b>0.9557</b>	0.9928
		Med	1.1987	1.0000	0.9957	1.0034	0.9915	1.0459	<b>0.9693</b>	0.9802
	IW	Avg	1.2823	1.0000	0.9944	1.0053	0.9865	1.0814	<b>0.9633</b>	0.9581
		Med	1.2114	1.0000	0.9971	1.0067	0.9937	1.0675	<b>0.9697</b>	0.9684
QLIKE	RW	Avg	1.4379	1.0000	0.9962	1.0067	0.9888	1.1309	<b>0.9787</b>	1.1212
		Med	1.4399	1.0000	0.9958	1.0127	0.9934	1.1180	<b>0.9820</b>	1.0673
	IW	Avg	1.6204	1.0000	0.9979	1.0257	0.9787	1.1920	<b>0.9353</b>	0.9639
		Med	1.5941	1.0000	0.9995	1.0258	0.9833	1.1616	<b>0.9395</b>	0.9532
Top 5% $RQ_t$										
S&P 500										
MSE	RW	0.8992	1.0000	0.9099	0.9578	0.8302	0.7995	0.8186	<b>0.7789</b>	
	IW	1.2410	1.0000	0.9663	0.9702	0.8995	0.9573	<b>0.8930</b>	0.9337	
QLIKE	RW	1.4311	1.0000	1.0615	<b>0.9869</b>	1.0049	1.2250	1.0310	1.2755	
	IW	1.3650	1.0000	0.9703	0.9691	0.9397	1.0594	<b>0.9357</b>	0.9456	
Individual Stocks										
MSE	RW	Avg	1.1426	1.0000	1.0228	1.0112	1.0163	0.9461	<b>0.9268</b>	1.0222
		Med	1.1591	1.0000	1.0212	1.0175	1.0073	0.9614	<b>0.9217</b>	1.0383
	IW	Avg	1.1933	1.0000	1.0053	0.9981	0.9983	1.0243	<b>0.9476</b>	1.0124
		Med	1.1984	1.0000	1.0052	1.0007	0.9979	1.0417	<b>0.9479</b>	0.9677
QLIKE	RW	Avg	1.3380	1.0000	1.0308	<b>0.9633</b>	1.0056	1.3161	1.0916	1.3535
		Med	1.3408	1.0000	0.9998	<b>0.9377</b>	1.0097	1.3052	1.0846	1.3250
	IW	Avg	1.3112	1.0000	0.9564	<b>0.9350</b>	1.0130	1.2864	1.0464	1.1301
		Med	1.3049	1.0000	0.9636	<b>0.9371</b>	1.0111	1.2186	1.0180	1.0045

*Note:* The table reports the same loss ratios given in Table 4.4 stratified according to  $RQ$ . The bottom panel shows the ratios for days following a value of  $RQ$  in the top 5%. The top panel shows the results for the remaining 95% of the days. The lowest ratio in each rows is indicated in bold.

$RQ$  is high. On the contrary, most of the gains in the QLIKE loss for the individual stocks appear to come from “normal” days and the increased persistence afforded by the HARQ model on those days. The results for the HARQ-F model underscores the difficulties in accurately estimating all of the parameters for that model, with the poor performance mostly stemming from the high  $RQ$  forecast days. These results also demonstrate that the HARQ model is distinctly different from the benchmark models. The CHAR and HAR-J models primarily show improvements on “high”  $RQ$  days, whereas most of the SHAR model’s improvements occur for the quieter “normal” days.

If the measurement error variance is constant over time, the HARQ model reduces to

Figure 4.4: Individual Stock Loss Ratios



*Note:* The graph plots the rolling window forecast MSE and QLIKE loss ratios of the HARQ model to the HAR model against the standard deviation of  $RQ$  ( $StDev(RQ)$ ) for each of the individual stocks.

the standard HAR model. Correspondingly, it is natural to expect that the HARQ model offers the greatest improvements when the measurement error is highly heteroskedastic. The results in Figure 4.4 corroborates this idea. The figure plots the MSE and QLIKE loss ratio of the HARQ model relative to the HAR model for the rolling window (RW) forecasts against the standard deviation of  $RQ$  ( $StDev(RQ)$ ) for each of the 27 individual stocks. Although  $StDev(RQ)$  provides a very noisy proxy, there is obviously a negative relation between the improvements afforded by the HARQ model and the heteroskedasticity in the measurement error variance. This is true for both of the loss ratios, but especially so for the QLIKE loss.<sup>19</sup>

### 4.3.5 Longer Forecast Horizons

Our analysis up until now has focused on forecasting daily volatility. In this section, we extend the analysis to longer weekly and monthly horizons. Our forecast will be based on direct projection, in which we replace the daily  $RV$ s on the left-hand-side of the different models with the weekly and monthly  $RV$ s.

Our previous findings indicate that for the one-day forecasts, the daily lag is generally the most important and receives by far the largest weight in the estimated HARQ models. Anticipating our results, when forecasting the weekly  $RV$  the weekly lag increases in importance, and similarly when forecasting the monthly  $RV$  the monthly lag becomes relatively more important. As such, allowing only a time-varying parameter for the daily lag may be suboptimal for the longer-run forecasts. Hence, in addition to the HARQ and HARQ-F models previously analyzed, we also consider a model in which we only adjust

<sup>19</sup>These same negative relations between the average gains afforded by the HARQ model and the magnitude of the heteroskedasticity in the measurement error variance hold true for the increasing window (IW) based forecasts as well.

**Table 4.6:** In-Sample Weekly and Monthly Model Estimates

	$h = 5$				$h = 22$			
	HAR	HARQ	HARQ-F	HARQ-h	HAR	HARQ	HARQ-F	HARQ-h
$\beta_0$	0.1717 (0.0432)	0.0977 (0.0429)	0.0576 (0.0392)	0.0170 (0.0466)	0.3417 (0.0276)	0.2914 (0.0297)	0.2845 (0.0339)	0.2930 (0.0346)
$\beta_1$	0.1864 (0.0597)	0.4078 (0.0717)	0.3408 (0.0699)	0.1898 (0.0492)	0.1049 (0.0502)	0.2547 (0.0595)	0.2124 (0.0617)	0.1043 (0.0495)
$\beta_2$	0.3957 (0.0768)	0.3159 (0.0762)	0.5623 (0.1056)	0.6825 (0.0980)	0.3342 (0.0662)	0.2802 (0.0658)	0.4537 (0.0959)	0.3364 (0.0678)
$\beta_3$	0.2709 (0.0655)	0.2172 (0.0659)	0.0862 (0.0852)	0.1609 (0.0725)	0.2695 (0.0540)	0.2332 (0.0557)	0.1122 (0.0658)	0.3225 (0.0509)
$\beta_{1Q}$		-0.2182 (0.0420)	-0.1488 (0.0415)			-0.1476 (0.0278)	-0.1032 (0.0309)	
$\beta_{2Q}$			-0.4404 (0.1514)	-0.5648 (0.1246)			-0.3158 (0.1132)	
$\beta_{3Q}$			0.2173 (0.2508)				0.2458 (0.1519)	-0.1847 (0.1353)
Mean Lag								
S&P500	5.2626	4.0952	3.0520	3.9564	5.9369	4.9173	3.6799	6.3185
Stocks	6.2593	5.0054	4.6427	4.4950	7.1939	6.0099	5.7646	7.9598

*Note:* The top panel reports the in-sample parameter estimates for the S&P 500 for the standard HAR model and the various HARQ models for forecasting the weekly ( $h = 5$ ) and monthly ( $h = 22$ ) RVs. Newey and West (1987) robust standard errors allowing for serial correlation up to order 10 ( $h = 5$ ), and 44 ( $h = 22$ ), respectively, are reported in parentheses. The bottom panel reports the mean lag implied by the estimated S&P 500 models, as well as the mean lags averaged across the models estimates for each of the individual stocks.

the lag corresponding to the specific forecast horizon. We term this model the HARQ-h model. Specifically, for the weekly and monthly forecasts analyzed here,

$$RV_{t+4|t} = \beta_0 + \beta_1 RV_{t-1} + \underbrace{(\beta_2 + \beta_{2Q} RQ_{t-1|t-5}^{1/2})}_{\beta_{2,t}} RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t \quad (4.18)$$

and

$$RV_{t+21|t} = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + u_t + \underbrace{(\beta_3 + \beta_{3Q} RQ_{t-1|t-22}^{1/2})}_{\beta_{3,t}} RV_{t-1|t-22} + u_t, \quad (4.19)$$

respectively. Note that for the daily horizon, the HARQ and HARQ-h models coincide. Table 4.6 reports the in-sample parameter estimates for the S&P 500 index for each of the different specifications. The general pattern is very similar to that reported in Table 4.3. Compared to the standard HAR model, the HARQ model always shifts the weights to the shorter lags. Correspondingly, the HARQ-h model shifts most of the weight to the lag that is allowed to be time-varying. Meanwhile, the HARQ-F model increases the relative weight of the daily and weekly lags, while reducing the weight on the monthly lag, with

the estimates for  $\beta_{1,t}$  and  $\beta_{2,t}$  both statistically significant and negative. The mean lags reported in the bottom panel also shows, that aside from the monthly HARQ-22 model, all of the HARQ specifications on average assign more weight to the more immediate and shorter lags than do the standard HAR models. The estimated HARQ-F models have the shortest mean lags among all of the models.

Turning to the out-of-sample forecast results for the weekly and monthly horizons reported in Tables 4.7 and 4.8, respectively, the CHAR and HAR-J benchmark models now both struggle to beat the HAR model. The SHAR model, on the other hand, offers small improvements over the HAR model in almost all cases. However, the simple version of the HARQ model substantially outperforms the standard HAR model in all of the different scenarios, except for the monthly rolling window MSE loss. The alternative HARQ-F and HARQ-h specifications sometimes perform even better, although there does not appear to be a single specification that systematically dominates all other.

The HARQ-F model, in particular, performs well for the increasing estimation window forecasts, but not so well for the rolling window forecasts. This again underscores the difficulties in accurately estimating the extra adjustment parameters in the HARQ-F model based on “only” 1,000 observations. The HARQ-h model that only adjusts the lag parameter corresponding to the forecast horizon often beats the HARQ model that only adjusts the daily lag parameter. At the weekly horizon, the HARQ and HARQ-h models also both perform better than the standard HAR model. For the monthly horizon, however, it appears more important to adjust the longer lags, and as a result the HARQ-F and HARQ-h models typically both do better than the HARQ model. Of course, as the forecast horizon increases, the forecasts become smoother and closer to the unconditional volatility, and as such the relative gains from adjusting the parameters are invariably reduced.

## 4.4 Robustness

### 4.4.1 Alternative Realized Variance Estimators

The most commonly used 5-minute  $RV$  estimator that underly all of our empirical results discussed above provides a simple way of mitigating the contaminating influences of market microstructure “noise” arising at higher intraday sampling frequencies.<sup>20</sup> However, a multitude of alternative  $RV$  estimators that allow for the use of higher intraday sampling frequencies have, of course, been proposed in the literature. In this section we consider some of the most commonly used of these robust estimators, namely: sub-sampled, two-scales, kernel, and pre-averaged  $RV$ , each described in more detail below. Our imple-

---

<sup>20</sup>As previously noted, the comprehensive comparisons in Liu, Patton, and Sheppard (2015) also show that HAR type models based on the simple 5-minute  $RV$  generally perform quite well in out-of-sample forecasting.

**Table 4.7:** Weekly Out-of-Sample Forecast Losses

		AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ	HARQ-F	HARQ-h	
S&P500											
MSE	RW	1.1450	1.0000	1.4030	0.9919	<b>0.9018</b>	1.0798	0.9475	1.2138	0.8884	
	IW	1.3509	1.0000	1.1549	0.9673	<b>0.8365</b>	1.0861	0.9031	0.9171	0.9232	
QLIKE	RW	1.5589	1.0000	1.3047	1.0417	0.9350	1.1892	<b>0.9159</b>	1.2529	0.9491	
	IW	1.8801	1.0000	1.0898	0.9870	0.8735	1.3717	0.8537	<b>0.7540</b>	0.7996	
Individual Stocks											
MSE	RW	Avg	1.2902	1.0000	1.0580	0.9960	0.9864	1.0985	0.9838	1.0234	<b>0.9765</b>
		Med	1.2859	1.0000	1.0504	0.9948	0.9904	1.1109	0.9806	1.0051	<b>0.9517</b>
	IW	Avg	1.4259	1.0000	1.0500	1.0003	0.9955	1.2126	0.9627	0.9601	<b>0.9477</b>
		Med	1.4322	1.0000	1.0435	1.0005	0.9922	1.2110	0.9596	0.9378	<b>0.9311</b>
QLIKE	RW	Aveg	1.6564	1.0000	1.1034	1.0124	0.9820	1.2111	<b>0.9309</b>	1.0665	0.9873
		Med	1.6554	1.0000	1.0980	1.0156	0.9827	1.2010	<b>0.9422</b>	1.0673	0.9869
	IW	Avg	1.9062	1.0000	1.0894	1.0279	0.9770	1.4147	<b>0.9066</b>	0.8529	0.8530
		Med	1.8762	1.0000	1.0721	1.0265	0.9781	1.4044	<b>0.9186</b>	0.8420	0.8579

*Note:* The table reports the same loss ratios for the weekly forecasting models previously reported for the one-day-ahead forecasts in Table 4.4. The top panel shows the results for the S&P 500, while the bottom panel gives the average and median ratios across the individual stocks. The lowest ratio in each row is indicated in bold.

**Table 4.8:** Monthly Out-of-Sample Forecast Losses

		AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ	HARQ-F	HARQ-h	
S&P500											
MSE	RW	1.1407	1.0000	0.9841	0.9642	<b>0.9558</b>	1.0964	1.0708	1.3485	1.2191	
	IW	1.2411	1.0000	1.0312	1.0107	1.0119	1.1456	0.9667	<b>0.9339</b>	0.9832	
QLIKE	RW	1.2455	1.0000	1.0552	0.9919	<b>0.9532</b>	1.0518	0.9808	1.1150	1.0450	
	IW	1.4159	1.0000	1.0773	0.9937	0.9842	1.2144	0.9368	<b>0.8448</b>	0.8843	
Individual Stocks											
MSE	RW	Avg	1.2246	1.0000	1.0173	1.0159	0.9924	1.0969	0.9953	1.0198	<b>0.9756</b>
		Med	1.2613	1.0000	1.0118	1.0105	0.9949	1.1005	0.9965	0.9963	<b>0.9619</b>
	IW	Avg	1.4127	1.0000	1.0181	1.0123	0.9907	1.2366	0.9770	<b>0.9723</b>	0.9815
		Med	1.4052	1.0000	1.0172	1.0145	0.9927	1.2182	0.9692	<b>0.9480</b>	0.9705
QLIKE	RW	Avg	1.4125	1.0000	1.0385	1.0143	0.9909	1.1335	0.9485	0.9127	<b>0.8804</b>
		Med	1.4300	1.0000	1.0367	1.0125	0.9928	1.1228	0.9481	0.8778	<b>0.8635</b>
	IW	Avg	1.6612	1.0000	1.0360	1.0257	0.9885	1.3519	0.9371	<b>0.8185</b>	0.8278
		Med	1.6294	1.0000	1.0224	1.0296	0.9912	1.3619	0.9442	<b>0.8245</b>	0.8442

*Note:* The table reports the same loss ratios for the monthly forecasting models previously reported for the one-day-ahead forecasts in Table 4.4. The top panel shows the results for the S&P 500, while the bottom panel gives the average and median ratios across the individual stocks. The lowest ratio in each row is indicated in bold.

## 4. EXPLOITING THE ERRORS

mentation of these alternative estimators will be based on 1-minute returns.<sup>21</sup> We begin by showing that the HARQ model based on the simple 5-minute  $RV$  outperforms the standard HAR models based on these alternative 1-minute robust  $RV$  estimators. We also show that despite the increased efficiency afforded by the use of a higher intraday sampling frequency, the HARQ models based on these alternative  $RV$  estimators still offer significant forecast improvements relative to the standard HAR models based on the same robust  $RV$  estimators. To allow for a direct comparison across the different estimators and models, we always take daily 5-minute  $RV$  as the forecast target. As such, the set-up mirrors that of the CHAR model in equation (4.16) with the different noise-robust  $RV$  estimators in place of the jump-robust  $BPV$  estimator.

The subsampled version of  $RV$  ( $SS-RV$ ) was introduced by Zhang, Mykland, and Aït-Sahalia (2005). Subsampling provides a simple way to improve on the efficiency of the standard  $RV$  estimator, by averaging over multiple time grids. Specifically, by computing the 5-minute  $RV$  on time grids with 5-minute intervals starting at 9:30, 9:31, 9:32, 9:33 and 9:34, the  $SS-RV$  estimator is obtained as the average of these five different  $RV$  estimators. The two-scale  $RV$  ( $TS-RV$ ) of Zhang, Mykland, and Aït-Sahalia (2005), bias-corrects the  $SS-RV$  estimator through a jackknife type adjustment and the use of  $RV$  at the highest possible frequency. It may be expressed as,

$$TS-RV = SS-RV - \frac{M}{M^{(all)}}RV^{(all)}, \quad (4.20)$$

where  $M^{(all)}$  denotes the number of observations at the highest frequency (here 1-minute), and  $RV^{(all)}$  refers to the resulting standard  $RV$  estimator. The realized kernel ( $RK$ ), developed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), takes the form,

$$RK = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^M r_{t,i}r_{t,i-|h|}, \quad (4.21)$$

where  $k(x)$  is a kernel weight function, and  $H$  is a bandwidth parameter. We follow Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) and use a Parzen kernel with their recommended choice of bandwidth. Finally, we also implement the pre-averaged  $RV$  ( $PA-RV$ ) estimator of Jacod, Li, Mykland, Podolskij, and Vetter (2009), defined by,

$$PA-RV = \frac{1}{\sqrt{M}\theta\psi_2} \sum_{i=1}^{M-H+2} \bar{r}_{t,i}^2 - \frac{\psi_1}{2M\theta^2\psi_2}RV, \quad (4.22)$$

where  $\bar{r}_{t,i} = \sum_{j=1}^{H-1} g(j/H)r_{t,j}$ . For implementation we follow Jacod, Li, Mykland, Podolskij, and Vetter (2009) in choosing the weighting function  $g(x) = \min(x, 1-x)$ , along

---

<sup>21</sup>Since we only have access to 5-minute returns for the S&P 500 futures contract, our results in this section pertaining to the market index will be based on the SPY ETF contract. We purposely do not use the SPY in the rest of the paper, as it is less actively traded than the S&P 500 futures for the earlier part of our sample.

**Table 4.9:** HAR Models based on Noise-Robust  $RV$ s versus HARQ Model

		RV	SS-RV	TS-RV	RK	PA-RV	
S&P 500							
MSE	RW	1.2574	1.0801	1.3472	1.3443	1.3521	
	IW	1.1290	1.1882	1.2468	1.1769	1.1604	
QLIKE	RW	1.0025	1.0149	1.1476	1.0493	1.0331	
	IW	1.1487	1.1424	1.2637	1.4182	1.3608	
Individual Stocks							
MSE	RW	Average	1.0714	1.0523	1.1641	1.0446	1.0476
		Median	1.0628	1.0430	1.1665	1.0604	1.0542
	IW	Average	1.0531	1.0502	1.1319	1.0603	1.0579
		Median	1.0499	1.0544	1.1206	1.0695	1.0479
QLIKE	RW	Average	1.0100	1.0438	1.0890	1.0615	1.1107
		Median	1.0064	1.0457	1.0923	1.0587	1.0988
	IW	Average	1.0552	1.0535	1.1297	1.1557	1.1425
		Median	1.0471	1.0491	1.1240	1.1520	1.1458

*Note:* The table reports the loss ratios of the HAR model using 1-minute noise-robust  $RV$  estimators and the 5-minute  $RV$  used previously, compared to the loss of the HARQ model using 5-minute  $RV$ . The S&P 500 results are based on returns for the SPY.

with their recommendations for the data-driven bandwidth  $H$ , and tuning parameter  $\theta$ . The  $\psi_i$  parameters are all functionals of the weighting function  $g(x)$ .

Table 4.9 compares the out-of-sample performance of the standard HAR model based on these more efficient noise-robust estimators with the forecasts from the HARQ model that uses 5-minute  $RV$ . (For comparison, the first column of this table shows the performance of the HAR model using 5-minute  $RV$ ; the entries here are the inverses of those in the “HARQ” column in Table 4.4.) As the table shows, the HARQ model based on the 5-minute  $RV$  easily outperforms the forecasts from the standard HAR models based on the 1-minute robust  $RV$  estimators. In fact, in line with the results of Liu, Patton, and Sheppard (2015), for most of the series the HAR models based on the noise-robust estimators do not systematically improve on the standard HAR model based on the 5-minute  $RV$ . While still inferior, the HAR model based on  $SS - RV$  gets closest in performance to the HARQ model, while the standard HAR model based on the other three estimators generally perform far worse. Of course, the  $TS - RV$ ,  $RK$ , and  $PA - RV$  estimators were all developed to allow for the consistent estimation of  $IV$  through the use of ever finely sampled returns. Thus, it is possible that even finer sampled  $RV$ s than the 1-minute frequency used here might outweigh the additional complexity of the estimators and result in better out-of-sample forecasts.

Further along these lines, all of the noise-robust 1-minute  $RV$  estimators are, of course, still subject to some measurement errors. To investigate whether adjusting for these errors remains useful, we also estimate HARQ models based on each of the alternative  $RV$  measures.<sup>22</sup> Table 4.10 shows the ratios of the resulting losses for the  $HARQ$  to  $HAR$

<sup>22</sup>Since the measurement error variance for all of the estimators are proportional to  $IQ$  (up to a

## 4. EXPLOITING THE ERRORS

**Table 4.10:** HARQ versus HAR Models based on Noise-Robust  $RV$ s

		RV	SS-RV	TS-RV	RK	PA-RV	
S&P 500							
MSE	RW	0.7953	1.0059	0.8606	0.7873	0.7896	
	IW	0.8857	0.8837	0.9749	0.8956	0.8952	
QLIKE	RW	0.9975	1.0585	0.9776	0.9711	1.0504	
	IW	0.8705	0.9195	0.9317	0.8971	0.9262	
Individual Stocks							
MSE	RW	Average	0.9333	0.9496	0.9547	1.0072	0.9698
		Median	0.9409	0.9593	0.9521	1.0012	0.9719
	IW	Average	0.9496	0.9582	0.9723	0.9755	0.9700
		Median	0.9525	0.9590	0.9730	0.9710	0.9771
QLIKE	RW	Average	0.9901	0.9462	0.9804	1.0874	0.9748
		Median	0.9936	0.9600	0.9831	0.9973	0.9793
	IW	Average	0.9477	0.9474	0.9665	0.9492	0.9443
		Median	0.9550	0.9447	0.9624	0.9431	0.9445

*Note:* The table reports the loss ratios of the HARQ models to the HAR models for the 1-minute noise-robust estimators. The S&P 500 results are based on returns for the SPY.

models using a given realized measure.<sup>23</sup> For ease of comparison, we also include the previous results for the simple 5-minute  $RV$  estimator. The use of higher 1-minute sampling and the more efficient  $RV$  estimators, should in theory result in smaller measurement errors. Consistent with this idea, the improvements afforded by the HARQ model for the 1-minute noise-robust estimators are typically smaller than for the 5-minute  $RV$ . However, the HARQ models for the 1-minute  $RV$ s still offer clear improvements over the standard HAR models, with most of the ratios below one.<sup>24</sup>

### 4.4.2 Alternative Quarticity Estimators

The integrated quarticity is notoriously difficult to estimate. To investigate the sensitivity of our results to this additional layer of estimation uncertainty, we consider a number of alternative estimators of  $IQ$  in place of the standard  $RQ$  estimator used in the HARQ models throughout the rest of the paper.

One such estimator is provided by the Tri-Power Quarticity of Barndorff-Nielsen and

---

small noise term, which is negligible at the 1-minute level), we continue to rely on the 5-minute  $RQ$  for estimating the measurement error variance in these HARQ models.

<sup>23</sup>We also investigated models for predicting the alternative volatility measures used in the prediction models in Table 4.10, resulting in qualitatively similar findings to ones for predicting  $RV$  reported here.

<sup>24</sup>Chaker and Meddahi (2013) have previously explored the use of  $RV^{(all)}$  as an estimator (up to scale) for the variance of the market microstructure noise in the context of  $RV$ -based volatility forecasting. Motivated by this idea, we also experimented with the inclusion of an  $RV^{(all)1/2} \cdot RV$  interaction term in the HAR and HARQ models as a simple way to adjust for the temporal variation in the magnitude of the market microstructure noise. The out-of-sample forecasts obtained from these alternative specifications were generally inferior to the forecasts from the basic HARQ model. Additional details of these results are available in Appendix D.

Shephard (2006),

$$TPQ_t \equiv M \mu_{4/3}^{-3} \sum_{i=1}^{M-2} |r_{t,i}|^{4/3} |r_{t+1,i}|^{4/3} |r_{t+2,i}|^{4/3}, \quad (4.23)$$

where  $\mu_{4/3} \equiv 2^{2/3} \Gamma(7/6) / \Gamma(1/2) = \mathbb{E}(|Z|^{4/3})$ . In contrast to the standard  $RQ$  estimator,  $TPQ$  remains consistent for  $IQ$  in the presence of jumps. Further along these lines, we also consider the jump-robust MedRQ estimator developed by Andersen, Dobrev, and Schaumburg (2012, 2014),

$$MedRQ \equiv \frac{3\pi}{9\pi + 72 - 52\sqrt{3}} \frac{M^2}{M-2} \sum_{i=1}^{M-2} \text{median}(|r_{t,i}|, |r_{t,i+1}|, |r_{t,i+2}|)^4, \quad (4.24)$$

as well as the Truncated RQ estimator based on the ideas of Mancini (2009), formally defined by,

$$TrRQ \equiv M \sum_{i=1}^M |r_{t,i}^4| \mathbb{I}_{\{|r_{t,i}| \leq \alpha_i M^\varpi\}}, \quad (4.25)$$

with the tuning parameters  $\alpha_i$  and  $\varpi$  implemented as in Bollerslev, Todorov, and Li (2013). It has also previously been suggested that the integrated quarticity may be more accurately estimated using a coarser time grid than that used for estimating integrated variance; see, e.g., Bandi and Russell (2008). To this end, we consider estimating  $RQ$  based on coarser 15-minute returns. The approximation in equation (4.4) that motivates the HARQ model is, of course, asymptotic (for  $M \rightarrow \infty$ ). Finally, we also consider the wild bootstrap of Gonçalves and Meddahi (2009). Specifically, re-sampling the high-frequency returns 999 times for each of the days in the sample, we use the sample variance of the relevant  $RV$  across the bootstrap replications as the estimator of the measurement error variance.

For ease of comparison, Table 4.11 summarizes the out-of-sample forecast losses from the HARQ models based on each of these different  $IQ$  estimators relative to the losses for the HARQ model based on the standard 5-minute  $RQ$  estimator used in the rest of the paper. As the results show, the performance of the HARQ models based on the different  $IQ$  estimators are generally close. The only alternative estimator that possibly improves on the standard  $RQ$  estimator in a systematic fashion is the bootstrapped variance of  $RV$ . However, the improvements are at best minor.

### 4.4.3 Alternative HARQ Specifications

The HARQ model is designed to allow for temporal variation in the degree of attenuation bias based on an estimate of the heteroskedastic measurement error variance. The exact specification of the model, and the interaction of  $RV$  with the square root of  $RQ$ , is, however, somewhat ad hoc and primarily motivated by concerns related to numerical

## 4. EXPLOITING THE ERRORS

**Table 4.11:** Alternative  $IQ$  estimators.

IQ-estimator		RQ	TPQ	MedRQ	TrRQ	RQ <sub>15min</sub>	Bootstrap	
S&P500								
MSE	RW	1.0000	1.0497	1.0254	1.0208	1.0590	<b>0.9925</b>	
	IW	1.0000	1.1635	1.0328	<b>0.9948</b>	0.9805	0.9981	
QLIKE	RW	1.0000	1.0933	1.1227	0.9971	1.0231	<b>0.9933</b>	
	IW	1.0000	<b>0.9814</b>	1.0548	1.2060	1.0414	0.9998	
Individual Stocks								
MSE	RW	Avg	1.0000	1.0403	1.0139	1.0531	1.0586	<b>0.9936</b>
		Med	<b>1.0000</b>	1.0497	1.0201	1.0378	1.0238	1.0003
	IW	Avg	1.0000	1.0211	1.0191	1.0491	1.0229	<b>0.9994</b>
		Med	1.0000	1.0220	1.0259	1.0558	1.0125	<b>0.9998</b>
QLIKE	RW	Avg	1.0000	1.0040	1.0089	1.0541	1.0254	<b>0.9995</b>
		Med	1.0000	0.9980	<b>0.9968</b>	1.0376	1.0015	0.9992
	IW	Avg	<b>1.0000</b>	1.0050	1.0047	1.0390	1.0014	1.0001
		Med	<b>1.0000</b>	1.0040	1.0065	1.0303	0.9989	1.0001

*Note:* The table reports the out-of-sample forecast losses from the HARQ model using different  $IQ$  estimators. All of the losses are reported relative to the losses from the HARQ model based on the standard  $RQ$  estimator used throughout the rest of the paper. The top panel shows the results for the S&P500, while the bottom panel reports the average and median ratios across each of the individual stocks. The lowest ratio in each row is indicated in bold.

stability. Equation (4.9), in particular, naturally suggests allowing the parameters of the model to vary with the inverse of  $RQ$ , as opposed to  $RQ^{1/2}$ . But,  $RQ$  may be close to zero, and as such the inverse of  $RQ$  is prone to amplify any estimation errors, resulting in inaccurate estimates of the inverse of  $IQ$ , and in turn unstable time-varying autoregressive parameter estimates. To further investigate this issue, we consider the out-of-sample forecasts from alternative HARQ specifications in which we substitute  $RQ$ ,  $RQ^{-1/2}$ ,  $RQ^{-1}$ , and  $\log(RQ)$  in place of  $RQ^{1/2}$  in equation (4.2.3). All of the HARQ specifications that we have considered so far have also restricted the intercept in the models, or  $\beta_0$ , to be constant. We now consider two alternative specifications, where we add  $RQ^{1/2}$  as an additional explanatory variable to the standard HAR and HARQ models, thereby allowing for a time-varying intercept in the HAR(Q) model. Table 4.12 reports the out-of-sample forecast results from each of these alternative HARQ specifications. For ease of comparison, we again normalize all of the losses relative to the losses of the HARQ model based on  $RQ^{1/2}$  analyzed throughout. The first set of columns show the results for the alternative  $RQ$  transformations. The models based on  $RQ^{1/2}$  and  $RQ$  perform roughly the same. Meanwhile, as expected, the modified HARQ models that rely on negative powers of  $RQ$ , or the log-transform of  $RQ$ , all result in numerically unstable parameter estimates and correspondingly inferior out-of-sample forecasts.<sup>25</sup>

The last two columns of the table give the results from directly including  $RQ^{1/2}$  in

<sup>25</sup>We also experimented with “higher-order” models, allowing the autoregressive parameters to depend on multiple  $RQ$  terms; e.g., both  $RQ^{1/2}$  and  $RQ$ . However, the difficulties in accurately estimating the parameters for these more elaborate models, again translate into inferior out-of-sample forecasts.

**Table 4.12:** Alternative HARQ Specifications

		Alternative $RQ$ Transformations					Adding $RQ^{1/2}$		
		$RQ$	$RQ^{1/2}$	$RQ^{-1/2}$	$RQ^{-1}$	$Log(RQ)$	HAR	HARQ	
S&P 500									
MSE	RW	1.0037	<b>1.0000</b>	1.2123	1.2334	1.3313	1.1552	1.0004	
	IW	1.0344	<b>1.0000</b>	1.1166	1.1357	1.0736	1.1166	1.1402	
QLIKE	RW	<b>0.9484</b>	1.0000	1.0952	1.0950	1.8104	0.9919	<b>0.9731</b>	
	IW	1.0222	<b>1.0000</b>	1.1327	1.3217	2.0107	1.0452	1.0089	
Individual Stocks									
MSE	RW	Avg	1.0108	<b>1.0000</b>	1.0808	1.0931	1.0329	1.0339	1.0207
		Med	1.0112	<b>1.0000</b>	1.0577	1.0664	1.0336	1.0050	<b>0.9904</b>
	IW	Avg	1.0189	<b>1.0000</b>	1.0495	1.0644	1.0143	0.9979	<b>0.9895</b>
		Med	1.0198	<b>1.0000</b>	1.0403	1.0598	1.0082	0.9986	<b>0.9863</b>
QLIKE	RW	Avg	<b>0.9973</b>	1.0000	1.0678	1.0814	1.3723	1.0377	1.0639
		Med	<b>0.9847</b>	1.0000	1.0458	1.0579	1.3324	1.0219	1.0133
	IW	Avg	1.0263	<b>1.0000</b>	1.0961	1.1155	1.2903	1.0394	1.0081
		Med	1.0241	<b>1.0000</b>	1.0778	1.0886	1.2084	1.0279	<b>0.9937</b>

*Note:* The table reports the out-of-sample forecast losses for HARQ models based on different transformations of  $RQ$ . All of the losses are reported relative to those for the HARQ model the rely on  $RQ^{1/2}$ . The left panel reports the results based on alternative  $RQ$  interaction terms. The right panel reports the results from models that include  $RQ^{1/2}$  as an explanatory variable. The top panel pertains to the S&P 500, while the bottom panel gives the average and median ratios across all of the individual stocks. The lowest ratio in each row is indicated in bold.

the HAR and HARQ models. While the HAR model with  $RQ^{1/2}$  tend to performs worse than the HARQ model, it does improve on the standard HAR model, as it is able to mimic the HARQ model and reduce the forecasts in situations with large measurement errors. However, the model does so less effectively than the HARQ model forecasts. The HARQ model that includes  $RQ^{1/2}$  does improve on the standard HARQ model for some of the forecast scenarios, but performs worse in others, and in no case is the forecast improvement very large. Overall, we conclude that the simple HARQ model in equation (4.2.3) appears the more stable and generally superior model compared to any of these alternative specifications.

#### 4.4.4 Alternative Q-Models

The motivation behind the HARQ model is equally applicable to other realized volatility based forecasting models, including the benchmark models analyzed in our forecast comparisons. In particular, the HAR-J model defined in equation (4.14) is readily modified in a manner completely analogous to the HARQ model, resulting in the HARQ-J model,

$$RV_t = \beta_0 + (\beta_1 + \beta_{1Q}RQ_{t-1}^{1/2})RV_{t-1} + \beta_2RV_{t-1|t-5} + \beta_3RV_{t-1|t-22} + \beta_JJ_{t-1} + u_t. \quad (4.26)$$

The CHAR model in (4.16) relies on the jump-robust bi-power variation ( $BPV$ ) measure in place of the realized volatility for predicting the future volatility. As shown by

## 4. EXPLOITING THE ERRORS

**Table 4.13:** Alternative Q-Model In-Sample Estimates

	HAR-J	HARQ-J	CHAR	CHARQ	SHAR	SHARQ
$\beta_0$	0.1208 (0.0606)	0.0045 (0.0561)	0.1361 (0.0595)	-0.0064 (0.0618)	0.0692 (0.0667)	-0.0766 (0.0613)
$\beta_1$	0.3599 (0.0891)	0.6035 (0.0882)	0.2657 (0.0958)	0.5834 (0.0967)		
$\beta_2$	0.4341 (0.1300)	0.3519 (0.1285)	0.4980 (0.1489)	0.4189 (0.1524)	0.4176 (0.1223)	0.3527 (0.1260)
$\beta_3$	0.1856 (0.1068)	0.1057 (0.1034)	0.1751 (0.1201)	0.1131 (0.1138)	0.1530 (0.1013)	0.0822 (0.0997)
$\beta_J$	-1.0033 (0.3668)	-0.3393 (0.2857)				
$\beta_1^+$					-0.3734 (0.1772)	-0.2027 (0.2054)
$\beta_1^-$					1.1282 (0.2773)	1.5723 (0.2658)
$\beta_{1Q}$		-0.3266 (0.0617)		-0.5410 (0.1800)		
$\beta_{1Q}^+$						-1.3227 (0.3632)
$\beta_{1Q}^-$						0.2485 (0.1316)
$R^2$	0.5376	0.5638	0.5347	0.5526	0.5751	0.5972
MSE	2.4908	2.3495	2.5064	2.4097	2.2887	2.1693
QLIKE	0.1538	0.1336	0.1442	0.1377	0.3315	0.2154
$\overline{R}^2$ Stocks	0.4913	0.5110	0.4891	0.5106	0.4986	0.5239
$\overline{MSE}$ Stocks	14.8224	14.0916	14.9265	14.1891	14.5431	13.6915
$\overline{QLIKE}$ Stocks	0.1492	0.1470	0.1509	0.1449	0.1496	0.1534

*Note:* The table reports the S&P 500 in-sample parameter estimates and measures of fit for the different benchmark models and HARQ-adaptations discussed in the main text. The bottom panel shows the average  $R^2$ s,  $MSE$ s, and  $QLIKE$ s across the individual stocks.

Barndorff-Nielsen and Shephard (2006), the asymptotic variance of  $BPV$  equals  $2.61\Delta IQ_t$ . This asymptotic variance is naturally estimated by the Tri-Power Quarticity ( $TPQ$ ) previously defined in equation (4.23), which remains consistent for  $IQ$  in the presence of jumps. Correspondingly, we define the CHARQ model as,

$$RV_t = \beta_0 + (\beta_1 + \beta_{1Q}TPQ_{t-1}^{1/2})BPV_{t-1} + \beta_2BPV_{t-1|t-5} + \beta_3BPV_{t-1|t-22} + u_t. \quad (4.27)$$

The asymptotic distribution of the  $RV_t^+$  and  $RV_t^-$  measures included in the SHAR model in equation (4.17) is unknown. However, a measure that is strongly correlated with their asymptotic variances should work well in terms of adjusting the parameters for measurement errors, as the estimated regression coefficients will automatically make up for any proportional differences. Hence, as a shortcut we simply rely on  $RQ$  to proxy the

**Table 4.14:** Alternative Q-Model Out-of-Sample Forecast Losses

		HARQ	HARQ-J	CHARQ	SHARQ	
S&P500						
MSE	RW	0.8266	0.9243	0.8951	1.4412	
	IW	0.8944	0.9335	1.0609	1.1027	
QLIKE	RW	0.9464	0.9653	1.0235	1.4576	
	IW	0.8809	0.9015	0.8825	1.2849	
Individual Stocks						
MSE	RW	Avg	0.9349	0.9397	0.9525	1.1308
		Med	0.9418	0.9513	0.9539	1.0840
	IW	Avg	0.9525	0.9666	0.9451	1.0870
		Med	0.9525	0.9662	0.9548	1.0554
QLIKE	RW	Avg	0.9902	0.9902	0.9879	1.1706
		Med	0.9916	0.9952	0.9900	1.1609
	IW	Avg	0.9487	0.9548	0.9306	1.1305
		Med	0.9550	0.9594	0.9277	1.1154

*Note:* The table reports the loss ratios for the alternative Q-model specifications discussed in the main text. All of the losses are reported relative the relevant baseline models without the Q-adjustment terms. The top panel shows the results for the S&P 500, while the bottom panel reports the average and median ratios across all of the individual stocks.

asymptotic variance of both  $RV^+$  and  $RV^-$ , defining the SHARQ model,

$$RV_t = \beta_0 + (\beta_1^+ + \beta_{1Q}^+ RQ_{t-1}^{1/2}) RV_{t-1}^+ + (\beta_1^- + \beta_{1Q}^- RQ_{t-1}^{1/2}) RV_{t-1}^- + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t. \quad (4.28)$$

Table 4.13 reports the parameter estimates for each of these Q-models and their baseline counterparts. The general pattern directly mirrors that of the estimates for the HARQ and HAR models. All of the models shift the weights from the weekly and monthly lags to the daily lag, with higher measurement error variances pulling the daily lag parameters closer to zero. Looking specifically at the HAR-J model, the parameter associated with the jump component is significantly negative: on days where part of the total  $RV$  is attributable to jumps, the next day's  $RV$  is reduced by -1.003 times the jump component. In the HARQ-J model, however, the measurement error subsumes a large portion of this jump variation: the  $\beta_J$  coefficient is reduced by two-thirds, and is no longer statistically significant. This same result holds true for the individual stocks, where the jump parameters are significant at the 5% level for 60% of the stocks for the HAR-J models, compared to only 10% of the stocks for the HARQ-J models. Also, comparing the SHAR and SHARQ models, the latter shifts the weight even further away from the positive part of  $RV$  to the negative part, so that only  $RV^-$  is significant in the SHARQ model.

The out-of-sample forecast results from each of these different Q-models are summarized in Table 4.14. For comparison purposes, we also include the results for the basic HARQ model. To facilitate the interpretation of the results, we report the loss ratios with respect to the relevant baseline models, that is the losses for the HARQ model is reported

relative to the losses for the standard HAR model, the CHARQ model losses relative to the CHAR model losses, and so forth. The improvements obtained for the HARQ-J and CHARQ models are generally in line with those for the basic HARQ model. This is true both for the S&P 500 losses given in the top panel and the average losses across the individual stocks reported in the bottom panel. The SHARQ model, however, does not improve on the standard SHAR model. Of course, in contrast to the other Q-models, our specification of the SHARQ model relies on an imperfect proxy for the asymptotic variance of the measurement errors. This might help explain the relatively poor performance of that model, and also indirectly highlight the importance of using a proper approximation for the distribution of the measurement errors to guide the adjustments of the autoregressive parameters and the forecasts from the models.

### 4.5 Conclusion

We propose a simple-to-implement new class of realized volatility based forecasting models. The models improve on the forecasts from standard volatility forecasting models, by explicitly accounting for the temporal variation in the magnitude of the measurement errors in the realized volatilities and the way in which the errors attenuate the parameters of the models. A particularly simple member of this new class of models, which we term the HARQ model, outperforms the forecasts from several other commonly used models. This holds true both in simulations and in- and out-of-sample forecasts of the volatility of the S&P 500 aggregate market portfolio and a number of individual stocks.

The new models developed here may usefully be applied in many other situations. The volatility risk premium, for example, defined as the difference between the so-called risk neutral expectation of the future volatility of the aggregate market portfolio and the actual statistical expectation of the market volatility, has recently received a lot of attention in the literature. The risk neutral expectation of the volatility is readily inferred from options prices in an essentially model-free manner. The actual volatility forecasts, however, invariably depends on the model used for constructing the forecasts. Bekaert and Hoerova (2014) and Conrad and Loch (2015) have both recently demonstrated how the use of different realized volatility based forecasting models, including versions of the HAR, HAR-J and CHAR models analyzed here, can materially affect the estimates of the volatility risk premium and the interpretation thereof. The HARQ models, of course, hold the promise of even more accurate forecasts and better volatility risk premium estimates, and in turn new insights and a deeper understanding of the economic mechanisms behind the temporal variation in the premium.

All of the forecasting models developed here are univariate. However, most practical questions related to risk measurement and management are intrinsically multivariate in nature, requiring the forecasts of both asset return variances and covariances. Building on the distributional results of Barndorff-Nielsen and Shephard (2004a), the realized volatility

based Vech HAR model of Chiriac and Voev (2011) may readily be extended to incorporate the effect of the measurement errors in the realized variances and covariance in a manner analogous to the one employed here for the univariate HARQ models. The multivariate HEAVY model of Noureldin, Shephard, and Sheppard (2012) and the Realized Beta GARCH model of Hansen, Lunde, and Voev (2014) may similarly be extended to allow the parameters of the models to vary with the degree of the measurement errors in the realized covariance matrix. We leave further work along these lines for future research.

## Appendix A: Simulation Design

Our simulations are based on the two-factor stochastic volatility model previously analyzed by Huang and Tauchen (2005) among others,

$$\begin{aligned}
 d \log S_t &= \mu dt + \sigma_{ut} \nu_t \left( \rho_1 dW_{1t} + \rho_2 dW_{2t} + \sqrt{1 - \rho_1^2 - \rho_2^2} dW_{3t} \right) \\
 \nu_t^2 &= \text{s-exp} \{ \beta_0 + \beta_1 \nu_{1t}^2 + \beta_2 \nu_{2t}^2 \} \\
 d\nu_{1t}^2 &= \alpha_1 \nu_{1t}^2 dt + dW_{1t} \\
 d\nu_{2t}^2 &= \alpha_2 \nu_{2t}^2 dt + (1 + \phi \nu_{2t}^2) dW_{2t} \\
 \sigma_{ut} &= C + Ae^{-at} + Be^{-b(1-t)},
 \end{aligned}$$

where s-exp denotes the exponential function with a polynomial splined at high values to avoid explosive behavior. We follow Huang and Tauchen (2005) in setting  $\alpha = 0.03$ ,  $\beta_0 = -1.2$ ,  $\beta_1 = 0.04$ ,  $\beta_2 = 1.5$ ,  $\alpha_1 = -0.00137$ ,  $\alpha_2 = -1.386$ ,  $\phi_1 = 0.25$ , and  $\rho_1 = \rho_2 = -0.3$ . We initialize the persistent factor  $\nu_1$  by drawing  $\nu_{1,0} \sim N(0, \frac{-1}{2\alpha_1})$  from its unconditional distribution. The  $\nu_2$  factor is initialized at 0. The intraday volatility pattern is modeled by means of the diurnal U-shape  $\sigma_{ut}$  function. Following Andersen, Dobrev, and Schaumburg (2012), we set  $A = 0.75$ ,  $B = 0.25$ ,  $C = 0.88929198$ , and  $a = b = 10$ , respectively. The simulations are generated using an Euler scheme based on 23,400 intervals for each of the  $T = 2,000$  “days” in the sample. We then aggregate these prices to sparsely sampled  $M = 39,78,390$  return observations per day, corresponding to 10-, 5- and 1-“minute” returns.

To allow for empirically more realistic high-frequency prices, we further add “noise” to the simulated price process. In line with the empirical evidence in Bandi and Russell (2006) and Hansen and Lunde (2006b), we allow the variance of the noise to increase with the volatility of the simulated efficient price. In particular, mirroring the design in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), on each day we generate an *i.i.d.* noise term  $u_{t,i} \sim N(0, \omega_t^2)$  with  $\omega_t^2 = \xi^2 \int_{t-1}^t \sigma_{us}^2 \nu_s^2 ds$ , so that the variance of the noise is constant throughout the day, but changes from day to day. This noise is then added to the  $S_{t,i}$  price process to obtain the time series of actual high-frequency simulate prices  $S_{t,i}^* = S_{t,i} + u_{t,i}$ .

## Appendix B: Out-of-sample forecast results

**Table 4.15:** MSE Ratios - Rolling Window

	AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ
MSFT	1.1934	1.0000	1.0071	1.0205	1.0182	1.0488	0.9878
KO	1.1780	1.0000	1.0665	1.0228	1.0045	0.9393	0.8889
DD	1.1730	1.0000	1.0299	1.0166	0.9867	1.0175	0.9652
XOM	1.0300	1.0000	1.3184	1.0558	0.9921	1.0212	0.9792
GE	1.0379	1.0000	1.0941	1.1415	1.1253	0.9958	0.9564
IBM	1.2765	1.0000	1.0171	1.0357	0.9868	0.9681	0.9204
CVX	1.2671	1.0000	1.0023	1.0370	0.9478	0.9160	0.8825
UTX	1.1472	1.0000	1.0815	1.0395	0.9853	0.9094	0.9159
PG	1.1036	1.0000	0.9494	0.9489	1.0258	1.0175	1.0098
CAT	0.9671	1.0000	0.9488	0.9081	1.0047	0.9888	0.9571
BA	1.1498	1.0000	0.9663	0.9624	0.9796	1.0112	0.9456
PFE	1.1789	1.0000	1.0103	0.9995	1.0263	0.9896	0.9418
JNJ	0.9323	1.0000	0.9926	1.0158	1.0041	0.8428	0.9990
MMM	1.2289	1.0000	0.9328	1.0820	0.9370	0.8540	0.8706
MRK	1.0721	1.0000	0.7373	0.7634	0.7480	0.7277	0.7378
DIS	1.0038	1.0000	0.9019	0.9735	1.0704	0.9586	0.9093
MCD	1.1102	1.0000	1.1189	1.1168	1.1028	0.9412	0.9388
JPM	1.0389	1.0000	1.0733	1.0047	0.9924	0.8247	0.8342
WMT	1.2096	1.0000	1.0236	0.9895	1.0020	1.0502	1.0403
NKE	1.2602	1.0000	1.0376	1.0401	1.0049	1.0804	0.9614
AXP	1.3382	1.0000	1.0115	1.0219	0.9872	0.8759	0.8572
INTC	1.2063	1.0000	1.0007	0.9971	0.9917	0.9953	0.9400
TRV	1.1022	1.0000	1.0188	1.0107	1.2486	1.1935	1.1221
VZ	1.2448	1.0000	1.0043	0.9909	1.0488	0.9570	0.8631
HD	1.2732	1.0000	1.0560	1.0458	1.0018	0.9263	0.8758
CSCO	1.1999	1.0000	0.9865	0.9784	1.0070	1.0415	0.9662
UNH	1.1402	1.0000	1.0193	0.9974	0.9952	0.9864	0.9750

*Note:* The table reports the loss ratios of the HARQ to HAR models for each of the individual stocks.

**Table 4.16:** MSE Ratios - Increasing Window

	AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ
MSFT	1.2165	1.0000	1.0082	1.0125	1.0129	1.0941	0.9690
KO	1.2337	1.0000	0.9761	0.9644	1.0311	1.0015	0.9209
DD	1.1998	1.0000	1.0028	0.9968	0.9805	1.0962	0.9910
XOM	1.1875	1.0000	1.0012	1.0519	0.9550	1.0270	0.9278
GE	1.1065	1.0000	1.0047	0.9914	0.9519	0.9615	0.9484
IBM	1.2835	1.0000	1.0038	1.0165	0.9919	1.0701	0.9447
CVX	1.2589	1.0000	1.0048	1.0401	0.9352	0.9666	1.0317
UTX	1.1922	1.0000	1.0082	1.0145	0.9692	1.0396	0.9525
PG	1.2092	1.0000	1.0423	1.0081	1.0028	1.0724	0.9794
CAT	1.1902	1.0000	0.9739	0.9523	1.0003	1.1058	1.0036
BA	1.2571	1.0000	0.9623	0.9620	0.9936	1.0872	0.9538
PFE	1.2161	1.0000	0.9989	0.9942	1.0060	1.0197	0.9399
JNJ	1.2026	1.0000	0.9949	0.9905	1.0042	1.0207	0.9478
MMM	1.2182	1.0000	1.0123	1.0233	0.9556	0.9522	0.9242
MRK	1.1207	1.0000	1.0925	1.0811	1.0656	0.8663	0.8485
DIS	1.1535	1.0000	0.9985	0.9625	1.0765	1.0879	0.9793
MCD	1.1813	1.0000	0.9901	0.9797	0.9908	1.0236	0.9515
JPM	1.0496	1.0000	1.0289	1.0115	0.9605	0.8818	0.8537
WMT	1.3765	1.0000	1.0061	0.9901	0.9768	1.1509	0.9851
NKE	1.3517	1.0000	1.0008	1.0012	0.9969	1.1293	0.9525
AXP	1.3136	1.0000	0.9892	1.0010	0.9758	0.9586	0.8794
INTC	1.2275	1.0000	0.9896	0.9923	0.9872	1.0323	0.9541
TRV	1.1002	1.0000	1.0242	1.0111	1.0283	1.0517	1.0134
VZ	1.2547	1.0000	0.9911	0.9831	1.0055	1.0928	0.9614
HD	1.2470	1.0000	1.0134	1.0142	1.0012	0.9756	0.9006
CSCO	1.2340	1.0000	0.9802	0.9745	1.0051	1.0820	0.9755
UNH	1.1690	1.0000	1.0096	1.0130	0.9968	1.1536	1.0282

*Note:* The table reports the loss ratios of the HARQ to HAR models for each of the individual stocks.

**Table 4.17:** QLIKE Ratios - Rolling Window

	AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ
MSFT	1.3690	1.0000	0.9984	1.0264	0.9970	1.1894	0.9900
KO	1.4975	1.0000	0.9762	0.9595	1.0113	1.1143	0.9389
DD	1.3252	1.0000	1.0087	1.0014	0.9816	1.1994	1.0444
XOM	1.3568	1.0000	0.9866	1.0049	1.0019	1.1434	0.9828
GE	1.4602	1.0000	1.0552	1.0113	0.9565	0.9702	0.8334
IBM	1.5559	1.0000	0.9860	1.0025	0.9824	1.2305	0.9789
CVX	1.3802	1.0000	0.9858	1.0209	1.0142	1.2134	0.9971
UTX	1.3484	1.0000	0.9836	1.0017	0.9645	1.2397	1.0662
PG	1.4715	1.0000	0.9601	0.9572	0.9817	1.0758	0.9863
CAT	1.3267	1.0000	0.9976	1.0071	0.9939	1.1781	1.0224
BA	1.2353	1.0000	1.0031	0.9945	1.0001	1.2240	1.0497
PFE	1.2521	1.0000	0.9896	0.9868	0.9995	1.0744	0.9881
JNJ	1.4044	1.0000	1.0330	1.0052	1.0061	1.1279	1.0389
MMM	1.3493	1.0000	1.0068	1.0282	0.9527	1.0553	0.9599
MRK	1.3383	1.0000	1.0325	0.9855	0.9887	1.0005	0.9210
DIS	1.3685	1.0000	1.0082	1.0527	0.9955	1.1561	1.0147
MCD	1.4891	1.0000	0.9418	0.9317	0.9677	1.0341	0.9128
JPM	1.4688	1.0000	0.9926	0.9803	0.9977	1.1916	1.1107
WMT	1.4935	1.0000	0.9929	0.9831	0.9830	1.2255	1.0203
NKE	1.4670	1.0000	0.9960	1.0068	0.9870	1.2282	1.0004
AXP	1.7802	1.0000	1.0762	0.9868	0.9511	1.2099	0.9435
INTC	1.4305	1.0000	0.9966	1.0004	0.9884	1.1133	0.9698
TRV	1.4883	1.0000	1.0036	1.0308	1.0194	1.1947	0.9276
VZ	1.3952	1.0000	0.9908	0.9771	0.9941	1.1832	1.0224
HD	1.5055	1.0000	1.0014	1.0184	1.0057	1.1130	1.0177
CSCO	1.4370	1.0000	1.0445	1.0144	1.0107	1.2513	1.0054
UNH	1.3574	1.0000	0.9996	1.0218	1.0033	1.1082	0.9916

*Note:* The table reports the loss ratios of the HARQ to HAR models for each of the individual stocks.

## 4. EXPLOITING THE ERRORS

---

**Table 4.18:** QLIKE Ratios - Increasing Window

	AR	HAR	HAR-J	CHAR	SHAR	ARQ	HARQ
MSFT	1.5419	1.0000	1.0111	1.0391	0.9929	1.2737	0.9714
KO	1.6397	1.0000	0.9845	0.9978	0.9985	1.2566	0.9441
DD	1.4166	1.0000	0.9964	1.0080	0.9806	1.1544	0.9726
XOM	1.4475	1.0000	1.0030	1.0316	0.9857	1.0781	0.9550
GE	1.6703	1.0000	0.9929	1.0281	0.9694	1.1592	0.9028
IBM	1.7969	1.0000	0.9780	0.9881	0.9441	1.3047	0.9039
CVX	1.3868	1.0000	0.9898	1.0409	1.0019	1.1856	0.9888
UTX	1.4787	1.0000	0.9827	1.0111	0.9621	1.1716	0.9606
PG	1.7407	1.0000	0.9609	0.9711	0.9559	1.2534	0.9059
CAT	1.3894	1.0000	1.0001	1.0261	0.9934	1.2128	1.0264
BA	1.4558	1.0000	0.9959	1.0100	0.9915	1.1569	0.9550
PFE	1.4837	1.0000	0.9842	0.9961	1.0007	1.1430	0.9318
JNJ	1.5782	1.0000	1.0068	1.0565	0.9795	1.1359	0.9241
MMM	1.4569	1.0000	0.9996	1.0338	0.9574	1.1669	0.9694
MRK	1.3740	1.0000	0.9918	0.9918	0.9965	1.0626	0.9217
DIS	1.5515	1.0000	0.9902	0.9716	0.9997	1.2901	0.9654
MCD	1.8009	1.0000	0.9773	0.9931	0.9767	1.3586	0.9087
JPM	1.5565	1.0000	1.0022	1.0200	0.9887	1.1505	1.0101
WMT	1.9259	1.0000	0.9587	0.9298	0.9338	1.3100	0.8977
NKE	1.8064	1.0000	0.9922	1.0166	0.9901	1.4343	0.9302
AXP	1.7946	1.0000	1.0097	1.0065	0.9608	1.1732	0.8869
INTC	1.5868	1.0000	0.9987	1.0163	0.9916	1.1464	0.9922
TRV	1.4615	1.0000	1.0124	1.1004	1.0198	1.2069	0.9632
VZ	1.5604	1.0000	0.9843	0.9972	0.9948	1.2401	0.9584
HD	1.6611	1.0000	1.0026	1.0447	0.9845	1.0779	0.9732
CSCO	1.7259	1.0000	1.0046	1.0384	0.9791	1.1942	0.9193
UNH	1.3785	1.0000	1.0009	1.0362	1.0077	1.1677	0.9767

*Note:* The table reports the loss ratios of the HARQ to HAR models for each of the individual stocks.

## Appendix C: Individual stock in-sample estimates

**Table 4.19:** Average Parameter Estimates Across Stocks

	AR	HAR	AR-HL	ARQ	HARQ	HARQ-F
$\beta_0$	1.1252 (0.3612)	0.3050 (0.1011)		0.4600 (0.2617)	0.0960 (0.1155)	-0.0295 (0.1161)
$\beta_1$	0.6263 (0.0844)	0.2661 (0.0908)	0.9350 (0.0271)	0.8874 (0.0974)	0.5233 (0.1142)	0.4710 (0.1188)
$\beta_2$		0.3629 (0.1205)			0.2929 (0.1188)	0.4386 (0.1325)
$\beta_3$		0.2686 (0.0852)			0.1764 (0.0893)	0.1381 (0.1090)
$\beta_{1Q}$				-0.2324 (0.1371)	-0.1504 (0.0929)	-0.1300 (0.0930)
$\beta_{2Q}$						-0.1142 (0.1301)
$\beta_{3Q}$						0.0250 (0.1466)
	HAR-J	HARQ-J	CHAR	CHARQ	SHAR	SHARQ
$\beta_0$	0.3184 (0.1062)	0.1102 (0.1171)	0.3704 (0.1181)	0.1631 (0.1077)	0.2741 (0.1090)	0.0388 (0.1198)
$\beta_1$	0.3284 (0.1001)	0.5392 (0.1090)	0.2996 (0.1036)	0.5700 (0.1118)		
$\beta_2$	0.3477 (0.1227)	0.2915 (0.1191)	0.3980 (0.1282)	0.3257 (0.1244)	0.3434 (0.1161)	0.2845 (0.1183)
$\beta_3$	0.2575 (0.0844)	0.1746 (0.0883)	0.2659 (0.0932)	0.1738 (0.0863)	0.2572 (0.0820)	0.1608 (0.0863)
$\beta_J$	-0.3813 (0.3489)	-0.1637 (0.3005)				
$\beta_1^+$					-0.0052 (0.2559)	0.1758 (0.2367)
$\beta_1^-$					0.6307 (0.3179)	0.9797 (0.3125)
$\beta_{1Q}$		-0.1463 (0.0878)		-0.2262 (0.1510)		
$\beta_{1Q}^+$						-0.3539 (0.3446)
$\beta_{1Q}^-$						-0.0473 (0.2052)

*Note:* This table reports the average parameter estimates across the individual stocks. The standard deviations of the estimates obtained across the stocks are reported in parentheses.

Appendix D: Models including  $RV^{(all)}$ Table 4.20: HAR(Q) Models Including  $RV^{(all)}$ 

Baseline Model		HARQ	HAR	HAR	HARQ	HARQ	
Added Variable			$RV^{(All)1/2}$	$RV^{(All)1/2}$	$RV^{(All)1/2}$	$RV^{(All)1/2}$	
			$RV$		$RV$		
S&P500							
MSE	RW	0.7953	0.8175	0.8396	0.8188	0.8079	
	IW	0.8857	0.9269	1.0049	0.8904	0.8886	
QLIKE	RW	0.9975	0.9251	1.0246	1.0710	1.2329	
	IW	0.8705	0.8694	1.0120	0.8579	0.9091	
Individual Stocks							
MSE	RW	Average	0.9333	0.9403	0.9964	0.9807	0.9678
		Median	0.9409	0.9451	0.9888	0.9353	0.9560
	IW	Average	0.9496	0.9611	1.0054	0.9832	0.9751
		Median	0.9525	0.9646	1.0042	0.9593	0.9752
QLIKE	RW	Average	0.9901	1.0081	1.0197	1.1126	1.0350
		Median	0.9936	1.0057	1.0122	0.9995	1.0151
	IW	Average	0.9477	0.9671	1.0127	0.9647	0.9721
		Median	0.9550	0.9596	0.9964	0.9647	0.9485

*Note:* The table reports the loss ratios for each of the different models relative to the HAR model. The augmented HAR(Q) models include  $RV^{(All)1/2}$  or  $RV^{(All)1/2} \cdot RV$  as an additional explanatory variable, as indicated in the table.  $RV^{(all)}$  is estimated with 1-minute returns. The results for the S&P500 are based on the SPY ETF contract.

## Appendix E: Forecasting alternative IV measures

**Table 4.21:** HARQ Ratios with the Same Left- and Right-Hand-Side Volatility Measure

			SS-RV	TS-RV	RK	PA-RV	BPV
S&P500							
MSE	RW		1.0714	1.2985	0.8753	0.9384	0.8864
	IW		0.8174	0.9864	0.9204	0.7892	0.8843
QLIKE	RW		1.0793	0.9706	1.0185	1.0163	1.0019
	IW		0.8720	0.9583	0.9205	0.9747	0.9010
Individual Stocks							
MSE	RW	Average	0.9504	0.9722	0.9825	0.8075	0.9542
		Median	0.9703	0.9765	0.9904	0.7765	0.9694
	IW	Average	0.9463	0.9846	1.0052	0.7709	0.9463
		Median	0.9612	0.9825	0.9867	0.7404	0.9524
QLIKE	RW	Average	0.9346	0.9874	0.9894	0.9392	0.9651
		Median	0.9535	0.9928	0.9821	0.9448	0.9660
	IW	Average	0.9079	0.9719	0.9644	0.9416	0.9350
		Median	0.9168	0.9740	0.9619	0.9339	0.9384

*Note:* The table reports the loss ratios of the HARQ model to the HAR model where the same *IV* estimator is used on the left- and right-hand side. This table contrasts with Table 4.10, where *RV* is always on the left-hand side.

# Chapter 5

## (Un)Reliable Realized Risk Measures: Portfolio Compositions and Conditional Betas

### 5.1 Introduction

The covariation among asset returns is of critical importance for many economic and financial applications, such as portfolio management and hedging decisions. In practice, the covariance matrix has to be estimated and forecasted from historical data, and the out-of-sample performance can be significantly impacted by parameter uncertainty and estimation errors.

A large component of the estimation error is the result of time-variation in covariances. This results in a trade-off between using many observations to accurately estimate the conditional covariance matrix, and only using recent history to better capture the short-term dynamics, likely increasing estimation error. The wide availability of high-frequency data has ameliorated this trade-off as it allows for more accurate estimation and forecasting of covariance matrices, even on a daily basis (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003; Fleming, Kirby, and Ostdiek, 2003).

Although high-frequency measures lead to better estimation of both (co)variances and derived measures of risk such as betas, they are still estimates of latent population quantities and therefore subject to estimation error. The use of these estimates in dynamic models leads to the classical errors-in-variables problem, where parameters are attenuated towards zero. The attenuation intensity is increasing in the variance of the measurement error. If measurement errors are heteroskedastic, the degree of attenuation is proportional to the average measurement error variance. This means that all observations are treated equally, while some estimates are likely to be close to the true values and others are not. An observation with below average measurement error provides a stronger signal and the autoregressive parameter for this observation would ideally be attenuated less and vice

versa. Hence, for the purpose of forecasting, a constant parameter is suboptimal as the degree of attenuation is never in line with the degree of measurement error.

In this paper I propose dynamic models for realized covariances, which explicitly take into account their measurement error. The models use the asymptotic distribution of the high-frequency estimates to allow for time-varying attenuation of the autoregressive parameters. The parameters are high on days where the covariance is estimated precisely and they are attenuated more heavily on days when the measurement error is large and the signal is weak. The models parsimoniously allow for different dynamics for each individual element of the covariance matrix. The concept was used in the simpler univariate setting by Bollerslev, Patton, and Quaedvlieg (2016), who found promising results. By studying the multivariate problem, I can consider the impact of dynamic attenuation in a number of economically relevant applications that are not possible in the univariate case.

I evaluate the models both statistically and economically, using a variety of out-of-sample methods. Firstly, I consider a standard statistical evaluation of the models, where I find that the models provide significantly better forecasts compared to their constant attenuation benchmarks. I show that the models work well, because the autoregressive parameters are tailored for the information content in past estimates. They allow for more responsive short-term dynamics when the signal is strong, and limit the propagation of potentially large estimation errors when the signal is weak. Using various loss functions I document improved forecast accuracy and find that improvements are made both after precise and after noisy covariance estimates, demonstrating that dynamic attenuation offers benefits in both directions.

Since measurement error has a large impact on economic decision making, the dynamic attenuation models have widespread practical applicability. I provide evidence of their empirical qualities by using them for several financial applications. In the first application I use the models for portfolio allocation. The problem of measurement error has received significant attention in this literature, as it often leads to extreme positions. The extreme portfolio weights are rarely the optimal allocation based on population values, but the result of inaccurate and poorly conditioned, estimated covariance matrices. The spurious nature of the positions leads to portfolios that are under-diversified, due to large weights in single stocks, and unstable, as the spurious positions have to be unwound in the next period (e.g. Li, 2015). This leads to poor out-of-sample performance, because of bad allocations and excessive turnover.

Two of the most popular techniques to reduce the impact of estimation error on portfolio allocation are to either ‘shrink’ the covariance matrix or to impose constraints on the portfolio weights. For shrinkage, the portfolio allocation is based on a linear combination of the covariance estimate and some target matrix that implies stable portfolios, such as the identity or an equicorrelation matrix (Ledoit and Wolf, 2003, 2004a). Another strategy is to constrain the portfolio weights directly, as the extreme positions are typically the result of estimation error, not of the population values. For instance, Jagannathan

## 5. (UN)RELIABLE REALIZED RISK MEASURES

---

and Ma (2003) study no short-sale global minimum variance portfolios, and DeMiguel, Garlappi, Nogales, and Uppal (2009) and Brodie, Daubechies, De Mol, Giannone, and Loris (2009) propose L-norm constrained portfolios.

The time-varying parameter models proposed in this paper endogenously provide a dynamic alternative to the shrinkage estimators. Covariance forecasts are shrunk, with time-varying intensity, from the conditional to the unconditional covariance matrix, which leads to less extreme and more stable portfolio allocations. I find that the dynamic adjustment of the parameter leads to more accurate (i.e. lower ex-post portfolio volatility) and more stable forecasts, which reduces turnover by about 30% compared to their respective benchmark models. This result is economically significant. Using the Fleming, Kirby, and Ostdiek (2003) framework I find that a risk averse investor would be willing to sacrifice up to 150 basis points annually to switch to a dynamic attenuation model and, using realistic values, typically gains an additional 30 basis points due to reduced transaction costs. The improvements persist after applying other methods to counter estimation error, such as imposing weight constraints. Moreover, I find that despite the higher level of transaction costs and increased noise in the estimates, portfolios that are rebalanced daily using the models proposed in this paper outperform portfolios which are rebalanced weekly or monthly, which is not the case for the base models.

Similar to how Realized (Co)variances allow effective exploitation of the information inherent in high-frequency data, they also allow for direct reduced-form modeling of the otherwise latent quantities. Indeed, in multivariate applications the covariance is seldom the object of interest and an important derived quantity is the realized beta (Andersen, Bollerslev, Diebold, and Wu, 2006; Hansen, Lunde, and Voev, 2014). In the second application of the models I consider the forecasting of risk exposures. The ex-post estimates of beta are similarly impacted by heteroskedastic measurement error. To obtain beta forecasts, I innovate by comparing forecasts of beta implied by the covariance matrix to beta forecasts obtained by directly modeling realized beta. I find that by taking into account the heteroskedasticity of the measurement error, betas appear to be more persistent than previously thought. I consider various hedging applications to evaluate the forecasts. I find that modeling beta directly using realized betas outperforms those implied by covariance forecasts and more importantly, that explicitly taking into account the time variation in estimation uncertainty of realized beta offers additional improvements in the forecasts.

The remainder of the paper is organized as follows. Section 5.2 introduces notation and the various models. In Section 5.3 I report the result of statistical evaluation of covariance forecasts. Sections 5.4 and 5.5 report the results of simulations and empirical application on the performance of the models proposed in this paper for the purpose of portfolio selection and beta forecasting respectively. Section 5.6 concludes.

## 5.2 Dynamic Modeling with Heteroskedastic Measurement Error

In this section I introduce the models for forecasting the conditional covariance matrix of daily returns. I build on popular models in the literature and extend the models to take into account the measurement error in realized covariances. I propose to model the dynamics of the covariance matrix as autoregressive processes, where the coefficients are allowed to vary throughout time as a function of the degree of measurement error, of which we can obtain estimates using the asymptotic theory of the high-frequency estimators. The responsiveness to recent observations is reduced when their measurement error is large and vice-versa.

The intuition for the models follows from the errors-in-variables literature. The time-series of the estimated (co)variances are the sum of two components, the latent quantity and measurement error. Since the latter is just noise, the series appear to be less persistent than they truly are, and the autoregressive parameters are attenuated towards zero. When measurement error is homoskedastic, the degree of attenuation is proportional to the measurement errors variance. However, if the measurement error is heteroskedastic, estimates will be attenuated based on the *average* measurement error variance. This results in attenuation which is never optimal. It is too strong when the past observation has little measurement error and too weak when it is estimated imprecisely. The dynamics of the estimated series can be more accurately modeled by taking into account the heteroskedastic nature of measurement error.

We know from the asymptotic theory of the realized covariance that its measurement error is in fact heteroskedastic, and moreover, we can feasibly estimate its magnitude every single day. I develop models that use these estimates to obtain daily varying parameters such that attenuation can be tailored specifically to each individual estimate. On some days there is sufficient information to allow for less attenuation, while on other days one might have to attenuate more dramatically than a constant parameter model would allow. By directly incorporating estimates of the asymptotic distribution, dynamic attenuation can be obtained in a robust, parsimonious, and easy to estimate manner.

In the following, I first develop some notation and provide intuition for the modeling choices I make, followed by the definition of the various models. As base models, I consider three popular models in the literature, the vech-HAR (Chiriac and Voev, 2011), the Exponentially Weighted Moving Average (EWMA) (used in, amongst others, Fleming, Kirby, and Ostdiek, 2003) and HEAVY (Noureddin, Shephard, and Sheppard, 2012) models. The former can easily be estimated using OLS and is potentially the most popular model in the literature. The latter two models are estimated using maximum likelihood techniques.<sup>1</sup>

---

<sup>1</sup>Pakel, Shephard, Sheppard, and Engle (2014) demonstrate that parameter estimates in large-dimensional covariance models may be severely biased with MLE estimation, they propose a technique

The purpose of using multiple models is not to run a horse-race between the various models, but to evaluate the relative performance of the models with dynamically attenuated parameters versus the ones with constant parameters. It also serves to highlight the fact that the principle is widely applicable. The same ideas could be applied to any other model using high-frequency estimates or where estimates of the measurement error variance are available.

### 5.2.1 Notation and Definitions

I assume the following log-price process

$$P(s) = \int_0^s \mu(u)du + \int_0^s \sigma(u)dW(u), \quad (5.1)$$

where  $W(u)$  is an  $N$ -dimensional vector of independent Brownian motions, with  $\sigma$  the volatility process such that  $\Sigma(u) = \sigma(u)\sigma(u)'$  is the spot covariance matrix.<sup>2</sup> The objective is to model the ex-post covariation for day  $t$ , called the Integrated Covariance, which for price-processes of this type equals

$$\Sigma_t = \int_{t-1}^t \Sigma(u)du. \quad (5.2)$$

This quantity is not directly observable, but can be consistently estimated with the use of high-frequency returns. Define  $r_{i,t} = P_{t-1+i\Delta} - P_{t-1+(i-1)\Delta}$  as the  $i$ -th return vector of day  $t$ . Returns for individual series  $n = 1, \dots, N$ , are denoted by  $r_{i,t}^{(n)}$ . We observe  $M = 1/\Delta$  uniformly spaced, intra-daily returns for each series. The vector of daily returns is obtained as  $r_t = \sum_{i=1}^M r_{i,t}$ .

The daily integrated covariance matrix,  $\Sigma_t$ , can be consistently estimated, denoted  $S_t$ , using a wide class of estimators. Although realized variances can be very accurately estimated at the 5-minute frequency, realized covariances suffer from the Epps (1979) effect, where correlation estimates tend to decrease when sampling at higher frequencies due to asynchronous trading and observation of the underlying price process. A number of alternatives to realized covariance exist that are robust to this effect and I use the Multivariate Kernel (MKern) of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) applied to 5-minute returns.<sup>3,4</sup>

---

known as composite likelihood, that significantly reduces the bias, and moreover greatly reduces computation time. For this paper I use their 2MSCLE technique. More details are provided in Appendix C.

<sup>2</sup>For simplicity I only consider a data generating process without jumps. The general principle behind the models in this paper readily extend to a price process that includes a jump-component.

<sup>3</sup>One disadvantage of using the MKern is that it is theoretically slightly biased. In practice, especially at this sampling frequency, this term is negligible, and I disregard it throughout the rest of the paper.

<sup>4</sup>Other robust choices include estimator of Hayashi and Yoshida (2005) and the pre-averaging estimator of Christensen, Kinnebrock, and Podolskij (2010), but I choose the MKern as its asymptotic variance can easily be estimated.

Throughout the paper, I will vectorize the covariance matrix, and consider the  $N^* = N(N + 1)/2$  dimensional vector  $\varsigma_t = \text{vech } \Sigma_t$ . The MKernel is an estimator of  $\Sigma_t$  and is therefore subject to estimation error,  $s_t = \varsigma_t + \eta_t$ , with  $s_t = \text{vech } S_t$ . The asymptotic theory of the MKernel provides us with the distribution of the measurement error  $\eta_t$ , which is mixed-normal with  $N^* \times N^*$  covariance matrix  $\Pi_t$ . The measurement error variance  $\Pi_t$  is proportional to a quantity called the Integrated Quarticity (IQ), defined as

$$IQ_t = \int_{t-1}^t \Sigma(u) \otimes \Sigma(u) du, \quad (5.3)$$

where  $\otimes$  denotes the kronecker product. The asymptotic variance of the variance, or IQ, can similarly be consistently estimated based on high-frequency data. Appendix A provides estimation details of the MKernel and its asymptotic distribution.

### 5.2.2 Vector Autoregressive Models with Measurement Error

The effect of measurement error on autoregressive models is extensively studied in the univariate case (e.g. Staudenmayer and Buonaccorsi, 2005). In the multivariate autoregressive setting the problem is less well-developed and typically looks at the effect of measurement error on identification and hypothesis testing (e.g Holtz-Eakin, Newey, and Rosen, 1988; Komunjer and Ng, 2014). It is however well-known that the estimates of autoregressive parameters are attenuated towards zero, both in uni- and multivariate case. In this section I review some of the results relevant for the forecasting of realized covariances.

In order to provide intuition for my modeling choices in the next sections, I assume a simple model for the integrated covariance and analyze the effect of measurement error. Specifically, I assume that  $\varsigma_t$  follows a VAR(1) model. Under this assumption the integrated covariance follows

$$\varsigma_t = \Phi_0 + \Phi_1 \varsigma_{t-1} + \epsilon_t, \quad (5.4)$$

with  $\Phi_0$  of dimension  $N^*$  and where  $\Phi_1$  is an  $N^* \times N^*$  matrix. In practice one would have to estimate the model on estimated quantities  $s_t = \varsigma_t + \eta_t$ ,<sup>5</sup>

$$s_t = \Theta_0 + \Theta_1 s_{t-1} + \epsilon_t. \quad (5.5)$$

Under the assumption that  $\epsilon_t$  and  $\eta_t$  are i.i.d. such that measurement error is homoskedastic and uncorrelated over time, it is straightforward to derive that the OLS estimate relates

---

<sup>5</sup>In the univariate case the population parameters of an AR( $p$ ) model with homoskedastic measurement error can be identified by estimating an ARMA( $p, p$ ) model. In the multivariate case the composite error term  $\epsilon_t - \eta_t + \Theta_1 \eta_{t-1}$  typically fails to be a finite order moving average process, and a VARMA will not identify the population parameter.

to the population value  $\Phi_1$  in the following way

$$\begin{aligned}\Theta_1 &= (\mathbf{s}'\mathbf{s})^{-1}(\boldsymbol{\varsigma}'\boldsymbol{\varsigma})\Phi_1 \\ &= (\boldsymbol{\varsigma}'\boldsymbol{\varsigma} + \boldsymbol{\eta}'\boldsymbol{\eta})^{-1}(\boldsymbol{\varsigma}'\boldsymbol{\varsigma})\Phi_1,\end{aligned}\tag{5.6}$$

where  $\mathbf{s} = (s_1, \dots, s_{T-1})$  is the  $T - 1 \times N^*$  matrix of lagged covariances, with  $\boldsymbol{\varsigma}$  and  $\boldsymbol{\eta}$  defined analogously.<sup>6</sup> The estimated parameter is the population parameter times the ratio of variation of the latent process over the variation of the estimated process, which is often called the reliability ratio. This result shows that in the general VAR(1) case, the attenuation in the multivariate case is more subtle than in the univariate case. Specifically, depending on the covariance structure of  $\boldsymbol{\eta}$ , certain parameters may be biased towards zero, while other parameters may be biased upwards. In contrast, in the univariate setting, parameters are always biased towards zero. Regardless of the direction, the degree of bias of the parameter is proportional to the degree of measurement error.

In practice, since the dimension of covariances, and therefore the parameter matrix, increases rapidly with the number of assets, it is customary in the covariance modeling literature to consider scalar models where the autoregressive parameter is restricted to be a scalar. One typically estimates the full dimension of the constant or uses covariance targeting for further parameter reduction. The scalar model removes the subtleties of the measurement error and its parameter's attenuation is proportional to the average of all measurement errors.<sup>7</sup>

The bias as quantified by (5.6), could be removed using estimates of the measurement error covariance matrix  $\boldsymbol{\eta}'\boldsymbol{\eta}$ , by simply reverse engineering the equation to retrieve  $\Phi_1$ . A similar result can be produced for the scalar model. Although this identifies the population parameter, which may often be of interest, this is not optimal for forecasting purposes, as attenuation serves a purpose by limiting the impact of noisy estimates.

When the measurement error is heteroskedastic, attenuation is proportional to the average measurement error variance. The attenuation serves the same purpose, but constant attenuation is not optimal. Equation (5.6) could similarly be used to obtain an optimally attenuated coefficient at each point in time by replacing the unconditional noise covariance matrix with an estimate of the conditional one. However, due to the difficulty of estimating the measurement error variance and the required inversion of the matrix, this will typically result in extremely poor and unstable parameter estimates.

In the following sections, I propose parsimonious dynamic specifications for the high-frequency covariance estimates that combine the features of the general VAR and re-

---

<sup>6</sup>In the univariate case this relationship simplifies to the easier to interpret  $\Theta_1 = \Phi_1(1 + \text{Var}(\eta_t)/\text{Var}(\varsigma_t))^{-1}$ , where the estimated parameter is the population parameter times a noise-to-signal ratio.

<sup>7</sup>To see this, note that to estimate the scalar model using OLS amounts to stacking the columns of the  $\mathbf{s}$  matrix, i.e.  $\mathbf{s}^* = \text{vec } \mathbf{s}$ . Hence, the scalar parameter has the same bias formula (5.6), where  $s, \varsigma$  and  $\eta$  are replaced with their stacked counterparts, and  $\boldsymbol{\eta}^*\boldsymbol{\eta}^*$  is simply the average measurement error variance.

stricted scalar models to deal with heteroskedastic measurement error in a robust manner. The autoregressive parameters are linear in the measurement error covariance matrix of the past estimates. This enables the model to allow for complicated and general dynamics with a small amount of parameters, where each element of the covariance matrix has different dynamic autoregressive parameters based on the degree of measurement error of individual series on each individual day.

### 5.2.3 HAR Models

The widely popular Heterogeneous Autoregressive (HAR) model of Corsi (2009) was first extended to the multivariate setting in Chiriac and Voev (2011) who apply the HAR approach to the vech of the covariance matrix, defined as

$$s_t = \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t, \quad (5.7)$$

where  $s_{t-h|t-1} = \frac{1}{h} \sum_{i=1}^h s_{t-i}$ .  $\theta_0$  is a  $N^*$  dimensional vector and the other  $\theta$ s are scalar. One could allow for more general specifications, by letting the different  $\theta$  coefficients be (possibly restricted)  $N^* \times N^*$  matrices. However, the more general specifications greatly complicate the model, and moreover, their forecasts are not guaranteed to be positive-definite. By considering scalar coefficients, the forecasts are sums of positive definite matrices and as such positive definite.

One way to allow for a more general specification whilst ensuring positive definite forecasts is to model the variances and correlations separately, as proposed in Oh and Patton (2015). I refer to this version as the HAR-DRD, where *DRD* refers to the decomposition of  $S$  into variances and correlations

$$S_t = D_t R_t D_t, \quad (5.8)$$

with  $D_t$  a diagonal matrix of standard deviations and  $R_t$  the correlation matrix. Each variance can be modeled using a univariate HAR, and the correlation matrix  $r_t = \text{vech } R_t$  can be modeled using the scalar HAR of Equation (5.7). I then recombine the forecasted correlation matrix with the forecasted variances to obtain the covariance forecasts. These forecasts are positive-definite as long as the forecasted variances are positive.

In order to exploit the heteroskedastic measurement error for forecasting purposes, Bollerslev, Patton, and Quaadvlieg (2016) propose a simple modification to the univariate HAR, which is Equation (5.7) with scalar  $s_t$ , by replacing the parameter for the daily lag by

$$\theta_{1,t} = (\theta_1 + \theta_{1Q} \Pi_t^{1/2}), \quad (5.9)$$

where  $\Pi_t$  is an estimate of twice the Integrated Quarticity, the asymptotic variance of the univariate Realized Variance. The specification can be seen as a linear approximation to the inverse of (5.6) in the univariate setting, which is more robust to the estimation error in

## 5. (UN)RELIABLE REALIZED RISK MEASURES

---

$\Pi_t$ . They call this model the HARQ, where the Q suffix refers to the Integrated Quarticity. One could also consider time-varying parameters for the weekly and monthly lag, but Bollerslev, Patton, and Quaedvlieg (2016) find that, due to the difficulty of estimating  $\Pi_t$ , these tend to add more noise than dynamic improvements, as the measurement error and its heteroskedasticity are significantly reduced when averaging over five or twenty-two days.

There are many potential ways of defining a multivariate analogue. Although measurement error in  $s_t$  is expected to be of greater importance in the multivariate setting, it is even more difficult to accurately estimate its distribution  $\Pi_t$ , which restricts the choice of specification significantly. It turns out that estimates of the full  $\Pi_t$  matrix are too noisy to successfully be incorporated in the models. Its estimation error leads to attenuation in the dynamics of the autoregressive parameter itself, which then essentially remains constant. Improvements can be made by restricting the way  $\Pi_t$  is incorporated in the dynamic specification.

One way too extract useful information from  $\Pi_t$  is to only use a subset of its elements. Similar to the covariance matrix  $s_t$  itself, the variance elements of  $\Pi_t$  are much more accurately estimated than the covariance elements. Hence, for robustness purposes, I choose to define  $\pi_t = \sqrt{\text{diag}(\Pi_t)}$ , i.e. the matrix that has the asymptotic standard deviation of each individual element of the covariance matrix on its diagonal.<sup>8</sup> I define the vech HARQ as

$$\begin{aligned} s_t &= \theta_0 + \theta_{1,t} \circ s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t \\ \theta_{1,t} &= \theta_{1\iota} + \theta_{1Q} \pi_{t-1}, \end{aligned} \tag{5.10}$$

where  $\circ$  denotes the Hadamard product and  $\iota$  an  $N^*$  dimensional vector of ones. In the HARQ, as in the HAR,  $\theta_1$  and  $\theta_{1Q}$  are scalar. However, as  $\pi_t$  is a vector,  $\theta_{1,t}$  is a vector as well, and the different elements of  $s_t$  obtain different degrees of persistence. The  $\theta_{1,t}$  parameter vector thus allows for different dynamics for each individual element of the covariance matrix based on their relative measurement error variance using only two estimated parameters,  $\theta_1$  and  $\theta_{1Q}$ . One could easily allow for different, and possibly more general, dynamics by incorporating  $\Pi_t$  in a less restrictive way, but I found this specification of  $\pi_t$  to work best empirically, as it balances the information content on the heteroskedasticity with the estimation error in  $\Pi_t$ .

For the HARQ-DRD, I simply model the variances using the univariate HARQ of Bollerslev, Patton, and Quaedvlieg (2016) as described in Equation (5.9). I use the normal scalar HAR for the correlations. One could also consider a HARQ specification on the correlations, but I refrain from a dynamic specification for the parameter as the

---

<sup>8</sup>In practice I demean  $\pi_t$  to make the  $\theta_1$  coefficient interpretable as the value at average measurement error.

heteroskedasticity in the measurement error of the correlations is limited.<sup>9,10</sup> The potential benefits of the marginal variability in the parameters would be outweighed by the estimation noise in  $\pi_t$ .

The main advantage of the HAR(Q) and HAR(Q)-DRD is that their estimation is very straightforward as the models can be easily estimated via OLS. For the HAR-DRD this is a 2-step estimation where the univariate models and the correlation model are separated.

### 5.2.4 EWMA Filters

Another popular method for forecasting covariance matrices which incorporates realized covariances is the Exponentially Weighted Moving Average (EWMA) used by, amongst others, Fleming, Kirby, and Ostdiek (2003). Contrary to the HAR model that directly models the realized covariance, the EWMA, and the HEAVY model discussed in the next subsection, forecast the covariance matrix of daily returns, by incorporating realized information.

They consider an exponentially weighted moving average representation for the daily covariance matrix  $V_t = E(r_t r_t' | \mathcal{F}_{t-1})$ , with  $\mathcal{F}_{t-1}$  the information set available at time  $t - 1$ , and incorporate Realized Covariance in their rolling estimator to improve forecasts. The specification they propose is

$$v_t = \exp(-\alpha)v_{t-1} + \alpha \exp(-\alpha)s_{t-1}, \quad (5.11)$$

with  $v_t = \text{vech } V_t$ . Here  $\alpha$  captures the decay rate of the lags. When  $\alpha$  is low the process is very persistent and when  $\alpha$  is high more of the new information coming from  $s_{t-1}$  is incorporated in the fitted values for  $v_t$ . The model guarantees positive-definite forecasts, which is essential for many applications. I estimate the optimal decay rate under the assumption that daily returns are conditionally normal using composite maximum likelihood as described in Appendix C.

Similar to the HARQ, I propose to let the parameter  $\alpha$  vary with estimates of the uncertainty in  $s_t$ . Specifically, the EWMAQ is defined as

$$\begin{aligned} v_t &= \exp(-\alpha_t) \circ v_{t-1} + \alpha_t \circ \exp(-\alpha_t) \circ s_{t-1} \\ \alpha_t &= (\alpha_l + \alpha_Q \pi_{t-1}). \end{aligned} \quad (5.12)$$

Typically, the  $\alpha$  parameter in the EWMAQ will be higher than in the EWMA, allowing for more short term dynamics stemming from  $s_{t-1}$ , with  $\alpha_Q$  negative such that the impact of highly uncertain estimates of the covariance are not incorporated in  $v_t$ . Instead, if  $s_t$

<sup>9</sup>Using Proposition 3 in Barndorff-Nielsen and Shephard (2004a), and assuming constant spot volatility, the correlation coefficient between two assets has asymptotic standard deviation of approximately  $1 - \rho_t^2$ . The heteroskedasticity of measurement error therefore only depends on the variability of the square of the correlation which is by definition limited compared to (co)variances, since  $|\rho_t| \leq 1$ .

<sup>10</sup>The asymptotic variance of the correlation matrix can be derived more generally using the delta method, i.e.  $\tilde{\Pi}_t = \nabla \Pi_t \nabla'$ , where  $\nabla = \frac{\partial \text{vech } D^{-1} \Sigma D^{-1}}{\partial \text{vech } \Sigma}$ .

is estimated imprecisely, the model shifts weight towards the weighted average of past covariances  $v_{t-1}$ .

### 5.2.5 HEAVY Models

The HEAVY model was introduced in Noureldin, Shephard, and Sheppard (2012). The HEAVY model is similar to the EWMA in that it models  $v_t$  by incorporating high-frequency covariance estimates, but is more richly parametrized. Similarly, the HEAVY model allows for short response times to changes in volatility by incorporating new information from the high-frequency estimates. Here, I introduce the HEAVYQ, which again differentiates the response to the high-frequency estimate by incorporating the strength of the signal of  $s_t$  as measured by its asymptotic variance.

In this paper I consider the most parsimonious HEAVY specification, which is the scalar version with covariance targeting, but the principle can be applied to any HEAVY specification.

$$v_t = (I_{d^*} - B - A\kappa)\lambda_V + Bv_{t-1} + As_{t-1}, \quad (5.13)$$

where A and B are scalar. Denote the unconditional expectation of the high-frequency covariance matrix by  $\Lambda_\Sigma$  and the unconditional expectation of the low-frequency daily covariance matrix by  $\Lambda_V$ .  $\kappa = L_N(\bar{\kappa} \otimes \bar{\kappa})D_N$ , with  $\bar{\kappa} = \Lambda_\Sigma^{1/2}\Lambda_V^{-1/2}$  and  $L_N$  and  $D_N$  the elimination and duplication matrices respectively (Magnus and Neudecker, 1980). The  $\kappa$  term is an adjustment to match the expectation of the high-frequency covariance with the daily covariance. Finally,  $\lambda_V = \text{vech } \Lambda_V$ . Again, the models are estimated using composite likelihood.

Like the previous two models, the HEAVYQ model simply allows the impact of  $s_{t-1}$  to vary over time with the degree of measurement error and is defined as

$$\begin{aligned} v_t &= (I_{d^*} - B - A_t\kappa)\lambda_V + Bv_{t-1} + A_t \circ s_{t-1} \\ A_t &= (A_t + A_Q\pi_{t-1}), \end{aligned} \quad (5.14)$$

where again  $A_Q$  is scalar.

## 5.3 Forecasting the Covariance Matrix

In this section I evaluate the empirical qualities of the Q models using statistical loss functions. Section 5.3.2 provides in-sample estimates and fit, and Section 5.3.3 presents the results of an out-of-sample model evaluation.

### 5.3.1 Data

The empirical investigation focuses on the 27 Dow Jones constituents as of September 20, 2013 that traded continuously from the start to the end of the sample. I use 5-minute

returns retrieved from the TAQ database over the period February 1993 until December 2013, for a total of 5,267 daily observations. For the purpose of constructing betas, I supplement the individual stocks with high-frequency counterparts of the four risk factors of Fama and French (1993) and Carhart (1997). The market is proxied by the S&P500 ETF, the SPY. The other three factors, Small-minus-Big (SMB), High-minus-Low (HML) and Momentum (MOM), are constructed as the difference between the returns of the highest and lowest decile portfolios of the full TAQ database.<sup>11</sup> These risk factors are used in the two empirical applications presented in Sections 5.4 and 5.5. In the former I use them to obtain portfolios to track the market index and in the latter I forecast the exposures to the four risk factors.

The analysis focuses on open-to-close covariance matrices, where the noisy overnight returns are not included.<sup>12</sup> This is often done in the high-frequency covariance literature, even in the context of portfolio allocation (De Lira Salvatierra and Patton, 2015; Hautsch, Kyj, and Malec, 2015; Lunde, Shephard, and Sheppard, 2015). Moreover, Liu (2009) finds that including the overnight return in covariance estimation leads to greater out-of-sample portfolio turnover. Omitting the overnight return is in line with Andersen, Bollerslev, Frederiksen, and Nielsen (2010) who treat overnight returns as deterministic jumps, which have a transitory effect in the dynamics of variances. Similarly, I measure the vector of daily returns using open-to-close returns.

I provide some descriptives of the data in Appendix B. In the remainder of this Section, as well as in Section 5.4, I analyze a subset of ten of the most liquid stocks to keep estimation feasible. In order to effectively model covariance matrices of higher dimension more structure needs to be applied to the model, which I leave for future research. The subset consists of stocks with the following tickers: AXP, BA, CVX, DD, GE, IBM, JPM, KO, MSFT and XOM. In Section 5.5, I separately model the covariance matrix of each of the 27 stocks with the four risk-factors, whose limited dimension remains feasible.

### 5.3.2 In-Sample Results

First consider the in-sample estimation results for the ten stock sub-sample. Parameters with robust standard errors are reported in Table 5.1 for each of the models. For the

---

<sup>11</sup>As the construction of the high-frequency risk-factors slightly different from the standard method, their daily returns do not perfectly match the factors typically used in empirical work. However, the correlation between the high-frequency factors and the daily factors obtained from Kenneth French's website is between 70% and 90%, which is mainly due to the exclusion of the overnight return. When included the correlation increases to values between 90% and 95%.

<sup>12</sup>The single overnight return covers seventeen and a half hours compared to the five-minute returns during trading hours. As such, the estimate of the close-to-open variation is very noisy. Moreover, the estimation of the asymptotic variance for this part of the daily variation is equally poor. In order to incorporate the overnight return, one could use an auxiliary specification to predict the overnight variation and open-to-close variation separately. For the overnight variation, a Q model is unlikely to provide great benefits due to the noise in its asymptotic variance estimates. Another possibility to deal with the overnight return is to scale up the intraday variation to reflect the variation for the whole day, similar to the  $\kappa$  adjustment term in the HEAVY model.

## 5. (UN)RELIABLE REALIZED RISK MEASURES

---

HAR(Q)-DRD I report the average of the parameters and standard errors over each of the individual variance specifications to save space. Finally, in the bottom two rows I report the Frobenius distance and QLIKE loss of the fitted values  $H_t$  with respect to the ex-post covariance  $S_t$ . The Frobenius distance is defined as

$$L_t^{Frobenius} = \sqrt{\text{Tr}(H_t - S_t)}, \quad (5.15)$$

and the quasi-likelihood (QLIKE) as

$$L_t^{QLIKE} = \log |H_t| + \text{Tr}(H_t^{-1}S_t), \quad (5.16)$$

which is the negative of the log-likelihood of a multivariate normal density. Both functions measure loss, and hence lower values are preferable.<sup>13</sup>

As expected, all the Q coefficients are negative and significant, which is consistent with the intuition that as the measurement error variance  $\pi_t$  increases, the informativeness of the past covariance estimate decreases, and the model endogenously increases the degree of attenuation. To put the magnitude of the parameters in perspective, the cross-sectional average of the composite parameter  $\theta_{1,t}$  varies between roughly 0.21 and 0.87 for the HARQ model.

As in Bollerslev, Patton, and Quaadvlieg (2016) for the HAR there is a distinct redistribution of weight from the monthly lag to the daily lag. This effect is more pronounced in the HAR-DRD model, due to its increased flexibility in the time-variation of parameters. For the HARQ-DRD all individual variances have separate dynamics, and they can allow for individual sensitivities to the degree of measurement error. As such the average increase in  $\theta_1$  compared to the standard models is greater for the HARQ-DRD than for the HARQ, and similarly the  $\theta_{1Q}$  parameter is greater in absolute value. This in itself is an attenuation story, as the measurement error in  $\pi_t$  aggregates in the cross-section for the HARQ, making the dynamic parameter less effective. Relative to the EWMA, the EWMAQ has a higher value of  $\alpha$  which shows it allows for a greater impact of the MK-ernel estimates on the sample path of  $v_t$ . However, given the negative value of  $\alpha_Q$ , when estimation error becomes greater the weight is shifted towards the long-run weighted average away from the noisy estimate. The table displays a similar picture for the HEAVYQ, where the  $A$  parameter of the HEAVYQ is higher than that of the HEAVY model, and high estimation uncertainty in  $s_{t-1}$  decreases the responsiveness to the high-frequency estimate in favor of the long term weighted average  $v_{t-1}$ .

---

<sup>13</sup>Both loss functions offer consistent rankings of models, despite the use of an ex-post estimate as a proxy for the true covariance matrix when evaluating the forecasts (Laurent, Rombouts, and Violante, 2013; Patton, 2011b).

**Table 5.1:** In-Sample Estimates

	HAR	HARQ	HAR- DRD	HARQ- DRD		EWMA	EWMAQ		HEAVY	HEAVYQ	
Average Variances											
$\theta_1$	0.247 (0.040)	0.541 (0.040)	$\theta_1$	0.260 (0.064)	0.660 (0.065)	$\alpha$	0.098 (0.005)	0.118 (0.005)	$A$	0.106 (0.009)	0.148 (0.009)
$\theta_2$	0.410 (0.038)	0.333 (0.038)	$\theta_2$	0.395 (0.016)	0.223 (0.022)	$\alpha_Q$		- 0.007 (0.003)	$B$	0.876 (0.004)	0.825 (0.004)
$\theta_3$	0.244 (0.038)	0.113 (0.038)	$\theta_3$	0.240 (0.028)	0.101 (0.028)				$A_Q$		- 0.026 (0.012)
$\theta_{1Q}$		- 0.043 (0.018)	$\theta_{1Q}$		- 0.067 (0.026)						
Correlation											
			$\theta_1$	0.049 (0.007)	0.049 (0.007)						
			$\theta_2$	0.159 (0.003)	0.159 (0.003)						
			$\theta_3$	0.560 (0.003)	0.560 (0.003)						
Frob.	12.282	12.190		11.987	11.952		12.295	12.242		12.539	12.532
QLIKE	15.484	15.392		15.289	15.205		15.439	15.441		15.517	15.505

*Note:* The table provides in-sample parameter estimates and measures of fit for the various benchmark and Q models. For the HAR-DRD the reported parameter estimates and standard errors for the univariate variance specifications are the average across the ten assets. The bottom panel shows the in-sample fit of the various models as measured by the Frobenius distance and QLIKE.

### 5.3.3 Out-of-Sample Results

In this section I consider out-of-sample forecasts of the covariance matrix derived from the various models. I focus on one-day-ahead forecasts of the ten-dimensional covariance matrix. For the forecasts, the parameters of the different models are re-estimated each day using a rolling window sample of 1,000 days.<sup>14</sup> The forecasts are evaluated based on the Frobenius distance and QLIKE defined in Equations (5.15) and (5.16) respectively, where the fitted value,  $H_t$ , is replaced with the forecasts based on the  $t - 1$  information set, denoted  $H_{t|t-1}$ .

Table 5.2 presents the forecasting results for the different models. The Q models produce improvements on their standard models for both loss functions. Unreported results show that both the standard and Q models forecast the right covariance matrix on average, with average bias of each element close to zero, but the Q models make great

<sup>14</sup>Similar to the univariate case in Bollerslev, Patton, and Quaedvlieg (2016), due to the sometimes erratic behavior of the difficult to estimate  $\pi_t$ , the HARQ and HARQ-DRD may sometimes produce negative definite forecasts. To ensure positive-definite forecasts, I apply the same insanity filter as Bollerslev, Patton, and Quaedvlieg (2016) and replace the entire covariance forecast with the time-series average of the estimation sample if the forecast is non positive-semidefinite. This happens for one or two forecasts per series throughout the empirical section.

## 5. (UN)RELIABLE REALIZED RISK MEASURES

**Table 5.2:** Out-of-Sample Forecast Results

	Full Sample		Lower 95% $\ \pi_t\ $		Upper 5% $\ \pi_t\ $	
	Frob.	QLIKE	Frob.	QLIKE	Frob.	QLIKE
HAR	12.305	14.382	9.274	13.275	<b>69.723</b>	32.651
HARQ	<b>12.107*</b>	14.159*	9.161*	13.206*	<b>69.703</b>	<b>32.190*</b>
HAR-DRD	<b>12.134</b>	14.140	<b>8.940</b>	<b>13.022</b>	70.467	32.603
HARQ-DRD	<b>11.976*</b>	<b>13.896*</b>	<b>8.888*</b>	<b>12.990*</b>	<b>70.056*</b>	<b>31.050*</b>
EWMA	12.379	14.105	9.303	13.109	71.104	32.686
EWMAQ	<b>12.178*</b>	<b>14.091*</b>	9.172*	13.122	70.633*	32.643*
HEAVY	12.473	<b>14.051</b>	9.299	13.041	71.789	32.263
HEAVYQ	<b>12.161*</b>	<b>14.004*</b>	<b>9.093*</b>	13.050	<b>70.170*</b>	32.258

*Note:* This table provides the loss of out-of-sample forecasts for the various models. The full-sample results are split up in days where the measurement error variance was low, where  $\|\pi_t\|$  was in the lower 95%, and when estimation error was severe, the upper 5%. Entries in bold depict the models that are part of the 90% Model Confidence Set for each column. Q models that significantly improve relative to their non-Q benchmark are indicated by an asterisk.

improvements in terms of the variance of the forecast error, which is significantly reduced.

To better understand whether most of the improvements are coming from the increased short-term dynamics when the estimates are precise or the increased attenuation when the estimates are poorly estimated, I split the results up in the 95% of days with lowest measurement error variance and the 5% highest. Since the asymptotic variance is a large-dimensional matrix, I summarize the total uncertainty in  $s_t$  by means of the Frobenius norm of  $\pi_t$ . For all models, the forecasts are more accurate both when the high-frequency estimates are precise and when they are imprecise. However, the largest percentage improvements appear to stem from the increased responsiveness when the estimates are precise.

To test whether the quality of the forecasts differ significantly across models, I apply the Model Confidence Set (MCS) of Hansen, Lunde, and Nason (2011). This procedure determines the (sub)set of models that contains the best forecasting model with 90% confidence. For both the full series, as well as the two uncertainty-split samples, I compare each Q model with its basemodel separately (significance indicated by \*), as well as all eight forecasts jointly (members of the set indicated in bold). For the former, the MCS procedure finds that each of the Q models significantly outperforms their respective benchmark in the full sample based on either loss function. For the two sub-sample split based on the degree of uncertainty, the Q models fail to significantly improve for a small set of models. Overall, this provides significant evidence that the dynamic attenuation models produce better forecasts.

The results of the MCS applied to all eight models simultaneously are reported in Table 5.2. The models that are part of the MCS are highlighted in bold. The only model that is consistently in the confidence set is the HARQ-DRD. In the full sample, the EWMAQ and HEAVYQ are also in the MCS based on either loss-function. The HARQ, on the other hand, appears to work particularly well when there is high uncertainty in

covariance estimates. Overall, the results of the eight-model MCS show that the Q models jointly outperform the non-Q models from a statistical point of view. In the next sections I evaluate their relative merit from a practical point of view by using them in a dynamic portfolio allocation setting and for beta forecasting.

## 5.4 Global Minimum Variance Portfolios

In this section the empirical qualities of the various models are evaluated by using their forecasts to obtain Global Minimum Variance (GMV) portfolios. The GMV portfolio is a popular tool to evaluate the accuracy of covariance forecasts as its optimal weights are a function of the covariance matrix only. Any other portfolio application would also depend on the estimation error of the mean. Moreover, Jagannathan and Ma (2003) and DeMiguel, Garlappi, Nogales, and Uppal (2009) demonstrate that mean-variance efficient portfolios do not perform as well as the global minimum variance portfolios in terms of out-of-sample Sharpe ratios, exactly because of the noisiness of the mean estimates. There is a large literature based on the performance of various methods for obtaining the covariance matrix for portfolio allocation using daily data (Brodie, Daubechies, De Mol, Giannone, and Loris, 2009; Chan, Karceski, and Lakonishok, 1999; Fan, Fan, and Lv, 2008; Jagannathan and Ma, 2003), and more recently the benefits of using high-frequency data for portfolio allocations have been extensively researched (Bandi, Russell, and Zhu, 2008; Fleming, Kirby, and Ostdiek, 2003; Hautsch, Kyj, and Malec, 2015; Liu, 2009; Pooter, Martens, and Dijk, 2008; Varneskov and Voev, 2013). Without exception, the papers find that integrating high-frequency data leads to superior out-of-sample portfolio performance. However, these papers typically focus on estimation considerations, such as which sampling frequency or estimator to use. Instead, in this application, I take the estimate as given, and compare different forecasting methods.

Implementation of the GMV portfolio remains difficult despite all the econometric advances and data availability. Covariances vary over time, and are subject to significant forecasting and measurement error. The former leads to the optimal portfolio allocation shifting over time, and the latter often leads to extreme allocations. Indeed, turnover in portfolios arises from two sources. Part of turnover is fundamental and induced by the time-variation in covariances, and part of this turnover is spurious, induced by forecast and measurement error.

To counter the effect of estimation error, a technique called *shrinkage* is often applied. Optimal allocations are based on a convex combination of the estimated, or forecasted, covariance matrix, and some target matrix. The target is often taken to be an equicorrelation matrix or some factor-model based matrix (Ledoit and Wolf, 2003, 2004b). The degree to which they are shrunk is typically based on some optimality result that shows that the shrinkage should be stronger when the estimation uncertainty is greater. The Q models directly incorporate this same intuition and endogenously offer an empirically

more realistic *conditional* shrinkage, making them ideally suited towards the goal of portfolio allocation.

Indeed, by allowing for time-variation in parameters as a function of measurement error, the models produce better forecasts when the signal is strong, and endogenously shrink the forecasts towards a long-term average when the signal is weak. As such, not only do they reduce the spurious turnover, but they also more accurately model the fundamental turnover when possible. To illustrate this I conduct a small simulation where I can distinguish between fundamental and spurious turnover, and an empirical application where I show the significant improvements in out-of-sample performance of the Q models.

### 5.4.1 Setup

I consider a risk-averse investor who allocates funds into  $N$  risky assets and uses conditional mean-variance analysis to make daily allocation decisions. Since the models described in this paper mainly improve on their basemodels by modeling the short-term dynamics, the merits of the Q model are better highlighted on the daily horizon than on the slower weekly or monthly horizons. Moreover, daily rebalancing is an often studied and important problem, and it is the ideal frequency for volatility-timing strategies (e.g. Fan, Li, and Yu, 2012; Fleming, Kirby, and Ostdiek, 2003).

The portfolio is rebalanced using forecasts of the daily covariance matrix. To minimize conditional volatility the investor solves the global minimum variance portfolio:

$$\begin{aligned} w_t &= \arg \min w_t' H_{t|t-1} w_t \\ \text{s.t. } & w_t' \iota = 1, \end{aligned} \tag{5.17}$$

where  $\iota$  is a vector of ones of appropriate size. The solution to this problem is the weight vector

$$w_t = \frac{H_{t|t-1}^{-1} \iota}{\iota' H_{t|t-1}^{-1} \iota}. \tag{5.18}$$

whose elements are denoted by  $w_t^{(n)}$ , which is the allocation to asset  $n$ .

One feature of the Q models is that they should lead to more stable forecasts. Hence, of particular interest for the forecasting models developed in this paper is the distinction between fundamental and spurious turnover, and the effect on portfolio performance net of transaction costs. To evaluate the impact of the turnover implied by the different models I assume a fixed transaction cost  $c$  on each dollar traded for any stock. Total transaction costs are proportional to the turnover rates of the portfolios.

I use a standard set-up for transaction costs due to turnover, used in amongst others Han (2006), Liu (2009) and DeMiguel, Nogales, and Uppal (2014). At day  $t$ , the investor has portfolio weights  $w_t^{(n)}$  in stock  $n$ . Before re-balancing, the return on stock  $n$  has pushed the size of the investment in stock  $n$  to  $w_t^{(n)}(1 + r_t^{(n)})$ , and the total value of the

portfolio has increased by a factor  $1 + w'_t r_t$ . As such, the actual weight of stock  $n$  at the end of the day equals  $w_t^{(n)} \frac{1+r_t^{(n)}}{1+w'_t r_t}$ .

Taking into account the shifts in portfolio composition on the relative portfolio weights, the total turnover of the portfolio is equal to

$$TO_t = \sum_n^N \left| w_{t+1}^{(n)} - w_t^{(n)} \frac{1 + r_t^{(n)}}{1 + w'_t r_t} \right|. \quad (5.19)$$

Total transactions are assumed to be proportional to the turnover, and defined as  $cTO_t$ , where  $c$  is a transaction cost on each traded dollar. Portfolio excess returns net of transaction cost are given by  $r_{pt} = w'_t r_t - cTO_t$ . I consider values for  $c$  ranging from 0 to 2% chosen to cover reasonable transaction cost levels for the recent sample, in line with amongst others Fleming, Kirby, and Ostdiek (2003) and Brown and Smith (2011).

The other main quality of the models is that they do not propagate highly uncertain covariances which likely have large estimation error. Since estimation error leads to extreme portfolio allocations, particularly if the matrix is ill-conditioned, I also report details on the composition of the portfolios. To summarize these features, I report the portfolio concentration

$$PC_t = \left( \sum_n w_t^{(n)2} \right)^{1/2} \quad (5.20)$$

and the total portfolio short position

$$PS_t = \sum_n w_t^{(n)} \mathbb{I}_{\{w_t^{(n)} < 0\}}. \quad (5.21)$$

Finally, in order to evaluate the economic significance of the forecasting models, I use the framework of Fleming, Kirby, and Ostdiek (2001, 2003), who consider an investor with quadratic utility, who places a fixed amount of wealth in the set of assets each day. The realized daily utility generated by the portfolio for model  $k$ ,

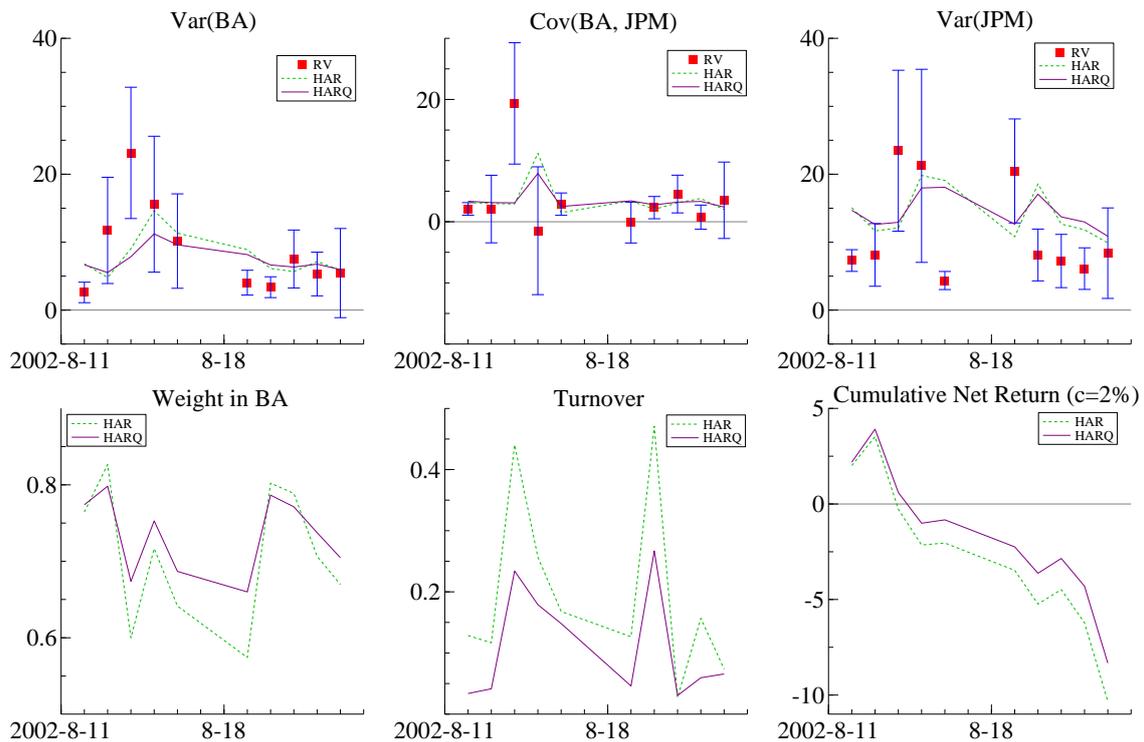
$$U(r_{pt}^k, \gamma) = (1 + r_{pt}^k) - \frac{\gamma}{2(1 + \gamma)} (1 + r_{pt}^k)^2.$$

To determine the economic value of switching from one model to another I solve for  $\Delta_\gamma$  in

$$\sum_{t=1}^T U(r_{pt}^k, \gamma) = \sum_{t=1}^T U(r_{pt}^l - \Delta_\gamma, \gamma), \quad (5.22)$$

where  $\Delta_\gamma$  can be interpreted as the return an investor with risk aversion  $\gamma$  would be willing to sacrifice to switch from model  $k$  to model  $l$ .

Figure 5.1: Dynamic Shrinkage



*Note:* The top row of panels shows the estimated (co)variance for BA and JPM, along with the forecasted values of the HAR and HARQ model. The bottom left panel shows the weights in the first asset for the HAR and HARQ model, the middle panel shows the turnover, and the right panel the cumulative portfolio return net of transaction costs over the two weeks.

#### 5.4.2 HARQ for portfolio allocation

First I provide some intuition on how the models work by means of an illustrative example using the HAR and HARQ. I consider a simple bivariate example (using BA and JPM) to better visualise the dynamics of the models. I plot the (co)variance and portfolio dynamics of both models over a two week period during the dot-com bubble in Figure 5.1. The top row shows ex-post estimated (co)variances along with 95% confidence bounds. In the same graphs, I show the ex-ante forecasts for the same dates based on the  $t-1$  information set. What is evident from the top row is that, in line with intuition, the HARQ is less responsive to new information when estimation uncertainty in those estimates is large. As a result the forecasts are more stable as large uncertainty is often accompanied by transitory spikes in volatility. This leads to a more stable portfolio allocation, as evidenced by two of the bottom graphs which display the portfolio weight in one of the assets, and the implied turnover. The allocations of the HAR and HARQ do not differ substantially, as both models on average have about a 70% allocation to BA, but the variability around this weight is much lower for the HARQ, which leads to uniformly lower turnover. The

reduction in transaction costs is significant, as shown by the bottom right graph, which displays the cumulative net return over the two weeks based on  $c = 2\%$  transaction costs. The daily return on the HARQ portfolio does not differ much from those of the HAR, as the portfolio composition is similar, but the reduced turnover ensures that the daily net return on the HARQ portfolio is higher every single day. Over the two week period this adds up to a significantly higher return on the HARQ portfolio. While the daily improvements may appear to be small, the empirical analyses will show that they amount to up to 70 basis points annually.

### 5.4.3 Simulation Results: Fundamental and Spurious Turnover

In the previous sections I argued that the Q models are ideally suited to the purpose of portfolio management. The better forecasts of the HARQ should lead to portfolio allocations closer to the fundamental weights implied by the true covariance matrix and hence lower ex-post portfolio volatility. Second, the more stable forecasts that the Q models obtain by reducing the impact of noisy estimates should lead to a more stable portfolio allocation and a reduction in turnover.

To analyze these predictions, this section presents the results from a small simulation study. I consider an out-of-sample conditional portfolio allocation, where I make one-day-ahead predictions of the covariance matrix using the HAR and HARQ. I then compare the estimated GMV portfolio with the fundamental one based on the population value in terms of the portfolio composition, ex-post variance and degree of turnover.

In order to simulate realistic sample paths of the covariance, for each iteration I sample a random block of 2,000 consecutive estimated covariance matrices from five random series of the dataset outlined in the previous section. On each day I simulate returns at one-second frequency with a diurnal volatility pattern (Andersen, Dobrev, and Schaumburg, 2012), such that the Integrate Covariance of the simulated data matches that of the empirical estimate.<sup>15</sup> I then aggregate the paths to 1, 5 and 15-minute returns ( $M = 390, 78, 26$ ), and estimate the MKernel and its asymptotic distribution. The different sampling frequencies allow me to attribute the gains to the HARQ's ability to better cope with the heteroskedastic measurement error, as the overall level of measurement error is decreasing in sampling frequency. To set a base-level of performance, I also consider HAR forecasts using the true population covariances ( $M = \infty$ ), which I will denote  $\text{HAR}_\infty$ . In this case there is no measurement error, only forecast error. By comparing the results based on estimated quantities we can directly infer the effect of estimation error on portfolio allocation.

As in the previous section I use a rolling window of 1,000 observations to obtain one-step-ahead forecasts. Each day, I obtain the GMV portfolio based on the forecasts and

---

<sup>15</sup>Although the simulated process is clearly oversimplified (i.e. no stochastic volatility), the diurnal pattern ensures that the asymptotic variance of the estimates is not simply their square, as would be the case if volatility were constant.

analyze their properties over the out-of-sample period. I consider the portfolio turnover as in (5.19) and the average portfolio standard deviation based on the population covariance  $\sqrt{\hat{w}_t \Sigma_t \hat{w}_t'}$ . To compare the estimated portfolio composition with the fundamental one, I report the average distance of the portfolio weights with respect to those obtained from the population value  $\sqrt{\sum_n (\hat{w}_t^{(n)} - w_t^{(n)})^2}$ . Finally, to demonstrate that the forecasts of the HARQ are truly shrunk based on the degree of measurement error, I also show the distance of the portfolio weights to those implied by the HAR $_{\infty}$ . Results are based on 1,000 simulations.

The results of the simulation are reported in Table 5.3. First consider the results for the HAR model. The column marked  $\infty$  directly models the latent population covariances, while the other columns use estimates based on  $M$  intraday observations. Note that even if the true covariance is observed, there will still be forecast errors, but it sets a baseline for the HAR's performance with estimated covariances. As expected, the estimation error in the MKernel has large consequences on the quality of the forecasts. When estimates are based on just 26 observations, the portfolio performance is significantly reduced compared to the HAR $_{\infty}$  as evidenced by the higher portfolio standard deviation and distance to the fundamental weights. Importantly, the estimation error also induces much more variability in the portfolio allocation, as turnover increases by over 60%. When using more observations, the estimates become more precise and the portfolio allocations converge to the fundamental weights and the ones of HAR using population quantities.

In contrast, the HARQ obtains much better performance using small estimation samples. Compared to the HAR, the HARQ reduces the spurious turnover induced by estimation error by over 50% when  $M = 26$ , and even more than 60% when  $M = 78$  or  $M = 390$ . The greater improvements relative to the HAR for larger  $M$  are due to the fact that the degree of measurement error also needs to be estimated, and its estimate similarly becomes better when  $M$  increases. In addition to the reduced excess turnover, the HARQ portfolio obtains lower ex-post portfolio variance than the HAR, and portfolio weights that are significantly closer to the fundamental and HAR $_{\infty}$  weights. Moreover, as the final row demonstrates, the difference in HARQ and HAR covariance forecasts is truly driven by the measurement error, as the optimal weights are shrunk in the direction of HAR forecasts based on population quantities, and not some other more accurate forecast.

### 5.4.4 Empirical Results: Minimum Variance and Tracking Error Portfolios

In this section I evaluate the out-of-sample performance of the various models in the context of a portfolio selection problem. Of particular interest is whether the Q models improve allocation with lower portfolio variance and reduced turnover. I consider the standard GMV problem, as well as a tracking error portfolio, where the SPY is tracked.

**Table 5.3:** Simulation Results: Fundamental and Spurious Turnover

M	HAR				HARQ		
	$\infty$	26	78	390	26	78	390
Turnover	0.1878	0.2999	0.2830	0.2379	0.2376	0.2257	0.2071
St. Dev	0.1829	0.1920	0.1879	0.1844	0.1887	0.1853	0.1837
Distance to population	0.0446	0.0601	0.0531	0.0471	0.0527	0.0488	0.0457
Distance to $HAR_\infty$		0.0231	0.0198	0.0104	0.0141	0.0102	0.0062

*Note:* The table provides the results of a simulation study of out-of-sample portfolio allocation. Forecasts are based on MKernel estimates based on different numbers of observations  $M$  using the HAR and HARQ model. The  $M = \infty$  column provides the result of the portfolio allocation using the HAR model applied to population values of the covariance. Turnover is as in (5.19), St. Dev is the portfolio standard deviation based on the population covariance, and Distance reports the distance between the portfolio allocations  $\sqrt{\sum_n (\hat{w}_t^{(n)} - w_t^{(n)})^2}$ .

Constructing the minimum tracking-error variance portfolio is equivalent to constructing the GMV portfolio using returns in excess of the SPY. The minimum tracking portfolio is of interest for at least two reasons. First, it is empirically relevant as many institutional investors are evaluated based on their performance relative to a benchmark and full replication of the benchmark is often not desired or practical. Second, the covariance matrices of returns, net of a benchmark, are typically more dispersed as assets correlate to a different degree to the benchmark. This makes the minimum tracking error portfolio more challenging and prone to extreme weights.

As in Section 5.3, I obtain one-step-ahead forecasts using a rolling window of observations, re-estimating the parameters every day. The portfolios are re-balanced daily and evaluated according to their out-of-sample performance. To statistically compare the different models I use the MCS procedure of Hansen, Lunde, and Nason (2011) to obtain the set of models with minimum ex-post portfolio variance. Additionally, I use the utility based framework of Fleming, Kirby, and Ostdiek (2003) to evaluate the economic gains of switching from the standard to the Q model and the economic impact of turnover through transaction costs.

#### 5.4.4.1 Portfolio Allocation Results

Table 5.4 compares the out-of-sample performance of the GMV and minimum tracking error portfolio for the various models. Along with the dynamic models I consider a standard random walk forecast and the 1/N portfolio. The latter has been shown to be empirically hard to beat out-of-sample in terms of Sharpe ratio (DeMiguel, Garlappi, Nogales, and Uppal, 2009).

Panel A shows the results of the GMV portfolios. Amongst all models considered the HEAVYQ has the lowest out-of-sample portfolio variance. However, based on the MCS results in bold, its portfolio variance is statistically indistinguishable from that of all other Q models and the standard HAR. Moreover, the out-of-sample Sharpe Ratio

## 5. (UN)RELIABLE REALIZED RISK MEASURES

**Table 5.4:** Unconstrained Portfolio Results

		HAR	HARQ	HAR- DRD	HARQ- DRD	EWMA	EWMAQ	HEAVY	HEAVYQ	RW	1/N
<i>Panel A: Global Minimum Variance</i>											
TO		0.522	0.385	0.391	0.339	0.135	0.096	0.172	0.122	2.506	0.009
CO		0.513	0.517	0.487	0.497	0.505	0.506	0.497	0.497	0.868	0.316
SP		-0.105	-0.109	-0.070	-0.082	-0.095	-0.096	-0.089	-0.089	-0.547	0.000
Mean Ret		3.176	3.274	3.726	4.270	3.565	4.105	3.969	4.354	3.741	1.039
Std Ret		<b>15.147</b>	<b>15.020</b>	15.253	<b>15.060</b>	15.208	<b>14.911</b>	15.157	<b>14.883</b>	19.158	18.871
$c = 0\%$	Sharpe	0.210	0.218	0.244	0.283	0.234	0.275	0.262	0.293	0.195	0.055
	$\Delta_1$		0.113		1.225		0.567		0.413		
	$\Delta_{10}$		0.280		1.481		0.958		0.772		
$c = 1\%$	Sharpe	0.120	0.151	0.178	0.291	0.211	0.259	0.232	0.271	-0.145	0.054
	$\Delta_1$		0.459		1.355		0.664		0.540		
	$\Delta_{10}$		0.626		1.611		1.055		0.899		
$c = 2\%$	Sharpe	0.031	0.085	0.111	0.233	0.188	0.242	0.203	0.250	-0.485	0.053
	$\Delta_1$		0.804		1.485		0.762		0.667		
	$\Delta_{10}$		0.972		1.742		1.153		1.026		
<i>Panel B: Minimum Tracking Error Variance</i>											
TO		0.173	0.102	0.134	0.114	0.063	0.045	0.080	0.057	1.128	0.009
CO		0.339	0.338	0.339	0.342	0.340	0.340	0.338	0.338	0.461	0.316
SP		-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.087	0.000
Mean TE		1.123	1.273	1.069	1.215	0.990	1.010	0.799	0.811	0.560	0.949
Std TE		7.119	<b>6.665</b>	7.086	<b>6.738</b>	7.107	<b>6.735</b>	7.122	<b>6.706</b>	9.268	7.616
$c = 0\%$	$\Delta_1$		0.880		0.276		0.797		0.829		
	$\Delta_{10}$		1.153		0.485		1.022		1.079		
$c = 1\%$	$\Delta_1$		1.057		0.325		0.843		0.889		
	$\Delta_{10}$		1.330		0.535		1.068		1.139		
$c = 2\%$	$\Delta_1$		1.234		0.375		0.889		0.949		
	$\Delta_{10}$		1.508		0.585		1.113		1.199		

*Note:* The top panel shows the portfolio results for the Global Minimum Variance Portfolio and the bottom panel shows the Minimum Tracking Error results under the constraint of no short sales. Each panel shows turnover (TO), portfolio concentration (CO), the average short position (SP), as well as the average annualized return (tracking error) and its volatility. Standard deviations in bold depict the models that belong to the 90% Model Confidence Set of lowest ex-post daily volatility. For both portfolios I show the economic gains of switching from the standard model to the Q-model in annual percentage points,  $\Delta_\gamma$ , for various transaction cost levels  $c$  and risk aversion coefficients  $\gamma$ , as outlined in Equation (5.22).

(without transaction cost) of all models, including the non-Q models, is higher than that of the 1/N portfolio. For most models, this result remains when accounting for differences in turnover with all reasonable values of  $c$ . This finding is at odds with DeMiguel, Garlappi, Nogales, and Uppal (2009), who report that strategies based on covariance forecasts cannot consistently outperform a naive diversification strategy. However, they consider unconditional Sharpe ratios, whereas I focus on conditional portfolio volatility.

The variability between turnover of the different base models is large. Interestingly, the increased turnover of, for instance, the HAR does not consistently lead to better Sharpe ratios or lower portfolio variance than the models that imply less turnover. On the other hand, all Q models obtain a reduction in variance, an increase in average returns and in Sharpe ratio in conjunction with a decrease in turnover relative to their respective benchmarks.<sup>16</sup> Turnover is reduced by a minimum of 14% for the HARQ-DRD up to

<sup>16</sup>This finding is in line with the low volatility anomaly documented in amongst others, Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003) and Baker, Bradley, and Wurgler (2011).

almost 30% for the HEAVYQ. On the downside, and contrary to expectation, the Q portfolios appear to be slightly more concentrated and ask for greater short-positions, although the effect is economically small.

The bottom of Panel A shows the economic gains of switching to the Q models compared to their standard counterparts using the framework of Equation (5.22). When  $c = 0\%$ , there are no transaction costs, and the gains of switching are limited from about 0.1% up to 1.5% annually depending on the model. However, as all models have less turnover, the higher the transaction costs the greater the economic gains of switching. Depending on the specification, going from  $c = 0\%$  to  $c = 2\%$ , increases the economic gains between 0.3 and 0.7% annually, purely due to decreased turnover. Under different assumptions on  $c$ , the economic gains can simply be extrapolated.

Panel B of Table 5.4 contains the results of the minimum tracking error portfolios. The results are similar to those of the Global Minimum Variance portfolio. The standard deviation of the tracking error is significantly reduced by the Q models, which jointly form the model confidence set of minimum variance models. Again, turnover is significantly reduced by the Q models with similar magnitude as in the GMV portfolios. The mean tracking error is slightly further from zero for the Q models than for the non-Q models. However, as the average return is positive this is not of concern, and actually shows that the models obtain higher returns in excess of the benchmark while obtaining lower volatility. As such, the models similarly offer economic gains as reported in the bottom part of Panel B. The gains are consistently positive and amount to between 0.3% and 1.5% annually.

#### 5.4.4.2 No Shortselling

As a robustness check, I consider a no short-sale portfolio (e.g. Frost and Savarino, 1986). This is an empirically relevant restriction, as unlimited short positions are practically infeasible. Moreover, as shown by Jagannathan and Ma (2003), imposing short-selling constraints may enhance portfolio performance by limiting the impact of estimation error. In fact, portfolio restrictions have been shown to be equivalent to shrinkage of the covariance (Jagannathan and Ma, 2003). Given that imposing this constraint also reduces the impact of forecast and estimation error it is expected to reduce the gains of the Q models.

For the no-short sell minimum variance portfolio, there is the additional restriction that each position  $w_t^{(n)}$  is non-negative.

$$\begin{aligned}
 w_t &= \arg \min w_t' H_{t|t-1} w_t \\
 \text{s.t. } & w_t' \iota = 1, \\
 & w_t^{(n)} \geq 0, \forall n,
 \end{aligned}
 \tag{5.23}$$

which can be solved numerically using standard quadratic programming tools.

The results for the GMV and minimum tracking error portfolios under a no short-sale

## 5. (UN)RELIABLE REALIZED RISK MEASURES

**Table 5.5:** No Short-sale Portfolio Results

		HAR	HARQ	HAR- DRD	HARQ- DRD	EWMA	EWMAQ	HEAVY	HEAVYQ	RW	1/N
<i>Panel A: Global Minimum Variance</i>											
TO		0.390	0.284	0.322	0.280	0.101	0.071	0.131	0.093	1.139	0.009
CO		0.477	0.479	0.466	0.471	0.475	0.475	0.469	0.469	0.572	0.316
Mean Ret		3.874	4.210	4.046	4.271	4.050	4.632	4.363	4.967	2.800	1.039
Std Ret		15.607	15.274	15.537	15.260	15.500	<b>15.108</b>	15.489	<b>15.075</b>	16.837	18.871
$c = 0\%$	Sharpe	0.248	0.276	0.260	0.279	0.261	0.307	0.282	0.330	0.166	0.055
	$\Delta_1$		0.375		0.98		0.623		0.647		
	$\Delta_{10}$		0.825		1.301		1.147		1.200		
$c = 1\%$	Sharpe	0.183	0.227	0.207	0.298	0.244	0.294	0.260	0.314	-0.010	0.054
	$\Delta_1$		0.642		1.035		0.697		0.744		
	$\Delta_{10}$		1.091		1.407		1.221		1.297		
$c = 2\%$	Sharpe	0.118	0.179	0.153	0.250	0.227	0.282	0.238	0.298	-0.185	0.053
	$\Delta_1$		0.908		1.141		0.771		0.840		
	$\Delta_{10}$		1.358		1.514		1.295		1.393		
<i>Panel B: Minimum Tracking Error Variance</i>											
TO		0.172	0.102	0.134	0.114	0.063	0.045	0.080	0.056	0.884	0.009
CO		0.339	0.337	0.339	0.342	0.340	0.340	0.338	0.338	0.416	0.316
Mean TE		1.118	1.269	1.068	1.214	0.990	1.008	0.794	0.810	0.542	0.949
Std TE		7.120	<b>6.667</b>	7.087	<b>6.742</b>	7.110	<b>6.735</b>	7.123	<b>6.708</b>	9.311	7.616
$c = 0\%$	$\Delta_1$		0.879		0.272		0.792		0.826		
	$\Delta_{10}$		1.150		0.482		1.021		1.077		
$c = 1\%$	$\Delta_1$		1.052		0.320		0.842		0.886		
	$\Delta_{10}$		1.325		0.533		1.067		1.134		
$c = 2\%$	$\Delta_1$		1.230		0.372		0.887		0.945		
	$\Delta_{10}$		1.507		0.584		1.110		1.195		

*Note:* The top panel shows the portfolio results for the Global Minimum Variance Portfolio and the bottom panel shows the Minimum Tracking Error results under the constraint of no short sales. Each panel shows turnover (TO), portfolio concentration (CO), as well as the average annualized return (tracking error) and its volatility. Standard deviations in bold depict the models that belong to the 90% Model Confidence Set of lowest ex-post daily volatility. For both portfolios I show the economic gains of switching from the standard model to the Q-model in annual percentage points,  $\Delta_\gamma$ , for various transaction cost levels  $c$  and risk aversion coefficients  $\gamma$ , as outlined in Equation (5.22).

constraint are reported in Table 5.5. They show that the Q models still improve on the non-Q models, even after doing one of the standard adjustments to deal with measurement error. Indeed, the gains of the Q models with respect to their standard counterparts are reduced compared to the unconstrained portfolios. For the GMV portfolio only the EWMAQ and HEAVYQ are in the set of lowest ex-post portfolio volatility while for the minimum tracking error volatility, all four Q models obtain roughly equal performance.

### 5.4.4.3 Weekly and Monthly Rebalancing

Up until now I have focused on daily re-balancing of the portfolios. While this is an empirically relevant problem, in practice not all investors would be willing to re-balance their portfolios on a daily basis. Moreover, at longer horizons more data is available to accurately estimate the covariance matrix, which is of course an alternative method to deal with measurement error.

In this section, I investigate the performance of the Q models for weekly and monthly forecasts. Specifically, the models are used to predict next week's total volatility  $\Sigma_{t|t+4}$

Table 5.6: Longer Horizon Portfolio Results

		Weekly				Monthly			
		HAR-DRD	HARQ-DRD	RW	1/N	HAR-DRD	HARQ-DRD	RW	1/N
Panel A: Global Minimum Variance									
TO		0.005	0.005	0.006	0.005	0.003	0.002	0.003	0.003
CO		0.458	0.477	0.545	0.324	0.444	0.456	0.494	0.340
SP		-0.058	-0.063	-0.148	0.000	-0.046	-0.047	-0.089	0.000
Mean Ret		1.364	1.741	1.801	1.213	1.966	1.992	1.087	1.328
Std Ret		16.565	<b>16.030</b>	16.968	19.278	16.667	<b>16.451</b>	16.991	20.019
c=0%	Sharpe	0.082	0.109	0.106	0.063	0.118	0.121	0.064	0.066
	$\Delta_1$		0.450				0.060		
	$\Delta_{10}$		1.212				0.371		
c=1%	Sharpe	0.082	0.108	0.105	0.063	0.118	0.121	0.064	0.066
	$\Delta_1$		0.451				0.060		
	$\Delta_{10}$		1.213				0.372		
c=2%	Sharpe	0.081	0.107	0.104	0.062	0.117	0.120	0.063	0.066
	$\Delta_1$		0.452				0.061		
	$\Delta_{10}$		1.214				0.372		
Panel B: Minimum Tracking Error Variance									
TO		0.004	0.003	0.004	0.004	0.002	0.002	0.002	0.002
CO		0.337	0.342	0.358	0.304	0.334	0.338	0.343	0.328
SP		0.000	0.000	-0.003	0.000	0.000	0.000	0.000	0.000
Mean TE		0.796	0.801	0.799	1.020	0.804	0.885	0.783	0.998
Std TE		<b>7.667</b>	<b>7.603</b>	7.981	8.178	<b>7.689</b>	<b>7.692</b>	7.778	8.253
c=0%	$\Delta_1$		0.239				0.058		
	$\Delta_{10}$		0.337				0.067		
c=1%	$\Delta_1$		0.238				0.057		
	$\Delta_{10}$		0.336				0.066		
c=2%	$\Delta_1$		0.236				0.056		
	$\Delta_{10}$		0.335				0.065		

*Note:* The top panel shows the portfolio results for the Global Minimum Variance Portfolio and the bottom panel shows the Minimum Tracking Error results where portfolios are re-balanced weekly or monthly based on forecasts of the weekly and monthly covariance matrix. Each panel shows turnover (TO), portfolio concentration (CO), as well as the average annualized return (tracking error) and its volatility. Standard deviations in bold depict the models that belong to the 90% Model Confidence Set of lowest ex-post daily volatility. For both portfolios and horizons I show the economic gains of switching from the standard HAR-DRD to the HARQ-DRD in annual percentage points,  $\Delta_\gamma$ , for various transaction cost levels  $c$  and risk aversion coefficients  $\gamma$ , as outlined in Equation (5.22).

on each Friday, and next month's volatility  $\Sigma_{t|t+21}$  on each last trading day of the month. The manager re-balances the portfolio to the weights implied by the forecasts and holds that portfolio for the next week or month, where the process is repeated.

I focus on the HAR(Q)-DRD as an illustration, as it is straightforward to extend the model to longer horizons. The simplest way to obtain longer horizon forecasts is by means of direct projection, where instead of Equation (5.7), for  $h$ -step ahead forecasts we use

$$s_{t|t+h} = \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t. \quad (5.24)$$

Motivated by the long-horizon forecast results in Bollerslev, Patton, and Quaedvlieg (2016), I adapt the HARQ to the forecast horizon. While for the daily forecasts the daily lag is of greatest importance, when forecasting the weekly or monthly covariance matrix, the weekly and monthly lag increase in importance. Instead of time-varying attenuation in the daily lag, where  $\theta_1$  is made time-varying, I allow for time-variation in  $\theta_2$

## 5. (UN)RELIABLE REALIZED RISK MEASURES

---

for the weekly forecasts, and in  $\theta_3$  for the monthly forecasts. Specifically, the weekly and monthly specifications for the univariate HARQ are

$$\begin{aligned} s_{t|t+4} &= \theta_0 + \theta_1 s_{t-1} + \theta_{2,t} s_{t-5|t-1} + \theta_3 s_{t-22|t-1} + \epsilon_t \\ \theta_{2,t} &= (\theta_{2t} + \theta_{2Q} \pi_{t-1|t-5}), \end{aligned} \quad (5.25)$$

and

$$\begin{aligned} s_{t|t+21} &= \theta_0 + \theta_1 s_{t-1} + \theta_2 s_{t-5|t-1} + \theta_{3,t} s_{t-22|t-1} + \epsilon_t \\ \theta_{3,t} &= (\theta_{3t} + \theta_{3Q} \pi_{t-1|t-22}), \end{aligned} \quad (5.26)$$

where  $\pi_{t-1|t-h} = \left( \frac{1}{h} \sum_{i=1}^h \text{diag}(\Pi_{t-i}) \right)^{1/2}$ . Again, I also consider the 1/N portfolio and portfolios based on random walk forecasts. In line with the out-of-sample horizon, the random walk forecast is based on the average covariance of the past week or month.

The portfolio allocation results are reported in Table 5.6. The table shows that also at weekly and monthly frequencies accurate forecasts of the covariance matrix lead to better portfolio characteristics out-of-sample compared to the robust 1/N strategy. However, the relative gains of the HARQ-DRD compared to the standard HAR-DRD decrease when the forecasting horizon increases. This again highlights the fact that, due to the measurement error, the standard models particularly fail to capture the short-term dynamics, while the long-term trend is modeled accurately. Regardless, also on the weekly and monthly horizon, the HARQ-DRD improves on the standard HAR-DRD for both the GMV and minimum tracking error portfolios.

Another interesting finding is that the Sharpe Ratio of the portfolios based on HARQ-DRD forecasts with daily re-balancing (as reported in Table 5.4), is higher than those based on weekly and monthly re-balancing, even after accounting for transaction costs. This is not uniformly true for the standard HAR-DRD model, which is more heavily affected by transaction costs. This shows that the degree of noise in the short-term dynamics, if appropriately taken into account, does not prohibit the implementation of daily managed portfolios, even with high transaction costs.

### 5.5 Beta Forecasting

The realized beta is an important derived quantity of realized covariances and plays a central role in many asset-pricing and risk management applications. For instance, Bollerslev and Zhang (2003) and Boguth, Carlson, Fisher, and Simutin (2011) use realized betas to assess multi-factor asset pricing models more accurately, Corradi, Distaso, and Fernandes (2013) use them to back out conditional alphas, and Patton and Verardo (2012) use high-frequency betas to obtain market-neutral portfolios.

If the objective is to forecast the exposure to risk factors, modeling the full covariance matrix can be overly complicated. Instead, it is much simpler to directly model the realized betas themselves. In this section I investigate the merits of direct reduced

form modeling of realized beta by comparing it to betas derived from covariance matrix forecasts. Similar to (co)variances, the realized beta also suffers from heteroskedastic measurement error, and reduced form modeling based on the ex-post beta estimates requires the same considerations.

Let  $r_{i,t} = \{r_{i,t}^{(n)}, r_{i,t}^{(MKT)}, r_{i,t}^{(HML)}, r_{i,t}^{(SMB)}, r_{i,t}^{(MOM)}\}$ , i.e. the returns of asset  $n$ , followed by the four risk factors. Its covariance matrix can be partitioned as follows

$$S_t = \begin{bmatrix} S_{nn,t} & S_{fn,t} \\ S_{nf,t} & S_{ff,t} \end{bmatrix}, \quad (5.27)$$

with  $S_{nn,t}$  the variance of asset  $n$ ,  $S_{fn,t} = S'_{nf,t}$  the covariance of asset  $n$  with each of the factors, and  $S_{ff,t}$  the covariance matrix of the factors. Then, the realized beta is defined as

$$b_t^{(n)} = S_{fn,t} S_{ff,t}^{-1}. \quad (5.28)$$

Modeling  $b_t$  rather than the covariance matrix  $s_t$ , has the benefit that the problem's dimension is significantly reduced. With just the market beta, the dimension reduction is limited from three to one, but for the four factor model, the reduction is significant from fifteen to just four.

In this section I use the HAR and HARQ models to directly forecast realized beta and compare its forecasting accuracy in a number of empirical applications with betas derived from forecasted covariance matrices. The vech HAR format directly carries over to the four dimensional beta vector. In order to obtain a HARQ specification, the asymptotic distribution of the beta estimate is needed. As beta is a function of the covariance estimate whose distribution is known, this is a straightforward application of the delta-method. Here I extend the result of Barndorff-Nielsen and Shephard (2004a) to the case when there is more than just one factor.

Let  $\Psi_t = D_N \Pi_t D'_N$ , with  $D_N$  the duplication matrix as before. Then

$$M^{1/k}(b_t - \beta_t) \rightarrow_{Ls} MN(0, \nabla_t \Psi_t \nabla'_t), \quad (5.29)$$

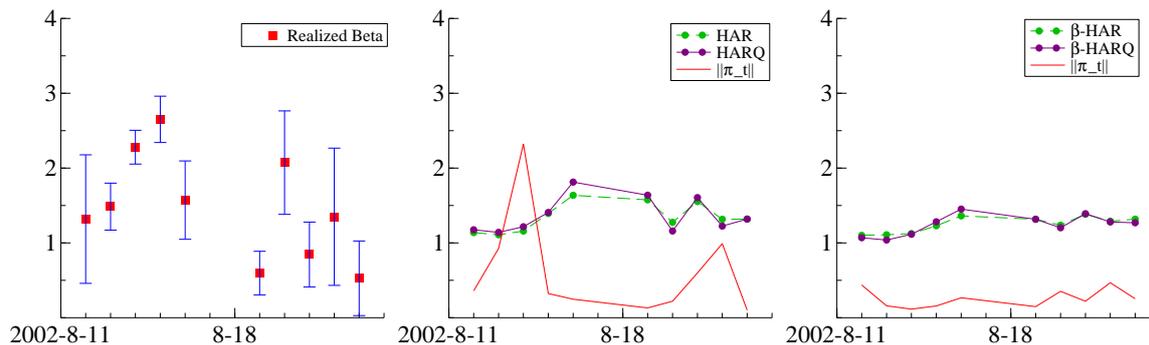
where  $k$  is dependent on the convergence rate of the estimator used and

$$\nabla_t = \frac{dvec(\beta_t)}{dvec(\Sigma_t')} = ((A \Sigma_t A')^{-1} A) \otimes (B - B \Sigma_t A' (A \Sigma_t A')^{-1} A) \quad (5.30)$$

with  $A = \begin{bmatrix} 0_{N-1} & I_{N-1} \end{bmatrix}$  and  $B = \begin{bmatrix} 1 & 0'_{N-1} \end{bmatrix}$ . The result is derived in Appendix D.<sup>17</sup>

In order to model  $\beta$  directly, we can apply the HAR to the realized beta vector. I

<sup>17</sup>Note that in the bivariate case this result reduces to the standard result for scalar realized beta. Indeed, in the bivariate case,  $\nabla = \begin{bmatrix} 0 & 0 & \Sigma_{22}^{-1} & \Sigma_{12} \Sigma_{22}^{-2} \end{bmatrix}$  which reduces to  $M^{1/k}(b_t - \beta_t) \rightarrow N \left[ 0, (\Sigma_{22,t})^{-2} \begin{bmatrix} 1 & -\beta_t \end{bmatrix} \Psi_t \begin{bmatrix} 1 & -\beta_t \end{bmatrix}' \right]$  with  $\beta_t = \Sigma_{12,t} \Sigma_{22,t}^{-1}$ .

**Figure 5.2:** Beta Estimates and Forecasts for General Electric


*Note:* The graph on the left displays the realized beta estimates with 95% confidence bounds for a two week period during the dot-com crisis. The middle graph provides HAR and HARQ forecasts along with the norm of the measurement error  $\pi_t$  which governs the dynamics of HARQ. The right graph similarly plots the forecasts from the  $\beta$ -HAR and  $\beta$ -HARQ along the degree of measurement error which governs the dynamics of  $\beta$ -HARQ.

define the  $\beta$ -HAR as

$$b_t = \theta_0 + \theta_1 b_{t-1} + \theta_2 b_{t-5|t-1} + \theta_3 b_{t-22|t-1} + \epsilon_t. \quad (5.31)$$

Using the previous result, the  $\beta$ -HARQ can be defined by replacing  $\theta_1$  with  $\theta_{1,t} = \theta_1 + \theta_{1Q}\pi_t$ , with  $\pi_t = \sqrt{\text{diag}(\nabla_t \Psi_t \nabla_t')}$ , analogous to the HARQ for covariances. Of course, forecasts for beta could be improved by allowing more general specifications and potentially even by modeling them all separately, but for now I consider models that are similarly parsimoniously parametrized to those of the covariance matrix to keep the comparison fair.

The distinction between modeling beta and the covariances is even more important for the HARQ and  $\beta$ -HARQ. Due to the non-linear nature of the beta transformation, the overall degree of measurement error of the covariance and beta can significantly diverge. Where the measurement error of the covariance may be above average, requiring greater attenuation, the beta estimate may be estimated with above average precision, allowing for strong short-term dynamics. This is especially important given the fact that small changes in covariances can have a large impact on betas.

### 5.5.1 HARQ for Beta Forecasting

The workings of the ( $\beta$ -)HAR(Q) are illustrated in Figure 5.2, where I plot General Electric's (GE) market beta estimates along with forecasts from the various models. The left-most graph plots the realized beta estimates along with 95% confidence bounds. The middle and right graph show the forecasts from the covariance and beta models respectively, along with the norm of the measurement error variance,  $\|\pi_t\|$  to proxy the total

degree of measurement error which governs the dynamic attenuation of the Q models.

The figure illustrates that the measurement error variance of the covariance matrix and the beta estimates is largely uncorrelated, and as a result the daily variation in responsiveness of beta differs across the two models. First, the overall responsiveness of the Q models is greater than of the non-Q models. An interesting data point to analyze is the forecast for the 15th of August, the fourth observation in the panels. This forecast follows large uncertainty in the covariance matrix, but a precisely estimated beta. Compared to the HAR, the HARQ is less responsive to this observation, and beta only rises the day after, when there is a precise covariance estimate. In contrast, compared to the  $\beta$ -HAR, the  $\beta$ -HARQ is immediately responsive to the high beta estimate of the previous day. Comparing the covariance with the beta models, I find that overall there is less short-term variability in the beta forecasts derived from the  $\beta$ -HAR models. This is due to the simple fact that small changes in the covariance can lead to large changes in betas.

In the remainder I investigate which model captures the dynamics of beta best. First I consider a small simulation study in which I evaluate the relative merits of the different models for forecasting beta, and investigate the effect of dimension reduction by considering a range of one to four betas, where the vech of the covariance matrix has dimension between three and fifteen. Next, I report the in- and out-of-sample performance of the HAR(Q) and  $\beta$ -HAR(Q) for the purpose of modeling beta and finally, I present the results of a number of empirical applications using the beta forecasts.

### 5.5.2 Simulation Results: The merits of dimension reduction

In the previous section I highlighted two reasons why modeling beta directly rather than inferring it from covariance forecasts might be advantageous. First, there is a significant dimension reduction of the vector to forecast, and second, the dynamics of the attenuation intensity of beta models are distinctly different from covariance models. Hence, the covariance models might attenuate too heavily on the wrong days for the purpose of beta forecasts.

To analyze these model characteristics, this section reports the results of a small simulation study. I consider the out-of-sample performance of one-step-ahead beta forecasts from the various models. I forecast beta vectors of dimension one to four, randomly selecting one of all possible permutations of the Fama and French (1993) and Carhart (1997) factors. When forecasting a single beta, the dimension reduction is limited from three to one, while for all four betas, the dimension is reduced from fifteen to four by directly modeling the betas.

Similar to the previous simulation, in order to simulate realistic sample paths of the covariance, for each iteration, I sample a random block of 2,000 consecutive estimated covariance matrices from a randomly selected Dow Jones constituent. Within this block of data I take the covariance matrix of the stock with a randomly selected subset of

## 5. (UN)RELIABLE REALIZED RISK MEASURES

**Table 5.7:** Simulation Results: Beta Forecasts

Dimension Beta		1		2		3		4	
	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
HAR	$\infty$	-0.004	0.473	0.009	0.777	-0.008	0.849	-0.003	0.912
	26	-0.006	0.489	0.005	0.806	-0.008	0.892	-0.003	0.952
	78	-0.005	0.488	0.006	0.801	-0.007	0.885	-0.002	0.951
	390	-0.004	0.483	0.006	0.800	-0.007	0.886	-0.002	0.951
HARQ	26	-0.004	0.479	0.004	0.783	-0.008	0.856	-0.001	0.920
	78	0.000	0.476	0.005	0.779	-0.006	0.852	-0.001	0.914
	390	0.000	0.474	0.005	0.778	-0.008	0.849	-0.010	0.913
$\beta$ -HAR	$\infty$	-0.008	0.469	-0.007	0.760	-0.008	0.827	-0.004	0.879
	26	-0.010	0.478	-0.008	0.769	-0.010	0.842	-0.004	0.899
	78	-0.009	0.476	-0.007	0.766	-0.009	0.839	-0.004	0.895
	390	-0.008	0.474	-0.007	0.764	-0.008	0.834	-0.004	0.889
$\beta$ -HARQ	26	-0.010	0.474	-0.007	0.764	-0.008	0.830	-0.004	0.882
	78	-0.009	0.471	-0.006	0.762	-0.008	0.828	-0.004	0.880
	390	-0.007	0.470	-0.006	0.761	-0.008	0.827	-0.004	0.879

*Note:* The table provides the results of a simulation study of out-of-sample beta forecasting. Forecasts are based on MKernel estimates of either the realized covariance or realized beta, based on different numbers of observations  $M$  using the HAR and HARQ models. The  $M = \infty$  row provides the result of the HAR models applied to population values of the covariance and beta.

between one and all four of the risk factors. On each day I simulate returns at one-second frequency with a diurnal volatility pattern (Andersen, Dobrev, and Schaumburg, 2012), such that the Integrate Covariance of the simulated data matches that of the empirical estimate. I then aggregate the paths to 1, 5 and 15-minute returns ( $M = 390, 78, 26$ ), and estimate the MKernel, the implied realized beta, and their asymptotic distributions.

Using a rolling window of 1,000 observations I make one-step-ahead forecasts of the covariances and betas and compare the forecasted betas with the population beta. Again, the different sampling frequencies allow us to attribute the gains to the Q models' ability to better cope with the heteroskedastic measurement error, while the different dimensions allow us to attribute gains to the degree of dimension reduction. To set a base-level of performance, I also consider HAR forecasts using the true population covariances and betas ( $M = \infty$ ).

The results of the simulation are reported in Table 5.7. First, the models' forecasts become less accurate as the dimension of the beta vector increases, with RMSE increasing by a factor two between the one- and four-dimensional betas. For both the covariance and beta models, we see improvements for the HARQ models compared to the HAR models. The HARQ models converge more quickly to the minimum obtainable bias and RMSE of the models applied to population values. Comparing the beta and covariance models, the simulations show that the beta models have slightly greater bias, but typically lower RMSE. Moreover, the percentage reduction in RMSE of the beta models increases with the dimension of beta showing that the dimension reduction plays a large role in the

**Table 5.8:** In-Sample Parameter Estimates Betas

	HAR		HARQ		$\beta$ -HAR		$\beta$ -HARQ	
	Average	StDev	Average	StDev	Average	StDev	Average	StDev
$\theta_1$	0.2355	0.0445	0.3481	0.0318	0.0231	0.0112	0.1629	0.0412
$\theta_2$	0.3609	0.0478	0.3464	0.0482	0.1071	0.0594	0.0870	0.0595
$\theta_3$	0.3001	0.0318	0.2526	0.0322	0.6197	0.0969	0.5586	0.0876
$\theta_{1Q}$			-0.0394	0.0182			-0.0122	0.0120

*Note:* The table shows the average and standard deviation of the parameter estimates across the 27 DJIA assets.

improvements of the models.

### 5.5.3 In-Sample Results

I consider the 27 assets described in Section 5.3.1 along with the four high-frequency risk factors. For each of the assets I estimate the 5-dimensional covariance matrix and compute realized betas. In this section I present in-sample parameter estimates for the HAR(Q) and  $\beta$ -HAR(Q) models. The parameter estimates are reported in Table 5.8. For each of the models I report the average and standard deviation of the parameter estimates across the 27 assets. Interestingly, the magnitude of the  $\theta_{1Q}$  coefficients relative to  $\theta_1$  is greater for these models than in the previous section where I modeled ten-dimensional matrices. This is in line with expectations as the measurement error in  $\pi_t$  itself aggregates across the cross-section, leading to attenuation of the Q parameter.

Andersen, Bollerslev, Diebold, and Wu (2006) report that the realized beta is less persistent than the covariances which is confirmed in my analysis. However, as the parameter estimates of the  $\beta$ -HARQ model show, part of this is due to the measurement error of  $b$ , as  $\theta_1$  increases significantly compared to the standard  $\beta$ -HAR. The model allows for far stronger short term dynamics when the beta estimates are precise, whereas the standard model is mostly driven by the long-term average.

### 5.5.4 Out-of-Sample Results

In this section I compare the four models in terms of their ability to forecast the beta vector. I use the same forecasting set-up as in the previous setting, re-estimating parameters daily using a rolling window of 1,000 observations. The forecasts are evaluated based on their bias and RMSE relative to the ex-post estimates for each beta separately, even though they are modeled jointly.

The forecast results are reported in Table 5.9. First, there is great variability in the forecasting accuracy across betas. The momentum factor loading is easier to forecast, while the SMB loading has the worst results. Overall we see that typically the  $\beta$ -HAR(Q) models have lower bias and also lower RMSE than the the beta forecasts implied by covariance forecasts. Additionally, the Q models reduce the RMSE of the

**Table 5.9:** Out-of-Sample Forecast Results

	HAR		HARQ		$\beta$ -HAR		$\beta$ -HARQ	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MKT	-0.0153	0.0289	-0.0086	<b>0.0263</b>	-0.0108	0.0281	-0.0109	<b>0.0261</b>
HML	0.0036	0.0254	0.0001	0.0253	-0.0168	<b>0.0247</b>	-0.0162	<b>0.0243</b>
SMB	-0.0231	0.0279	-0.0217	0.0274	-0.0200	0.0267	-0.0187	<b>0.0245</b>
MOM	-0.0037	0.0172	-0.0068	<b>0.0163</b>	0.0014	0.0168	0.0011	<b>0.0163</b>

*Note:* The table gives the average bias and RMSE of beta forecasts across time and the 27 DJIA assets. RMSE in bold depict the models that belong to the 90% Model Confidence Set of best forecasting models for that particular beta.

forecasts compared to the non-Q models without exception, while the bias is only slightly reduced, similar to the covariance results reported in Section 5.3.

To statistically determine whether the  $\beta$ -HAR and Q-models significantly improve the forecasts I again use the 90% MCS to determine per risk factor, which model forecasts most accurately. The models that make up the MCS are highlighted in bold in Table 5.9. Although the improvements in RMSE may seem small, they are consistent enough to reject the null of equality of models. The  $\beta$ -HARQ is part of the set of best models for all risk factors signifying the merits of modeling realized beta directly.

### 5.5.5 Economic Evaluation of Beta Forecasts

In this section I compare the beta forecasts produced by the HAR(Q) and  $\beta$ -HAR(Q) using two methods. First, I evaluate the different models by comparing their adequacy for hedging factor risk, which is captured by their out-of-sample residual variance. I then analyze these results more deeply by comparing them directly and jointly in a regression-based comparison.

#### 5.5.5.1 Beta Hedging

In order to economically evaluate the beta forecasts, I use them for beta hedging. The factor structure should most adequately capture the returns for the model with the most accurate beta forecasts. To compare the different models based on this premise, for each model and each asset, I compute the time series

$$Z_t^{(n)} = b_{MKT,t}^{(n)} r_{MKT,t} + b_{HML,t}^{(n)} r_{HML,t} + b_{SMB,t}^{(n)} r_{SMB,t} + b_{MOM,t}^{(n)} r_{MOM,t}, \quad (5.32)$$

where each model has its own  $Z_t^{(n)}$  implied by their respective beta forecasts. I then compute the implied factor tracking error

$$r_t^{(n)} - Z_t^{(n)}. \quad (5.33)$$

The best model should have the lowest tracking error variance. I compute the sample variance of their time-series, and use the Model Confidence Set of Hansen, Lunde, and Nason (2011) to identify the beta forecasting specification whose variance is significantly smaller than that of the competitors. A similar evaluation is performed in Hansen, Lunde, and Voev (2014) and Boudt, Laurent, Lunde, Quaevdlieg, and Sauri (2015).

The differences in tracking error variances between the models are large. The average residual variance across assets of the HAR is 1.40, compared to 1.31 for the HARQ. This large reduction is a result of the better conditioned forecasts produced by the HARQ, where the inversion needed for the beta forecasts is less affected by the measurement error. Both  $\beta$  models significantly reduce the tracking error variance relative to the covariance models, with the  $\beta$ -HAR obtaining just 1.14 and the  $\beta$ -HARQ modeling the factor structure slightly better at 1.13. To statistically evaluate the magnitude of these differences I construct the model confidence such that it contains the model with the smallest tracking error with a probability of no less than 90%. I compute the set for each of the 27 DJIA stocks and find that the HAR is included just once, while the HARQ is included for six of the assets. The  $\beta$ -HAR models lead to much lower tracking error, with the normal  $\beta$ -HAR being included in the set seventeen times, and the  $\beta$ -HARQ 25 times.

### 5.5.5.2 Regression Based Comparison

In the previous section, I considered the tracking errors based on the different models separately. In this section, I run the following regression to allow for a direct competition between the different models

$$r_t^{(n)} = \alpha^{(n)} + \delta_{HAR}^{(n)} Z_{t,HAR}^{(n)} + \delta_{HARQ}^{(n)} Z_{t,HARQ}^{(n)} + \delta_{\beta-HAR}^{(n)} Z_{t,\beta-HAR}^{(n)} + \delta_{\beta-HARQ}^{(n)} Z_{t,\beta-HARQ}^{(n)} + \epsilon_t^{(n)}. \quad (5.34)$$

$Z_t^{(n)}$  is as before in Equation (5.32), with the subscript denoting from which model the beta forecasts are derived. Perfect forecasts of a model would correspond to its  $\delta$  coefficient being equal to one, while all other models'  $\delta$  should be equal to 0. Engle (2015) uses this approach to find support for time-variation in beta by comparing betas obtained from a DCC model with static betas. I use it to investigate whether directly modeling beta offers advantages over modeling it indirectly through the covariance matrix, and whether the Q models lead to better forecasts compared to the non-Q models.

First, consider the full Equation (5.34), where we can compare all models at once. To

## 5. (UN)RELIABLE REALIZED RISK MEASURES

---

test whether any specification dominates all others I test the following four hypotheses:

$$\begin{aligned}
 H_{HAR} & : \delta_{HAR} = 1, \delta_{HARQ} = 0, \delta_{\beta-HAR} = 0, \delta_{\beta-HARQ} = 0 \\
 H_{HARQ} & : \delta_{HAR} = 0, \delta_{HARQ} = 1, \delta_{\beta-HAR} = 0, \delta_{\beta-HARQ} = 0 \\
 H_{\beta-HAR} & : \delta_{HAR} = 0, \delta_{HARQ} = 0, \delta_{\beta-HAR} = 1, \delta_{\beta-HARQ} = 0 \\
 H_{\beta-HARQ} & : \delta_{HAR} = 0, \delta_{HARQ} = 0, \delta_{\beta-HAR} = 0, \delta_{\beta-HARQ} = 1,
 \end{aligned}$$

using robust inference (White, 1980). These hypotheses are very strong, especially considering the strong correlation between the forecasts, and are expected to be rejected for almost every DJIA stock.

In order to better highlight the relative merits of the different models I split the analysis up in parts. First, in order to directly compare the Q models with their standard counterparts I estimate Equation (5.34) with either just the HAR and HARQ, or the  $\beta$ -HAR and  $\beta$ -HARQ. I then test the following hypotheses in the first regression

$$H'_{HARQ} : \delta_i^{HAR} = 0 \quad \text{and} \quad H'_{HAR} : \delta_i^{HARQ} = 0$$

and

$$H'_{\beta-HARQ} : \delta_i^{\beta-HAR} = 0 \quad \text{and} \quad H'_{\beta-HAR} : \delta_i^{\beta-HARQ} = 0$$

in the second regression. Note that for the hypothesis  $H'_{HARQ}$  I test whether the HAR model is still significant within the regression. Hence, a non-rejection is in favor of the HARQ, similar to the first set of hypotheses.

Next, in order to directly test the usefulness of the direct modeling of beta compared to using the forecasts for the covariance matrix, I estimate Equation 5.34 with just the HAR and  $\beta$ -HAR, and similarly with the HARQ and  $\beta$ -HARQ. Based on these two regressions we can test the following set of hypotheses:

$$\begin{aligned}
 H''_{\beta-HAR} : \delta_i^{HAR} = 0 & \quad \text{and} \quad H''_{HAR} : \delta_i^{\beta-HAR} = 0 \\
 H''_{\beta-HARQ} : \delta_i^{HARQ} = 0 & \quad \text{and} \quad H''_{HARQ} : \delta_i^{\beta-HARQ} = 0
 \end{aligned}$$

in the first and second regressions, respectively. I estimate the five different regressions and test all twelve hypotheses for each of the 27 assets. A summary of the parameter estimates, as well as the rejection frequencies, are reported in Table 5.10.

Panel A reports the point estimates of the various models. The top half of the panel shows descriptives of the parameter estimates across the assets for the full specification of Equation (5.34), while the bottom half shows the parameter estimates when one  $Z$  is included at a time. First, due to the high collinearity of the various  $Z$  variables, large variation can be seen in the point estimates. On average, the HARQ and  $\beta$ -HARQ obtain the point estimates closest to one. When only considering one model at a time, all models appear able to capture the betas relatively well, with average point estimates

**Table 5.10:** Regression Based Comparison of Betas

Panel A: Parameter Estimates				
	$\delta_i^{HAR}$	$\delta_i^{HARQ}$	$\delta_i^{\beta-HAR}$	$\delta_i^{\beta-HARQ}$
Full Model				
Mean	-0.183	0.618	-0.086	0.648
StDev	0.307	0.361	1.017	1.070
25% Quartile	-0.279	0.383	-0.681	-0.119
Median	-0.152	0.682	-0.143	0.840
75% Quartile	0.041	0.804	0.648	1.271
Individual				
Mean	0.853	0.917	0.991	0.992
StDev	0.074	0.060	0.095	0.095
25% Quartile	0.800	0.884	0.923	0.926
Median	0.873	0.917	0.995	0.997
75% Quartile	0.901	0.961	1.059	1.060
Panel B: Rejection Frequencies Hypothesis Testing				
Full Model				
	$H_{HAR}$ 27	$H_{HARQ}$ 25	$H_{\beta-HAR}$ 24	$H_{\beta-HARQ}$ 20
Q vs. Non-Q	$H'_{HAR}$ 26	$H'_{HARQ}$ 0	$H'_{\beta-HAR}$ 15	$H'_{\beta-HARQ}$ 5
Beta vs. Covariance	$H''_{HAR}$ 25	$H''_{HARQ}$ 23	$H''_{\beta-HAR}$ 3	$H''_{\beta-HARQ}$ 5

*Note:* This table reports the results of regression comparison based on Equation (5.34). Panel A shows parameter estimates of the full model and of the regression models which have one  $Z$  at a time. Panel B shows rejection frequencies of the null-hypotheses that each model best fits the data, hence lower numbers are better. The number reported is the frequency of rejections out of the possible 27 for each stock. The first row directly compares all models at once, and tests whether the models  $\delta$  parameter equals one and all others equal zero jointly. The second row compares the Q with the non-Q models, i.e.  $r_t = \alpha + \delta_{HAR}Z_{HAR,t} + \delta_{HARQ}Z_{HARQ,t}$  and analogous for the  $\beta - HAR$  models. Here the hypotheses are  $H'_{HAR} : \delta_{HARQ} = 0$  and vice versa, such that rejection is in favor of the model in the hypothesis subscript. The final row compares beta forecasts with betas derived from covariance forecasts using the regressions  $r_t = \alpha + \delta_{HAR}Z_{HAR,t} + \delta_{\beta-HAR}Z_{\beta-HAR,t}$ , and analogous for the Q-models. Here,  $H''_{HAR} : \delta_{\beta-HAR} = 0$  and vice versa.

close to one, and again the  $\beta-HAR(Q)$  significantly closer than the covariance models.

Panel B shows the rejection rates for the various hypotheses. First consider the results on the full specification in the top row. The hypothesis that the HAR forecasts the ideal time series of beta is rejected for every asset, leaving no doubt that the HAR is not the best model. The HARQ is not rejected for a small set of series, with even fewer rejections for the  $\beta-HAR$  and  $\beta-HARQ$  specification. The  $\beta-HARQ$  survives this stringent test the best, as the model is not rejected as the single best model for seven of the series.

Next consider the results of the direct comparison of Q models with non-Q models in the second row of Panel B. The HAR model is almost uniformly rejected, implying that the HARQ almost always obtains non-zero weight in the direct comparison between HAR

and HARQ. On the other hand, the HARQ hypothesis is never rejected, implying that there is no significant loading on HAR. We see a similar, though less extreme, picture for the  $\beta$ -HAR(Q) models, where the forecasts from the Q model also better describe the data.

Finally consider the bottom row in Panel B, where I test whether modeling beta directly leads to better results than modeling the covariance. Here, the HAR and  $\beta$ -HAR are directly compared. Both for the Q and non-Q models, the  $\beta$  specification dominates the covariance specification, with far greater rejection frequencies for the covariance models. This confirms the overall conclusion of this section that modeling beta directly rather than indirectly leads to improved forecasts, with a possibility of additional improvements through dynamic attenuation models.

### 5.6 Conclusion

This paper proposes a simple extension to multivariate volatility models that incorporate high-frequency covariance estimates. The models improve on their benchmarks by allowing the responsiveness to the high-frequency estimates to vary over time with the heteroskedasticity of the measurement error. The models use the asymptotic theory of the high-frequency estimator to parsimoniously allow for dynamic attenuation of the autoregressive parameters. The new high-frequency information is influential when the estimate is precise, and the impact is significantly reduced when the estimate is noisy. This allows the models to incorporate the information of high-frequency returns much more efficiently, compared to standard models, that have a fixed degree of attenuation. I illustrate the concept on three of the most popular multivariate volatility models, the vech HAR, EWMA and HEAVY. The models with time-varying parameters are dubbed HARQ, EWMAQ and HEAVYQ, as the adjustment is based on the Integrated Quarticity of high-frequency returns. The simple extension leads to significantly improved out-of-sample performance.

I evaluate the performance of the models in various financial applications. First, I consider a dynamic portfolio allocation for which the Q models are perfectly suited. I use Global Minimum Variance portfolios such that the performance of the various models is not contaminated by estimation error in the mean. Here the models can be interpreted as dynamic shrinkage models, where the covariance forecasts are shrunk towards the unconditional covariance matrix when estimation uncertainty is large. The more accurate forecasts of the Q models lead to better portfolio allocation, with reduced ex-post variance and higher Sharpe ratios. Moreover, by limiting the influence of noisy covariance estimates, the forecasts are more stable, which leads to reduced turnover. As a result, an investor implementing this volatility-timing strategy would be willing to pay up to 180 basis points per year to switch to the Q models from their respective benchmark.

In addition, I use the models to forecast betas. Here, I evaluate the forecasting accuracy of dynamically modeling realized betas, versus betas implied by forecasts from a

covariance model. I similarly derive a Q version for the dynamics of beta, incorporating the asymptotic distribution of the realized beta estimate. The main advantage of directly modeling beta compared to indirectly through covariances is the significant reduction in the dimension of the problem. Moreover, the attenuation dynamics of the Q models can differ significantly between the beta and covariance models. I find that out-of-sample forecasts of dynamic models of beta are more accurate from a statistical point of view, and also perform better in a number of risk management applications, with the single best model being the Q version of the dynamic model of beta.

## Appendix A: MKernel Theory and Estimation

The Multivariate Kernel is defined as

$$\text{MKernel}_t = \sum_{h=-M}^M k(h/H)\Gamma_h, \quad (5.35)$$

where  $\Gamma_h = \sum_{i=h+1}^M r_{i,t}r'_{i-h,t}$ ,  $\Gamma_h = \Gamma'_{-h}$  and  $k(\cdot)$  is an appropriate kernel function, where I use the parzen kernel. The asymptotic theory of the MKernel requires ‘jittering’, i.e. averaging out the first few and last few observations, which I do implement but leave out for notational convenience.

According to Theorem 2 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011),

$$M^{1/5}(\text{vech MKernel}_t - \text{vech } \Sigma_t) \rightarrow_{Ls} MN(\omega_t, \Pi_t) \quad (5.36)$$

The bias term,  $\omega_t$ , is negligibly small and ignored in this paper, and  $\Pi_t = 3.777IQ_t$ , where  $IQ_t$  is the Integrated Quarticity. Following the suggestions in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), I estimate  $\Pi_t$  using the estimator of Barndorff-Nielsen and Shephard (2004a), but on pre-averaged data. Defining  $x_{i,t} = \text{vech}(\bar{r}_{i,t}\bar{r}'_{i,t})$ , where  $\bar{r}_{i,t}$  is the pre-averaged data using bandwidth  $H$ , the  $IQ_t$  can be estimated as

$$\widehat{IQ}_t = M \sum_{i=1}^{M-H} x_{i,t}x'_{i,t} - \frac{M}{2} \sum_{j=1}^{M-H-1} (x_{i,t}x'_{i+1,t} + x_{i+1,t}x'_{i,t}). \quad (5.37)$$

## Appendix B: Descriptive Statistics

Table 5.11 provides a set of summary statistics for the daily variance estimates. The estimates in this table are obtained using the Realized Kernel estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), the univariate counterpart of the MKernel.

**Table 5.11:** Realized Kernel Summary Statistics

Name		Min	Mean	Median	Max	Skewness	Kurtosis
Factors							
Market	MKT	0.025	0.954	0.483	62.824	11.465	221.171
High-minus-Low	HML	0.022	0.563	0.265	36.710	13.226	315.023
Small-minus-Big	SMB	0.034	0.831	0.435	44.386	10.219	173.225
Momentum	MOM	0.041	1.865	0.608	92.309	9.003	121.663
Individual Stocks							
American Express	AXP	0.035	3.906	1.876	167.268	7.853	107.127
Boeing	BA	0.084	2.782	1.641	82.516	7.042	86.481
Caterpillar	CAT	0.086	3.329	1.961	107.856	6.681	84.979
Cisco Systems	CSCO	0.086	5.284	2.920	313.587	13.314	348.012
Chevron	CVX	0.069	1.975	1.249	118.871	14.147	319.889
DuPont	DD	0.023	2.723	1.579	97.924	7.072	102.556
Walt Disney	DIS	0.085	2.835	1.584	82.146	5.926	60.945
General Electric	GE	0.037	2.709	1.268	147.268	9.763	151.835
The Home Depot	HD	0.053	3.215	1.837	121.143	7.411	107.310
IBM	IBM	0.054	2.250	1.262	62.827	5.983	66.064
Intel	INTC	0.098	4.147	2.448	88.254	4.910	45.995
Johnson & Johnson	JNJ	0.028	1.582	0.956	54.406	7.455	106.876
JPMorgan Chase	JPM	0.046	4.329	2.021	200.125	9.030	128.000
Coca-Cola	KO	0.021	1.740	1.014	49.874	7.128	89.965
McDonald's	MCD	0.027	2.122	1.309	88.773	8.871	167.448
3M	MMM	0.050	1.929	1.121	113.378	13.476	363.764
Merck	MRK	0.041	2.346	1.431	108.603	10.357	201.485
Microsoft	MSFT	0.041	2.897	1.781	54.623	4.735	42.371
Nike	NKE	0.087	3.202	1.731	97.949	6.024	68.289
Pfizer	PFE	0.088	2.350	1.483	74.954	6.888	90.296
Procter & Gamble	PG	0.040	1.687	0.993	75.425	9.049	160.917
Travelers	TRV	0.073	2.805	1.294	134.942	9.020	135.165
UnitedHealth Group	UNH	0.070	4.097	2.231	156.822	8.046	106.611
United Technologies	UTX	0.066	2.436	1.354	136.111	11.758	288.459
Verizon	VZ	0.084	2.238	1.279	146.248	14.234	457.643
Wal-Mart	WMT	0.067	2.483	1.467	88.586	6.698	94.866
ExxonMobil	XOM	0.065	1.789	1.066	147.160	18.078	563.450

*Note:* The table provides summary statistics for the Realized Kernel for each of the 27 DJIA stocks and the high-frequency risk factors. The sample covers the period from February 1993 through December 2013.

## Appendix C: Composite Likelihood Estimators

Estimation of large-dimensional models often becomes very time-consuming due to the need to invert the large dimensional conditional covariance matrix.

Pakel, Shephard, Sheppard, and Engle (2014) propose a composite likelihood theory. The method is based on summing up quasi-likelihoods of subsets of assets. Each subset has a valid quasi-likelihood, but has only limited information on the parameters. By summing over many subsets the estimator has the desirable property that we do not have to invert large-dimensional matrices.

The quantity of interest is  $V_t = E(r_t r_t')$ , where  $V_t$  is modeled parametrically, by means of e.g. the EWMA or HEAVY model, with parameter vector  $\psi$ . Assuming normality, the quasi-likelihood is given by

$$\log L(\psi; r) = \sum_{t=1}^T \ell(\psi; r_t) \text{ with } \ell(\psi; r_t) = -\frac{1}{2} \log |V_t| - \frac{1}{2} r_t' V_t^{-1} r_t. \quad (5.38)$$

The main problem with this is that  $V_t$  is potentially of large dimension and it has to be inverted often with numerical optimization.

Composite likelihoods sidestep this issue by instead approximating the likelihood with a large number of lower dimensional marginal densities. The dimension of the problem is reduced from  $N$  to 2. While different options are available, I will consider contiguous pairs,  $Y_{1t} = (r_{1,t}, r_{2,t})$ ,  $Y_{2t} = (r_{2,t}, r_{3,t}), \dots, Y_{d-1,t} = (r_{d-1,t}, r_{d,t})$ , with composite likelihood

$$CL(\psi) = \frac{1}{T(N-1)} \sum_{j=1}^{N-1} \sum_{t=1}^T \ell_{jt}(\psi; Y_{jt}) \quad (5.39)$$

The estimator is consistent and asymptotically normal. Distributions of the estimates based on this theory are available in Pakel, Shephard, Sheppard, and Engle (2014).

## Appendix D: Distribution of $\beta$

Assuming returns are ordered as  $r_t = \{r_t^{(i)}, r_t^{(MKT)}, r_t^{(HML)}, r_t^{(SMB)}, r_t^{(MOM)}\}$ , i.e. the individual stock first and the risk factors afterwards with  $\Sigma$  its  $N \times N$  covariance matrix. Then, realized beta as defined in Equation (5.28), can be written as

$$\beta = (B\Sigma A')(A\Sigma A')^{-1},$$

with  $A = \begin{bmatrix} 0_{N-1} & I_{N-1} \end{bmatrix}$ , and  $B = \begin{bmatrix} 1 & 0'_{N-1} \end{bmatrix}$ . Here,  $I_N$  is an identity matrix of size  $N$  and  $0_N$  is a vector of zeros. To apply the delta method the derivative of  $\beta$  with respect to elements of  $\Sigma$  is needed.

Note that

$$d(A\Sigma A')^{-1} = -(A\Sigma A')^{-1}Ad\Sigma A'(A\Sigma A')^{-1}$$

and thus

$$\begin{aligned} d(B\Sigma A')(A\Sigma A')^{-1} &= Bd\Sigma A'(A\Sigma A')^{-1} + B\Sigma A'd(A\Sigma A')^{-1} \\ &= Bd\Sigma A'(A\Sigma A')^{-1} - B\Sigma A'(A\Sigma A')^{-1}Ad\Sigma A'(A\Sigma A')^{-1} \\ &= (B - B\Sigma A'(A\Sigma A')^{-1}A)d\Sigma A'(A\Sigma A')^{-1} \end{aligned}$$

Using  $\text{vec}(XYZ) = (Z' \otimes X)\text{vec}(Y)$ ,

$$\nabla = \frac{d\text{vec}(\beta)}{d\text{vec}(\Sigma)'} = ((A\Sigma A')^{-1}A \otimes (B - B\Sigma A'(A\Sigma A')^{-1}A))$$

Let  $\Psi_t = D_N \Pi_t D_N'$  is the distribution of  $\text{vec } S_t$ . Using the delta method we find that the MKernel estimate of  $\beta$ , denoted  $b$  has distribution

$$M^{1/5}(b - \beta) \rightarrow_{Ls} MN(0, \nabla\Psi\nabla'). \quad (5.40)$$

# Chapter 6

## Score-Driven Nelson Siegel: Hedging Long-Term Liabilities\*

### 6.1 Introduction

Hedging interest rate risk requires a model of how prices of bonds with different maturities react to shocks. There is a large literature that fits and models the time series of the term structure of bonds, returns and their yields. As the aim is to forecast a large cross-section of maturities, the literature imposes a factor structure on the yield curve to reduce the dimensionality of the problem. Most popular is the class of affine term structure models (Dai and Singleton, 2000; Duffee, 2002; Duffie and Kan, 1996), which is both tractable from an econometric perspective as well as consistent with no-arbitrage conditions from finance theory.

Nelson and Siegel (1987) propose a purely statistical model which has good empirical fit. The model has received quite some attention since Diebold, Rudebusch, and Boragan Aruoba (2006) who extend the standard model to a dynamic version, which describes the yield curve over time as a three factor model. The three factors are the well-known level, slope and curve factors, whose exact shape is determined by a single parameter  $\lambda$ .<sup>1</sup> As a further extension Christensen, Diebold, and Rudebusch (2011) showed that a simple additive constant term makes the model arbitrage free and a member of the affine class.

In this paper we use the Nelson-Siegel model for the purpose of risk management and specifically to hedge the risk of a long-term liability using shorter-term bonds. This empirical problem is of great interest to institutions like pension funds, who often have difficulty hedging long-term obligations. To this purpose, we transform the Nelson-Siegel to a factor model on excess bond returns, and obtain a model of the impact of shocks on bond prices. Since the Nelson Siegel model imposes a factor structure, hedging amounts to forming a portfolio with opposite factor exposure to the liability, and as such, for a successful hedge, a correct specification of the factor structure is vital. In the Nelson-Siegel

---

\*This chapter is co-authored with Peter Schotman (Maastricht University)

<sup>1</sup>For a textbook description, see Diebold and Rudebusch (2013).

model, the factor structure is fully controlled by the single  $\lambda$  parameter. This parameter has mostly been disregarded, and is often not even estimated, but set to a value chosen by the users. We focus on the  $\lambda$  and propose a Dynamic Conditional Score version of the model which provides an intuitive way to model time-variation in  $\lambda$ . Crucially, if  $\lambda$  is time-varying, the true factor exposure cannot be adequately hedged. We know of two other papers that allow for time-variation of  $\lambda$ . Koopman, Mallee, and Van der Wel (2010) estimate time-variation using an extended Kalman filter, while Creal, Koopman, and Lucas (2008) allow for time-variation as a function of the time-varying factor exposures.

Additionally, we allow for heteroskedasticity in the model. Bianchi, Mumtaz, and Surico (2009) model time-variation in the variances of the factors. Here, the factor loadings for the term structure are also used as weights for the volatility in the term structure. Koopman, Mallee, and Van der Wel (2010) assume a state-space model where the first principal component of the Nelson-Siegel error follows a GARCH(1,1) process. The same construction is used for the volatility of the measurement error of the state variables. That means that both the innovation and the measurement error are driven by a single GARCH(1,1) process. Hautsch and Ou (2012) propose a stochastic volatility specification of the Nelson-Siegel model. Similar to Koopman, Mallee, and Van der Wel (2010) we use a single GARCH model that drives volatility dynamics for all maturities.

In our model the time-variation in  $\lambda$  and the variance are interconnected as the covariance matrix directly influences the updates of  $\lambda$ . We find considerable time-variation in  $\lambda$  and show the importance of the heteroskedasticity. Importantly, when not taking into account the time-variation of the variance, the factor structure breaks down during the financial crisis of 2008:  $\lambda$  converges to zero, effectively making it a two-factor model. Taking into account the volatility,  $\lambda$  actually increases to levels not seen before. If not modelled, the large variation in returns simply obscures the factor structure. The relation between the factor loadings and volatility in our model prevents the factor structure from breaking down.

In an out-of-sample hedging exercise we show that the factor exposure of the hedge portfolio constructed using the Nelson Siegel with time-varying  $\lambda$  and volatility is reduced by almost 50% compared to the standard Nelson-Siegel model where both are constant. The gains mainly stem from the time-variation in  $\lambda$ , but also allowing for heteroskedasticity improves the hedging results further.

The remainder of the paper is organized as follows. In Section 6.2 we recap the Nelson-Siegel model and introduce how it can be used for hedging. Section 6.3 introduced the Dynamic Conditional Score version of the NS model with time-varying  $\lambda$ . Section 6.4 gives in-sample estimation results of the various models, which are used in an out-of-sample hedging exercise in Section 6.5.

## 6.2 Nelson Siegel model

Denote by  $y_t(\tau)$  the yields at time  $t$  for maturity  $\tau$ . At any time, the yield curve is some smooth function representing the yields as a function of maturity  $\tau$ . Nelson and Siegel (1987) (NS) provide a parsimonious factor description of these yields. The Diebold and Li (2006) formulation can be written as

$$y_t(\tau) = B(\tau)\tilde{f}_t, \quad (6.1)$$

where

$$B(\tau) = \begin{pmatrix} 1 \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} \\ \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \end{pmatrix}' \quad (6.2)$$

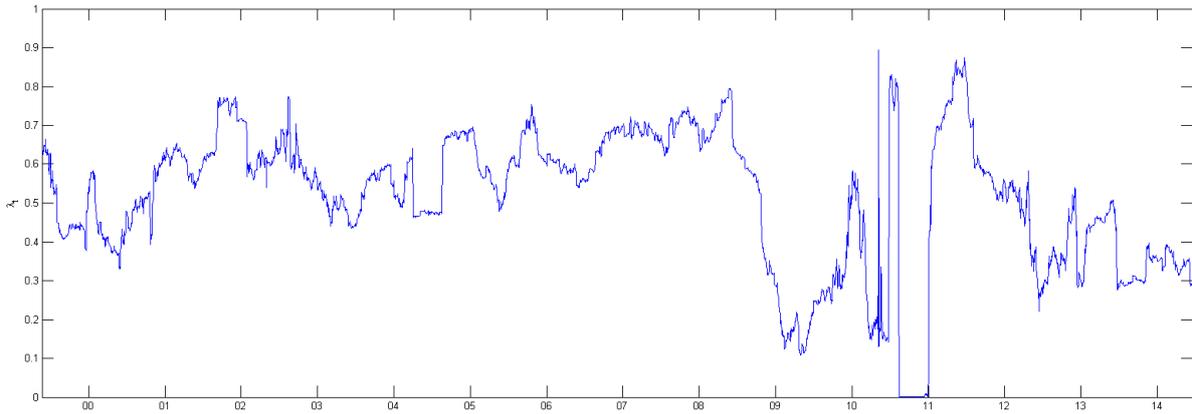
depends on the shape parameter  $\lambda > 0$ . The NS model is a three-factor model where the three factors are often interpreted as level, slope and curve respectively. The first component has constant factor loadings of 1. As such, it influences short and long term yields equally, and can be considered as an overall level curve. The second component converges to one as  $\tau \downarrow 0$  and converges to zero as  $\tau \rightarrow \infty$ . This component mostly affects short-term rates. The third factor converges to zero as  $\tau \downarrow 0$  and  $\tau \rightarrow \infty$ , but is concave in  $\tau$ . This component therefore mainly influences the medium-term rates.

Given a value for  $\lambda$  the factors are usually estimated by the cross-sectional regression model

$$y_t(\tau) = \sum_{i=1}^3 B_i(\tau)\tilde{f}_{i,t} + \epsilon_t(\tau), \quad (6.3)$$

for  $t = 1, \dots, T$ , where  $B_i(\tau)$  denotes the  $i$ -th column of  $B(\tau)$ . At each  $t$  the data consist of yields for  $K$  different maturities,  $\tau_j$ , ( $j = 1, \dots, K$ ). The estimated time series of  $\tilde{f}_t$  have empirically been shown to be strongly correlated over time, and are deemed forecastable. As such, Diebold and Li (2006) estimate the cross-sections of  $f_t$  and treat these estimates as time series to forecast future values of  $f_t$ , and thus the yield-curve, using a VAR model. Diebold, Rudebusch, and Boragan Aruoba (2006) and Koopman, Mallee, and Van der Wel (2010) put the model into a state-space framework, and estimate the dynamics using a Kalman filter. For hedging purposes, the dynamics of  $\tilde{f}_t$  are less important, since the aim is to obtain a portfolio that is orthogonal to the factor loadings, and hence not exposed to the factors.

In contrast to the factors  $\tilde{f}_t$ ,  $\lambda$  has not gained much attention. Generally, it is set to a constant value to maximize the curve at a certain maturity, without any estimation. However,  $\lambda$  is a crucial parameter, as it governs the exponential decay of the second and third term, and as such exposure to the different factors across maturities. Small values of  $\lambda$  produce slow decay, and vice versa. We are aware of only two efforts of modeling time-variation in  $\lambda$ . Koopman, Mallee, and Van der Wel (2010) allow  $\lambda$  to be time-varying in a state-space framework and Creal, Koopman, and Lucas (2008) let it evolve as a linear

**Figure 6.1:** Time-variation in  $\lambda$ 

*Note:* This graph plots the rolling-window NLS estimates of  $\lambda$  using 100 days of data.

combination of the  $f_t$ .

In this paper the time-variation of  $\lambda$  is of main interest. As a preliminary indication of the importance of modelling  $\lambda$  as a time-varying parameter we estimate the parameter on a rolling-window of 100 days using NLS, plotted in Figure 6.1. The data is described in detail in Section 6.4. The figure shows there is great variability in  $\lambda_t$  with typical values ranging between 0.1 and 0.8, whereas a full sample estimate yields  $\lambda = 0.3834$ . The factor structure appears to largely break down during the financial crisis. Not adequately modelling this leads to suboptimal fit where hedge portfolios will have unintended factor exposure.

### 6.2.1 Factor Hedging with NS

Our main empirical focus is on hedging long term liabilities. Suppose the yield curve can be described by the Nelson-Siegel model of Equation (6.1). The NS model implies discount factors of the form

$$\delta_t(\tau) = \exp(-\tau B(\tau) \tilde{f}_t). \quad (6.4)$$

These discount factors can be used for the valuation of a stream of liabilities with cash-flows  $Z_t(\tau)$ ,

$$L_t = \int_0^{\mathcal{T}} Z_t(\tau) \delta_t(\tau) d\tau. \quad (6.5)$$

The exposure of the liabilities with respect to the factors is given by

$$\frac{\partial L_t}{\partial \tilde{f}_t} = - \int_0^{\mathcal{T}} w_t(\tau) \tau B(\tau) d\tau \equiv -D_t, \quad (6.6)$$

where the weight of each of the liability cash-flows is given by

$$w_t(\tau) = \frac{Z_t(\tau)\delta_t(\tau)}{L_t}. \quad (6.7)$$

As  $B_1(\tau) = 1$ , the first element in  $D_t$  is the duration of the liabilities. Hedging the level factor in Nelson-Siegel is therefore equivalent to duration hedging. This result was similarly derived in Diebold, Ji, and Li (2004), and following them, we refer to  $D_t$  as the generalized duration of the liabilities. The second and third element are in this case exposures to the slope and curvature factors.

In practice hedging focuses on the first factor, ignoring the slope and curve factor, despite wide acceptance of the presence of at least three factors in interest rate curves. In our sample, the first factor explains about 82% of the variation in returns. With a second factor a total of 93% of variation is explained, and a third factor raises this to over 96%. Addressing only the first factor therefore inherently limits the hedging potential. Portfolios can easily be formed to hedge exposure to all three factors. Given that the factors are accurate, the risk of the hedge portfolio should be significantly reduced.

### 6.2.2 Estimation on yields versus returns

Since we are interested in reducing the risk of a hedge portfolio, ultimately we are interested in returns on the portfolio, not in changes in yield. We therefore propose to model factor structure of returns instead of yields. Importantly, excess returns have the same factor structure as yields.

Log-prices have the structure

$$p_t(\tau) = -\tau y_t(\tau) = -b(\tau)\tilde{f}_t, \quad (6.8)$$

where  $b(\tau) = \tau B(\tau)$ . Excess returns over a period of length  $h$  are defined as

$$\begin{aligned} r_{t+h}(\tau) &= p_{t+h}(\tau) - p_t(\tau + h) + p_t(h) \\ &= -b(\tau)\tilde{f}_{t+h} + (b(\tau + h) - b(h))\tilde{f}_t. \end{aligned} \quad (6.9)$$

Using the property

$$\begin{aligned} b(\tau + h) - b(h) &= b(\tau) \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\lambda h} & \lambda h e^{-\lambda h} \\ 0 & 0 & e^{-\lambda h} \end{pmatrix} \\ &\equiv b(\tau)A, \end{aligned} \quad (6.10)$$

where  $A$  does not depend on maturity  $\tau$ , excess returns can be written as

$$\begin{aligned} r_{t+h}(\tau) &= -b(\tau)(\tilde{f}_{t+h} - Af_t) \\ &= b(\tau)f_t, \end{aligned} \tag{6.11}$$

with  $f_t = -(\tilde{f}_{t+h} - Af_t)$ . As such, returns standardized by their maturity have the same factor structure as yields

$$\rho_{t+h} \equiv \frac{r_{t+h}(\tau)}{\tau} = B(\tau)f_t. \tag{6.12}$$

The standardization by  $\tau$  has the additional benefit of taking away a large part of the cross-sectional heteroskedasticity, since volatility of bond returns increases with maturity. We estimate all our models on these standardized returns. The fact that excess returns and yields have the same factor structure, is not just a coincidence for the NS model, but a general property of all affine term structure models. Although the transformation leaves the factor structure unchanged, it will affect the properties of the error term  $\epsilon_t$  of Equation (6.3). When estimated on yield levels, the errors are usually strongly autocorrelated. However, the errors of the model applied to returns will most likely not exhibit autocorrelation.

With constant factor loadings, a hedging problem would reduce to hedging the factor risk. If the liabilities have a factor exposure  $B^*$ , then a bond portfolio with the same factor exposure should be a perfect hedge. The need for a factor model does not, however, imply that factor loadings must be constant. Suppose the parameters  $\lambda$  in the factor loadings follow a stochastic process  $\lambda_t$ . Then the hedge portfolio should also consider changes in the factor structure. We can locally hedge using the current factors  $B_t$ , but the hedge needs to be rebalanced if factor loadings change. If they are different the return on the hedge portfolio could be exposed to factor risk.

$$\begin{aligned} r_{t+h}(\tau) &= -b_{t+h}(\tau)\tilde{f}_{t+h} - b_t(\tau)A_t\tilde{f}_t \\ &= b_t(\tau)f_{t+h} - (b_{t+h}(\tau) - b_t(\tau))\tilde{f}_{t+h}. \end{aligned} \tag{6.13}$$

A portfolio that is orthogonal to  $b_t(\tau)$  is exposed to the change in the factor structure, except when it is still orthogonal to  $b_{t+h}(\tau)$ . The hedge portfolio can be updated every period depending on a change in  $b_t(\tau)$ . Still the hedging problem becomes far more difficult, since the time-varying factor loadings will be estimated with less precision than fixed factor loadings. For fixed factor loadings, estimation accuracy will be close to perfect with a long enough sample of reasonably frequent observations. With time-varying factor loadings estimation error becomes important. Estimation error implies that the hedge will be imperfect, because the estimated factor loadings differ from the true factor loadings, which exposes the portfolio to factor risk.

Finally, time-variation in factor loadings could also be a sign that the factor model does not have a sufficient number of factors. Suppose the factor loadings depend on a

time-varying parameter  $\lambda_t$

$$b_t(\tau) = b(\lambda_t, \tau). \quad (6.14)$$

Then we can linearise the model for excess returns as

$$\begin{aligned} r_{t+h}(\tau) &= b(\lambda_t, \tau) f_{t+h} \\ &= \tilde{b}(\tau) f_{t+h} + C(\tau)(\lambda_t - \bar{\lambda}) f_{t+h}, \end{aligned} \quad (6.15)$$

where  $C$  is the derivative of  $b_t$  with respect to  $\lambda_t$  evaluated at  $\bar{\lambda}$ . This is again a factor model, but it has constant factor loadings and additional factors  $(\lambda_t - \bar{\lambda}) f_{t+h}$ . In the Nelson-Siegel model with just a single parameter  $\lambda_t$  and three-factors, the linearisation would lead to a model with six factors, since all elements in  $f_{t+h}$  need to be scaled by  $\lambda_t - \bar{\lambda}$ . It seems much more parsimonious to model the variation in  $\lambda_t$  directly, instead of specifying a larger factor model.

### 6.3 DCS-NS Model

We propose a version of the Nelson-Siegel model with time-variation in  $\lambda_t$ , in which its dynamics are governed by a Dynamic Conditional Score (DCS) model. DCS models were first proposed in their most general form in Creal, Koopman, and Lucas (2013). The class of models is now often used empirically, and theoretical results are derived in, amongst others, Blasques, Koopman, and Lucas (2014a) and Blasques, Koopman, and Lucas (2014b).

In the class of DCS models, the dynamics of parameters are driven by the score of the likelihood with respect to that parameter and provide a natural way to update a wide variety of (latent) parameters, when a functional form for their dynamics are not directly apparent. The DCS principle turns out to nest a lot of the models commonly used in time-series analysis. For instance, when assuming a gaussian likelihood, dynamics in the mean are given by an ARMA model, and the dynamics of volatility are given by the well-known GARCH model, which we will also use to allow for time-varying volatility.

Let  $r_t$  and  $\rho_t$  be  $K$ -vectors of (standardized) excess returns with different maturities. Consider the conditional observation density of returns  $p(r_t|\lambda_t)$ , and let  $\nabla_t$  the score with respect to  $\lambda_t$ ,  $\nabla_t = \frac{\partial \log p(r_t|\lambda_t)}{\partial \lambda_t}$ . Then the DCS model is defined as

$$\lambda_{t+1} = \phi_0(1 - \phi_1) + \phi_1 \lambda_t + \phi_2 s_t, \quad (6.16)$$

where  $s_t = S_t \nabla_t$  is the score times an appropriate scaling function. Time-variation of the parameter is driven by the scaled score of the parameter and, as such, the dynamics of the parameters are linked to the shape of the density. Intuitively, when the score is negative the likelihood is improved when the parameter is decreased, and the DCS updates the parameter in that direction. The NS-DCS model nests a model of constant  $\lambda$ .

The choice of  $S_t$  delivers flexibility in how the score  $\nabla_t$  updates the parameter. A natural candidate is a form of scaling that depends on the variance of the score:  $S_t = \mathcal{J}_{t|t-1}^{-1}$ , where  $\mathcal{J}_{t|t-1} = E_{t-1}[\nabla_t \nabla_t']$ . Another option to consider is  $S_t = \mathcal{J}'_{t|t-1} \mathcal{J}_{t|t-1} = \mathcal{J}_{t|t-1}^{-1}$ , such that it is the square root matrix of the inverse information matrix, or no scaling at all,  $S_t$  being the identity matrix. We choose  $S_t = \mathcal{J}_{t|t-1}$ . In our case, as  $\lambda$  is a scalar parameter,  $S_t$  is scalar too.

We use the DCS model to introduce time-variation in  $\lambda_t$  in the NS model. Let  $B_t$  denote a  $(K \times 3)$  time-varying matrix with as rows time-varying versions of  $B(\tau)$ , where the  $t$  subscript relates to time-variation in  $\lambda_t$ . We consider the following specification, which we call NS-DCS:

$$\begin{aligned}\rho_t &= B_t f_t + \epsilon_t \\ \lambda_t &= \phi_0(1 - \phi_1) + \phi_1 \lambda_{t-1} + \phi_2 s_{t-1},\end{aligned}\tag{6.17}$$

where we assume  $\epsilon_t \sim N(0, \Sigma_t)$ . The  $(K \times K)$  error covariance matrix will be specified below. For this model we have  $\nabla_t = \epsilon_t' \Sigma_t^{-1} G_t f_t$  and  $\mathcal{J}_{t|t-1} = f_t' G_t' \Sigma_t^{-1} G_t f_t$ , where  $G_t \equiv \frac{\partial B_t}{\partial \lambda_t}$ . The log-likelihood function is

$$\mathcal{L}_N = -\frac{1}{2} \ln |\Sigma_t| - \frac{1}{2} \epsilon_t' \Sigma_t^{-1} \epsilon_t,\tag{6.18}$$

Importantly, we do not model the time-variation in factors  $f_t$ . We simply define  $f_t = (B_t' B_t)^{-1} B_t' \rho_t$ . The reason for this is two-fold. First, as note before the  $f_t$  have unit roots, and as such the optimal forecast is the random walk forecast. Second, for hedging and forecasting purposes,  $B_t$  is of interest and not  $f_t$ . By concentrating  $f_t$  out of the likelihood we do not have to consider the dependence between  $\lambda_t$  and  $f_t$  in the observation density and in the specification of the DCS model. This greatly simplifies the model, and makes it far easier to estimate. Moreover, it makes our analysis robust against any misspecification in the time-series process of  $f_t$ .

Finally, we want to discuss the choice for a DCS model. An alternative way to model time-variation in the latent parameter is estimation by Kalman Filter of a state space model. The two are fundamentally different, and their difference is similar to Stochastic Volatility models versus GARCH-type models, where DCS models are similar to GARCH. We prefer the DCS method for a number of reasons. The main reason is that in the DCS model,  $\lambda_t$  is known conditional on  $\mathcal{F}_{t-1}$  as it is a deterministic function. As such, the likelihood is known analytically, making estimation straightforward. For a state-space model,  $\lambda_t$  is not deterministic conditional on the past, it remains unobserved. Evaluation of the log-likelihood therefore requires integrating over the path space of  $\lambda_t$ , and is much more computationally demanding.

Second, recent results by Blasques, Koopman, and Lucas (2014b) show that only parameter updates based on the score always reduce local Kullback-Leibler (KB) divergence between the true conditional density and the model implied conditional density. Un-

der some regularity conditions on the smoothness of the model, it leads to considerable improvements in global KB divergence as well. DCS models are thus optimal from an information-theoretic perspective.

### 6.3.1 Time-varying volatility

The ‘innovation’  $s_t$  is a function of  $\Sigma_t$ , the covariance matrix of  $\epsilon_t$ . Keeping the covariance matrix constant or time-varying is similar to OLS versus GLS. When we allow for dynamics in  $\Sigma_t$ , the score is down-weighted more in high volatility times, reducing the impact of large idiosyncratic errors. As such we consider two versions of the DCS. In the base case we consider  $\Sigma = \sigma^2 I_K$  as a constant. In the more general model we allow for a common GARCH-type process. In order to have a parsimonious specification we let  $\Sigma_t = \sigma_t^2 I_K$ , where

$$\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta \Gamma' \epsilon_{t-1}' \epsilon_{t-1} \Gamma, \quad (6.19)$$

and  $\Gamma$  is a loading vector of the cross-sectional dimension such that  $\Gamma' \Gamma = 1$ . The loading vector may be estimated or, as we do, set to equal weights for each maturity, i.e.  $\Gamma = \iota / \sqrt{K}$ . The volatility dynamics of all maturities are thus governed by a single GARCH process. Any full multivariate model would quickly suffer from the curse of dimensionality as the cross-sectional dimension typically exceeds ten or fifteen. This leads to two self-explanatory additional models, the NS-GARCH and NS-DCS-GARCH.

## 6.4 Empirical Analysis

### 6.4.1 Data

We obtain daily data on euro swap interest rates with maturities 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40 and 50 years from Datastream. Data is available from January 1999, except for the 40 and 50 years rates, which are available from March 2005 onwards. All rates are ask rates quoted at the close of London business. Euro swap rates are quoted on an annual bond 30/360 basis against 6 months Euribor with the exception of the 1 year rate which is quoted against 3 months Euribor. The main advantage of using swaps is that they are available and relatively liquid until long maturities, which allows us to verify the accuracy of the model for long-term liabilities.<sup>2</sup> The series are interpolated using a local cubic interpolation to obtain rates with annual intervals between maturities. We then use a bootstrap method to construct the yield curve. The resulting panel of yields is plotted in Figure 6.2.

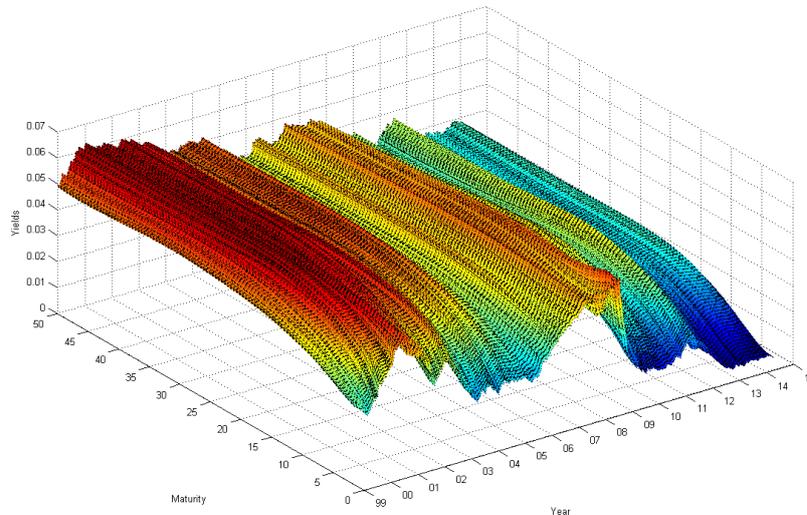
The next step is to compute returns. We first obtain the price from the yields as

$$P_t(\tau) = e^{-\tau y_t(\tau)} \quad (6.20)$$

---

<sup>2</sup>On April 1st 2004 there is almost surely a data error for the 6Y rate, which is exactly equal to the 5Y rate. This observation distorts all estimations and is replaced with an interpolated value.

Figure 6.2: Yield Curves



*Note:* This figure plots the yield curve of our euroswap data evolving over time for the different maturities.

and obtain returns as 100 times the log-difference over time. We subtract the risk-free rate defined as  $r_t^f = y_t(1)/252$ , the shortest-term yield.

In the remainder of the paper we will estimate the model on two different samples. First we use the full cross-section of seventeen maturities, and second we use a limited cross-section, with only the maturities up until twenty years. There are good reasons for excluding maturities longer than twenty years. The most important reason is the general assumption that the so-called ‘last liquid point’ of the market rates is at the twenty year maturity.<sup>3</sup> The prices of bonds with maturities greater than this are severely affected by their limited liquidity, which significantly impacts the yield curve and therefore estimates of the NS-model.

### 6.4.2 In-sample Results

Estimation results are reported in Table 6.1. The results of the full and limited cross-section are qualitatively similar, apart from the limited cross-section having higher estimates of  $\lambda$ , which puts more exposure on short maturities. On the full cross-section the shape parameter in the standard NS model is estimated at  $\lambda = 0.3834$ , putting the maximum of the curve factor between 4 and 5 years. Due to the presence of heteroskedasticity, the NS-GARCH has a far larger estimate, 0.5502, which maxes the curve out between 3 and 4 years. A quick look at the rolling window estimates of Figure 6.1 explains the difference between the constant and GARCH volatility models, as  $\lambda$  is lowest during the financial crisis when volatility was high.

<sup>3</sup>This is the point assumed by the European Insurance and Occupational Pensions Authority (EIOPA)

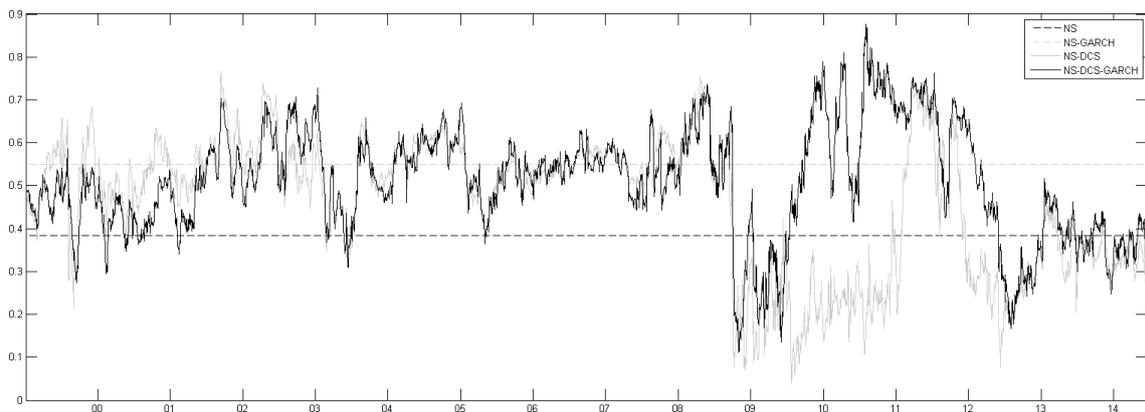
**Table 6.1:** In-Sample Parameter Estimates

	Full Cross-Section				Limited Cross-Section			
	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH	NS	NS-DCS	NS-GARCH	NS-DCS-GARCH
$\lambda$	0.3834 (0.0446)		0.5502 (0.0204)		0.5742 (0.0228)		0.6372 (0.0233)	
$\phi_0$		0.4108 (0.0262)		0.4161 (0.0221)		0.5799 (0.0130)		0.6531 (0.0278)
$\phi_1$		0.9888 (0.0955)		0.9859 (0.0759)		0.9659 (0.0759)		0.9614 (0.0716)
$\phi_2$		0.0345 (0.0188)		0.0208 (0.0102)		0.0288 (0.0082)		0.0145 (0.0045)
$\sigma^2(\times 10^5)$	6.7973 (0.6929)	6.3411 (0.6809)			5.0114 (0.5587)	4.9419 (0.5349)		
$\omega(\times 10^7)$			6.1422 (8.3806)	6.5799 (5.2184)			7.0818 (6.2002)	5.0953 (4.8337)
$\alpha$			0.8737 (0.1053)	0.8760 (0.0841)			0.8794 (0.1067)	0.8616 (0.0612)
$\beta$			0.1226 (0.0987)	0.1193 (0.0755)			0.1191 (0.0935)	0.1346 (0.0602)
$LL(\times 10^{-5})$	2.2720	2.2954	2.4945	2.5029	1.8158	1.8194	1.9748	1.9998

*Note:* This table provides in-sample parameter estimates for the four different models, both on the full cross-section of all seventeen maturities, as well as the model estimated on the limited cross-section which only includes bonds up until maturity of twenty years. Newey-West standard errors (22 lags) in brackets.

Next consider the DCS models. The DCS models improve over their standard counterparts: both the NS-DCS and NS-DCS-GARCH have a much larger value for the log-likelihood. The  $\lambda$ 's are highly persistent with autoregressive coefficient over 0.98. Importantly, the coefficient of the score innovation,  $\phi_2$ , is significantly different from zero. The models therefore produce significant dynamics in  $\lambda$ .

To illustrate the dynamics of the DCS model we plot the estimated  $\lambda_t$  in Figure 6.3 for the different models on the full cross-section. The main series of interest are the  $\lambda_t$  estimates of the NS-DCS and NS-DCS-GARCH. Interestingly, up until the second half of 2008, volatility does not play a large role and the fitted  $\lambda_t$  of both models are very similar, and more over, variation is limited. However, during the crisis period, the two models completely diverge. The NS-DCS' estimates of  $\lambda$  go down, with an absolute minimum at 0.04, which effectively reduces the model to two factors. Interestingly, the  $\lambda_t$  implied by the NS-DCS-GARCH actually increases to levels not seen in the entirety of the first half of our sample. After most of the volatility has died down, the two DCS models converge again.

**Figure 6.3:** Time Series of  $\lambda_t$  for various models.

## 6.5 Hedging long-term liabilities

The aim of this section is to use the various forms of the Nelson-Siegel model introduced in this paper for a hedging exercise. We use the four models highlighted in the previous section, and contrast their results with the hedging performance of simple duration hedging, in which we only hedge the level factor.

### 6.5.1 Setup

Let  $w_\tau$  denote the weight in bond with maturity  $\tau$ . We consider the following problem at any point in time  $t$ :

$$w_{50} = -1 \quad (6.21a)$$

$$\sum_i w_i = 0 \quad (6.21b)$$

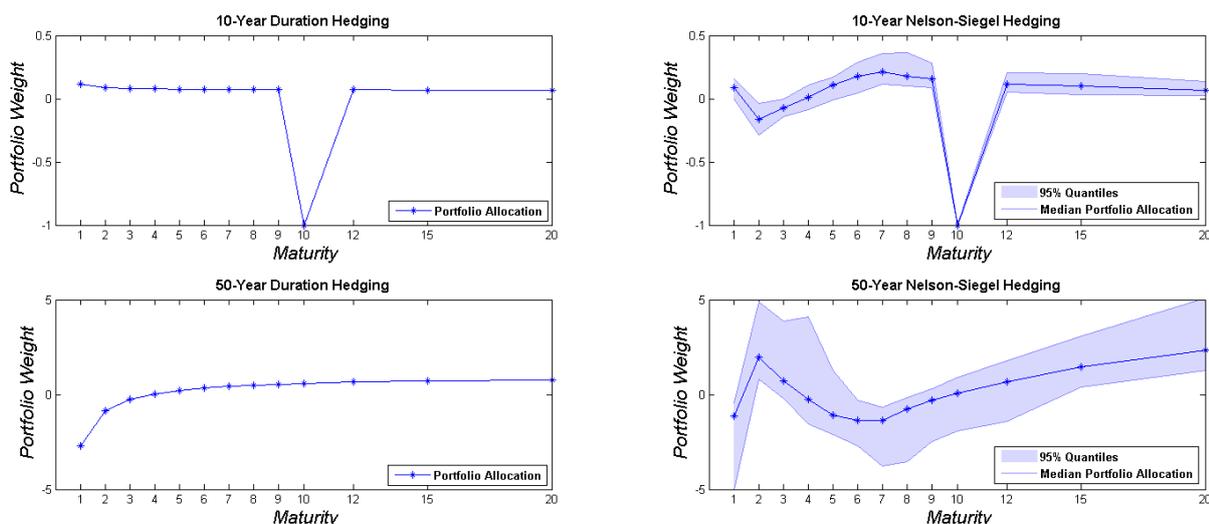
$$\sum_i w_i \tau_i B_k(\tau_i) = 0 \quad \forall k = 1, 2, 3 \quad (6.21c)$$

$$\min w' \mathcal{T}' \Sigma_t \mathcal{T} w, \quad (6.21d)$$

where  $\mathcal{T}$  is the vector of maturities  $\tau_i$ . Line *a* shows a short position in the long-maturity bond. Line *b* requires the portfolio to be self-financing, and line *c* implies zero factor exposure of the portfolio. If the factor loadings are estimated correctly, that is,  $\lambda_t$  is equal to the true  $\lambda_t$ , the only remaining risk is the idiosyncratic risk of the individual bonds. Finally, line *d* chooses the portfolio with minimum idiosyncratic risk out of the infinite set of portfolios satisfying the first three constraints. Note that  $\tau$  re-enters in the last two equations, as we now look at the true returns, not those standardized by  $\tau$ .

The hedging of long-term liabilities is of great interest for practitioners such as pension funds, but it also serves as a great illustration of the qualities and importance of accurately

Figure 6.4: Hedging Long-Term Liabilities



*Note:* The graphs depict the importance of hedging all factors for long-term hedging. The top graphs show the portfolio allocation for hedging a bond with 10-year maturity, and the bottom graphs show the portfolio allocation for hedging a 50-year bond. The two graphs on the left show Duration hedging and the right graphs show the results of hedging all three factors using  $\lambda_t$  from the NCS-DS model. For the three-factor hedge we show the empirical distribution of the portfolio weights over time.

modeling the factor structure. Hedging long-term liabilities is a much more intricate problem than short to medium term liabilities. To hedge a ten-year bond, taking an opposite position in a number of bonds with close maturity will work quite well as their factor loadings are similar to the ten year bond. For long-term bonds, such as a fifty year bond, no liquid bonds with close maturity exist, making this infeasible.

Hence, the 50-year bond needs to be hedged using large positions in bonds with much shorter maturities. In terms of the Nelson-Siegel model, the 50-year maturity bond essentially only loads on the level factor. However, the slope and curve factor play a crucial role in hedging the long-term liability, since the portfolio of bonds used to hedge the long-position do load significantly on the slope and curve factor. These shorter bonds have to be chosen carefully, since the overall hedge portfolio also needs to be neutral to these factors.

To illustrate the three-factor hedge for short and long-term liabilities we plot the empirical distribution of the ten- and fifty-year hedge portfolios in Figure 6.4. The top graphs depict the weights in the ten-year hedge portfolio for both duration hedging and the portfolios that hedge all three NS factors, where  $\lambda_t$  is derived from the NCS-DS model. The portfolios do not differ much, as the factor loadings change gradually over the neighboring maturities, and neutralizing the slope and level factor requires a marginal change in the portfolios. Moreover, there is very little time-variation, as changes in  $\lambda$  only affect the second and third factor, which have limited impact on portfolio composition.

On the other hand, the dynamics in the 50-year bond hedge are far more extreme. First the duration hedging portfolio takes a large positive position in the bond with closest maturity (20 in this case), but spreads out the weight over different maturities to diversify away as much of the idiosyncratic risk as possible. The three-factor hedge for the 50-year maturity shows that portfolio composition is, in fact, largely determined by the need to neutralize the second and third factor. Moreover, there is a lot more time-variation in the portfolio composition, as the changing  $\lambda$  only affects the slope and curve factors, which play a limited role for short-term hedges.

### 6.5.2 Hedging Results

We set up two out-of-sample forecasting exercises. In both we re-estimate the parameters of the model every Friday using a rolling window of 1000 observations, and use these parameters to make daily forecasts of  $\lambda_t$ . These forecasts are used to solve the system of equations of the minimum risk hedge portfolio given by Equations (26a-d). However, in one set we use the full cross-section of the 17 maturities including the 50 year maturity itself. In the second exercise, we use only the 13 maturities up until and including 20 years. This is an economically important restriction, as bonds with longer maturities are not as liquid, and the market is simply not deep enough to accommodate the needs of pension funds. Here, the 50 year maturity is out-of-sample both in terms of the time-series and the cross-sectional dimensional. In this case we do not estimate the full covariance matrix which makes minimizing (6.21d) infeasible. Instead, we minimize the variance of the portfolio of the first 13 maturities, with their combined weight standardized to one.

We report the mean absolute hedging error (MAE) as well as the variance of the hedging error (Var). The better the hedge, the smaller these two quantities will be. We also report the mean hedging error (Mean). If the mean is positive, the hedging portfolio is earning a higher average return than the target asset which is the more desirable of deviations from zero. The results are reported in Table 6.2. We report results for the full out-of-sample period, as well as three sub-periods, corresponding to the periods before, during and after the financial crisis. The top half shows the results for the hedging portfolios constructed using  $\lambda$  forecasts based on the sample including the maturities over twenty years, whereas the bottom panel is based on forecasts using only data with maturities up to twenty years.

The results for the full and partial cross-section are qualitatively similar, and models on the full cross-section perform slightly better. All models have average return close to zero, so none of the models has systematic bias. Therefore, what matters for hedging purposes is the variance of the hedging errors. The MAE and portfolio return variance show differences across the different models. For instance, looking at the full cross-section on the full sample, we see that the variance of hedging errors is 0.58 for duration hedging, 0.37 for the standard Nelson-Siegel and 0.26 for the DCS version. To put things into perspective, the actual variance of the 50-year maturity return is 5.30 over this period.

## 6. SCORE-DRIVEN NELSON SIEGEL

**Table 6.2:** Daily hedging error descriptives

	Duration Hedging	NS	NS-DCS	NS-GARCH	NS-DCS- GARCH
Full Cross-Section					
Full Sample					
Mean	-0.0024	0.0007	-0.0021	0.0004	0.0020
MAE	0.4053	0.3091	0.2009	0.2964	0.1997
Var	0.5876	0.3691	0.2570	0.3366	0.2215
2003-2007					
Mean	0.0061	0.0039	0.0039	0.0035	0.0086
MAE	0.1796	0.1595	0.1259	0.1537	0.1211
Var	0.1577	0.1654	0.1279	0.1563	0.1064
2008-2009					
Mean	0.0165	0.0097	-0.0062	0.0140	0.0049
MAE	0.8106	0.5531	0.3673	0.5829	0.2927
Var	1.7106	0.8066	0.5052	0.8873	0.4167
2010-2014					
Mean	-0.0207	-0.0070	-0.0070	-0.0093	-0.0068
MAE	0.4817	0.3708	0.2652	0.3830	0.3035
Var	0.5770	0.4067	0.2848	0.4262	0.3736
Limited Cross-Section					
Full Sample					
Mean	0.0031	0.0004	0.0042	0.0003	0.0019
MAE	0.4284	0.3174	0.2165	0.3193	0.2333
Var	0.6234	0.3892	0.3281	0.3947	0.3217
2003-2007					
Mean	0.0056	0.0035	0.0036	0.0034	0.0085
MAE	0.1976	0.1529	0.1238	0.1511	0.1282
Var	0.1899	0.1553	0.1250	0.1533	0.1047
2008-2009					
Mean	0.0211	0.0145	0.0031	0.0154	0.0116
MAE	0.8618	0.5867	0.3646	0.5944	0.3204
Var	1.8961	0.8992	0.6519	0.9245	0.5141
2010-2014					
Mean	-0.0196	-0.0095	0.0054	-0.0101	-0.0100
MAE	0.5247	0.3847	0.2913	0.3883	0.3163
Var	0.6110	0.4288	0.3516	0.4341	0.3967

*Note:* This table gives descriptives of the hedging performance of the different models in sub-periods. The Mean is the average hedging error, the MAE is the mean absolute hedging error, and the Var is the variance of the hedge portfolio returns. The top panel shows results when the models are estimated on the full cross-section of maturities, and all are used to hedge the 50-year maturity, the bottom panel shows the results when optimization occurs over assets with a maximum maturity of 20 years.

Duration hedging therefore takes out 89% of the variance and the NS improves by hedging 93% of the variance. Importantly, the NS-DCS has additional non-trivial improvements, and hedges 95% of the variance. Because of the poor performance of standard duration hedging, we will not discuss it further.

For the standard NS, the inclusion of time-varying volatility does not necessarily lead to superior performance. The NS-GARCH performs better in the early sample, whereas the NS model performs better in the latter part of the sample. This is in line with the value of  $\lambda$  implied by our DCS models in Figure 6.3. There the DCS-NS shows that  $\lambda$  is high in early sample, and decreases during the crises. The NS-GARCH estimate of  $\lambda$  is typically higher than the NS estimate, and therefore more accurate in the beginning of the sample.

The DCS models improve on the standard models without exception. Overall, the NS-DCS-(GARCH) lower the MAE and variance by approximately 30% compared to the NS(-GARCH). This is clearly an economically significant improvement. The NS-DCS-GARCH performs slightly better in the early sample and during the financial crisis, while the NS-DCS beats the time-varying volatility variant in the final part of our sample. A possible explanation for the latter result is that yields have hit the zero-lower bound in recent years, and variance is artificially low.

In the first part of the sample DCS' improvements are only about 20%, while during the crisis improvements are 40% to 50% and they are still about 30% after the crisis. As would be expected, variation in  $\lambda$  is greater in highly dynamic markets, and the DCS models offer the greatest improvements over the constant  $\lambda$  models. However, even in tranquil times the DCS offers improvements over the constant model.

Although the in-sample likelihood improvement of time-varying volatility is large, for our empirical application the gains are small. The reverse is true for the time-variation in  $\lambda$ , where the DCS model offers smaller likelihood improvements, but the empirical applications shows that benefits can be very large.

## 6.6 Conclusion

We have proposed a Dynamic Conditional Score version of the Nelson-Siegel in which time-variation of the shape-parameter  $\lambda$  is modeled. The parameter is of great importance as it completely specifies the factor loadings of bonds at different maturities. We show that the time-variation in the parameter is large, and assuming it to be constant is too limiting. In-sample estimation shows the parameter varies between 0.1 and 0.9, and even within a year the parameter has a range of up to 0.6. We document that the parameter is relatively stable in the beginning of the sample, but becomes highly volatile during the financial crisis.

In our empirical application, we use the DCS version of the NS model to hedge long-term liabilities using euro-swap data. We find that allowing for time-variation in  $\lambda$  leads

to significant improvements in hedging performance, greatly reducing the mean absolute hedging error and hedge portfolio variance. Specifically, hedging error variance is reduced by an additional 50% compared to the NS model's improvement over duration hedging. The variation in the shape parameter is of far greater importance than taking into account time-varying volatility, which only offers small improvements, both in the constant and DCS  $\lambda$  setting.

## Appendix A: Derivations of the Score and Information

In this section we show the derivations of the score and information which are used for the DCS model in (6.17). Let  $\epsilon_t = y_t - B_t f_t$ . Assume  $\epsilon_t \sim N(0, \Sigma_t)$ . Then the likelihood is

$$\mathcal{L}_N = -\frac{1}{2} \ln |\Sigma_t| - \frac{1}{2} \epsilon_t' \Sigma_t^{-1} \epsilon_t. \quad (6.22)$$

Defining  $G_t \equiv \frac{\partial B_t}{\partial \lambda_t} = \begin{pmatrix} 0 & \frac{-b_{t,3}}{\lambda_t} & \tau b_{t,2} - (\tau + \frac{1}{\lambda_t}) b_{t,3} \end{pmatrix}$  and  $H_t \equiv \frac{\partial^2 B_t}{\partial \lambda_t^2}$  the scores are

$$\begin{aligned} \nabla_t(\lambda_t) &= \frac{\partial \mathcal{L}_N}{\partial \lambda_t} = r_t' \Sigma^{-1} G_t f_t - f_t' B_t' \Sigma^{-1} G_t f_t \\ &= \epsilon_t \Sigma^{-1} G_t f_t. \end{aligned} \quad (6.23)$$

The second order derivatives are

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_N}{\partial \lambda_t^2} &= r_t \Sigma^{-1} H_t f_t - f_t' G_t' \Sigma^{-1} G_t f_t - f_t' B_t' \Sigma^{-1} H_t f_t \\ &= \epsilon_t \Sigma^{-1} H_t f_t - f_t' G_t \Sigma^{-1} G_t f_t \end{aligned} \quad (6.24)$$

from which we find the information matrix as

$$\begin{aligned} \mathcal{J}_{t|t-1}(\lambda_t) &= -\mathbb{E} \left( \frac{\partial^2 \mathcal{L}_N}{\partial \lambda_t^2} \right) \\ &= f_t' G_t' \Sigma^{-1} G_t f_t. \end{aligned} \quad (6.25)$$

and thus

$$\mathcal{J}_{t|t-1}(\lambda_t) = \sqrt{\frac{1}{f_t' G_t' \Sigma^{-1} G_t f_t}}. \quad (6.26)$$

# References

- ACHARYA, V., R. ENGLE, AND M. RICHARDSON (2012): “Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks,” *The American Economic Review*, 102(3), 59–64.
- ACHARYA, V., L. PEDERSEN, T. PHILIPPON, AND M. RICHARDSON (2010): “Measuring Systemic Risk,” *Working paper*.
- ADRIAN, T., AND M. K. BRUNNERMEIER (2014): “CoVaR,” *Working paper*.
- AIELLI, G. P. (2013): “Dynamic Conditional Correlation: on Properties and Estimation,” *Journal of Business & Economic Statistics*, 31(3).
- AÏT-SAHALIA, Y., J. FAN, AND D. XIU (2010): “High-Frequency Covariance Estimates with Noisy and Asynchronous Financial Data,” *Journal of the American Statistical Association*, 105(492), 1504–1517.
- AÏT-SAHALIA, Y., AND L. MANCINI (2008): “Out of Sample Forecasts of Quadratic Variation,” *Journal of Econometrics*, 147(1), 17–33.
- ANDERSEN, T. G., AND T. BOLLERSLEV (1998): “Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts,” *International Economic Review*, 39(4), 885–905.
- ANDERSEN, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2013): “Financial Risk Measurement for Financial Risk Management,” in *Handbook of the Economics of Finance* (eds. G. Constantinides, M. Harris and R. Stulz), pp. 1127–1220. Elsevier Inc.
- ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2007): “Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility,” *Review of Economics and Statistics*, 89(4), 701–720.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71(2), 579–625.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND G. WU (2006): “Realized Beta: Persistence and Predictability,” *Advances in Econometrics*, 20, 1–39.

## REFERENCES

---

- ANDERSEN, T. G., T. BOLLERSLEV, P. FREDERIKSEN, AND M. Ø. NIELSEN (2010): “Continuous-Time Models, Realized Volatilities, and Testable Distributional Implications for Daily Stock Returns,” *Journal of Applied Econometrics*, 25(2), 233–261.
- ANDERSEN, T. G., T. BOLLERSLEV, AND N. MEDDAHI (2004): “Analytical Evaluation of Volatility Forecasts,” *International Economic Review*, 45(4), 1079–1110.
- (2005): “Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities,” *Econometrica*, 73(1), 279–296.
- (2011): “Realized Volatility Forecasting and Market Microstructure Noise,” *Journal of Econometrics*, 160(1), 220–234.
- ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2012): “Jump-robust Volatility Estimation Using Nearest Neighbor Truncation,” *Journal of Econometrics*, 169(1), 75–93.
- (2014): “A Robust Neighborhood Truncation Approach to Estimation of Integrated Quarticity,” *Econometric Theory*, 30(01), 3–59.
- ASAI, M., M. MCALEER, AND M. C. MEDEIROS (2012): “Modelling and Forecasting Noisy Realized Volatility,” *Computational Statistics and Data Analysis*, 56(1), 217–230.
- BAJGROWICZ, P., AND O. SCAILLET (2012): “Trading Revisited: Persistence Tests, Transaction Costs, and False Discoveries,” *Journal of Financial Economics*, 106, 473–491.
- BAKER, M., B. BRADLEY, AND J. WURGLER (2011): “Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly,” *Financial Analysts Journal*, 67(1), 40–54.
- BANDI, F. M., AND J. R. RUSSELL (2006): “Separating Microstructure Noise from Volatility,” *Journal of Financial Economics*, 79(3), 655–692.
- (2008): “Microstructure Noise, Realized Variance, and Optimal Sampling,” *The Review of Economic Studies*, 75(2), 339–369.
- BANDI, F. M., J. R. RUSSELL, AND C. YANG (2013): “Realized Volatility Forecasting in the Presence of Time-Varying Noise,” *Journal of Business and Economic Statistics*, 31(3), 331–345.
- BANDI, F. M., J. R. RUSSELL, AND Y. ZHU (2008): “Using high-frequency data in dynamic portfolio choice,” *Econometric Reviews*, 27(1-3), 163–198.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): “Designing Realized Kernels to Measure the Ex Post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76(6), 1481–1536.

- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): “Realized Kernels in Practice: Trades and Quotes,” *The Econometrics Journal*, 12(3), C1–C32.
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2011): “Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading,” *Journal of Econometrics*, 162, 149–169.
- BARNDORFF-NIELSEN, O. E., S. KINNEBROCK, AND N. SHEPHARD (2010): “Measuring Downside Risk: Realised Semivariance,” in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle* (eds. T. Bollerslev, J. Russell and M. Watson), pp. 117–136. Oxford University Press.
- BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002): “Econometric Analysis of Realized Volatility and its use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society: Series B*, 64(2), 253–280.
- (2004a): “Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics,” *Econometrica*, 72(3), 885–925.
- (2004b): “Measuring the Impact of Jumps in Multivariate Price Processes using Bipower Covariation,” *Working Paper*.
- (2004c): “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, 2(1), 1–37.
- (2006): “Econometrics of Testing for Jumps in Financial Economics using Bipower Variation,” *Journal of Financial Econometrics*, 4(1), 1–30.
- BASEL COMMITTEE ON BANKING SUPERVISION (2013): “Global Systemically Important Banks: Updated Assessment Methodology and the Higher Loss Absorbency Requirement,” Bank for international settlements.
- BAUWENS, L., G. STORTI, AND F. VIOLANTE (2012): “Dynamic Conditional Correlation Models for Realized Covariance Matrices,” *Working Paper*.
- BEKAERT, G., AND M. HOEROVA (2014): “The VIX, the Variance Premium and Stock Market Volatility,” *Journal of Econometrics*, 183(2), 181–192.
- BENJAMINI, Y., AND Y. HOCHBERG (1995): “Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300.

## REFERENCES

---

- BENOIT, S., G. COLLETAZ, AND C. HURLIN (2014): “A Theoretical and Empirical Comparison of Systemic Risk Measures,” *Working paper*.
- BIANCHI, F., H. MUMTAZ, AND P. SURICO (2009): “The Great Moderation of the Term Structure of UK Interest Rates,” *Journal of Monetary Economics*, 56(6), 856–871.
- BIBINGER, M., N. HAUTSCH, P. MALEC, AND M. REISS (2014): “Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency,” *The Annals of Statistics*, 42(4), 1312–1346.
- BIBINGER, M., AND M. REISS (2014): “Spectral Estimation of Covolatility from Noisy Observations Using Local Weights,” *Scandinavian Journal of Statistics*, 41(1), 23–50.
- BLASQUES, F., S. J. KOOPMAN, AND A. LUCAS (2014a): “Maximum Likelihood Estimation for Generalized Autoregressive Score Models,” *Working Paper*.
- BLASQUES, F., S. J. KOOPMAN, AND A. LUCAS (2014b): “Optimal Formulations for Nonlinear Autoregressive Processes,” *Working Paper*.
- BOGUTH, O., M. CARLSON, A. FISHER, AND M. SIMUTIN (2011): “Conditional Risk and Performance Evaluation: Volatility Timing, Overconditioning, and New Estimates of Momentum Alphas,” *Journal of Financial Economics*, 102(2), 363–389.
- BOLLERSLEV, T., A. J. PATTON, AND R. QUAEDEVLIIEG (2016): “Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting,” *Journal of Econometrics*, 192(1), 1–18.
- BOLLERSLEV, T., V. TODOROV, AND S. Z. LI (2013): “Jump tails, extreme dependencies, and the distribution of stock returns,” *Journal of Econometrics*, 172(2), 307–324.
- BOLLERSLEV, T., AND B. Y. ZHANG (2003): “Measuring and Modeling Systematic Risk in Factor Pricing Models using High-Frequency Data,” *Journal of Empirical Finance*, 10(5), 533–558.
- BOUDT, K., S. LAURENT, A. LUNDE, R. QUAEDEVLIIEG, AND O. SAURI (2015): “Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity,” *Working Paper*.
- BRODIE, J., I. DAUBECHIES, C. DE MOL, D. GIANNONE, AND I. LORIS (2009): “Sparse and Stable Markowitz Portfolios,” *Proceedings of the National Academy of Sciences*, 106(30), 12267–12272.
- BROWN, D. B., AND J. E. SMITH (2011): “Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds,” *Management Science*, 57(10), 1752–1770.
- BROWNLEES, C., AND R. ENGLE (2012): “Volatility, Correlation and Tails for Systemic Risk Measurement,” *Working paper*.

- BROWNLEES, C. T., AND G. M. GALLO (2010): “Comparison of Volatility Measures: a Risk Management Perspective,” *Journal of Financial Econometrics*, 8(1), 29–56.
- CARHART, M. M. (1997): “On Persistence in Mutual Fund Performance,” *The Journal of Finance*, 52(1), 57–82.
- CASTRO, C., AND S. FERRARI (2014): “Measuring and Testing for the Systemically Important Financial Institutions,” *Journal of Empirical Finance*, 25, 1–14.
- CHAKER, S., AND N. MEDDAHI (2013): “Volatility Forecasting When the Noise Variance is Time-Varying,” *Working Paper*.
- CHAN, L. K., J. KARCESKI, AND J. LAKONISHOK (1999): “On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model,” *Review of Financial Studies*, 12(5), 937–974.
- CHAN, N., S. DENG, L. PENG, AND Z. XIA (2007): “Interval Estimation of Value-at-Risk Based on GARCH Models with Heavy-tailed Innovations,” *Journal of Econometrics*, 137(2), 556–576.
- CHEN, Q., AND R. H. GERLACH (2013): “The Two-Sided Weibull Distribution and Forecasting Financial Tail Risk,” *International Journal of Forecasting*, 29(4), 527–540.
- CHIRIAC, R., AND V. VOEV (2011): “Modelling and Forecasting Multivariate Realized Volatility,” *Journal of Applied Econometrics*, 26(6), 922–947.
- CHRISTENSEN, J. H., F. X. DIEBOLD, AND G. D. RUDEBUSCH (2011): “The Affine Arbitrage-Free Class of Nelson–Siegel Term Structure Models,” *Journal of Econometrics*, 164(1), 4–20.
- CHRISTENSEN, K., S. KINNEBROCK, AND M. PODOLSKIJ (2010): “Pre-Averaging Estimators of the Ex-post Covariance Matrix in Noisy Diffusion Models with Non-synchronous Data,” *Journal of Econometrics*, 159, 116–133.
- CHRISTENSEN, K., M. PODOLSKIJ, AND M. VETTER (2013): “On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes,” *Journal of Multivariate Analysis*, 120, 59–84.
- CHRISTOFFERSEN, P., AND S. GONÇALVES (2005): “Estimation Risk in Financial Risk Management,” *Journal of Risk*, 7(3), 1–28.
- CONRAD, C., AND K. LOCH (2015): “The Variance Risk Premium and Fundamental Uncertainty,” *Working Paper*.
- CORRADI, V., W. DISTASO, AND M. FERNANDES (2013): “Conditional Alphas and Realized Betas,” *Working Paper*.

## REFERENCES

---

- CORSI, F. (2009): “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 (2008): “A General Framework for Observation Driven Time-Varying Parameter Models,” *Working Paper*.
- (2013): “Generalized Autoregressive Score Models with Applications,” *Journal of Applied Econometrics*, 28(5), 777–795.
- DAI, Q., AND K. J. SINGLETON (2000): “Specification Analysis of Affine Term Structure Models,” *The Journal of Finance*, 55(5), 1943–1978.
- DANIELSSON, J., K. R. JAMES, M. VALENZUELA, AND I. ZER (2011): “Model Risk of Systemic Risk Models,” *Working paper*.
- DE LA VEGA, W. F. (1974): “On Almost Sure Convergence of Quadratic Brownian Variation,” *The Annals of Probability*, 2(3), 551–552.
- DE LIRA SALVATIERRA, I., AND A. J. PATTON (2015): “Dynamic Copula Models and High Frequency Data,” *Journal of Empirical Finance*, 30, 120–135.
- DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009): “A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms,” *Management Science*, 55(5), 798–812.
- DEMIGUEL, V., F. J. NOGALES, AND R. UPPAL (2014): “Stock Return Serial Dependence and Out-of-Sample Portfolio Performance,” *Review of Financial Studies*, 27(4), 1031–1073.
- DIEBOLD, F., AND R. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics*, 13, 253–263.
- DIEBOLD, F. X., L. JI, AND C. LI (2004): “A Three-Factor Yield Curve Model: Non-Affine Structure, Systematic Risk Sources, and Generalized Duration,” *Working Paper*.
- DIEBOLD, F. X., AND C. LI (2006): “Forecasting the Term Structure of Government Bond Yields,” *Journal of Econometrics*, 130(2), 337–364.
- DIEBOLD, F. X., AND G. D. RUDEBUSCH (2013): *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press.
- DIEBOLD, F. X., G. D. RUDEBUSCH, AND S. BORAGAN ARUOBA (2006): “The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach,” *Journal of Econometrics*, 131(1), 309–338.

- DOBREV, D. P., AND P. J. SZERSZEN (2010): “The Information Content of High-Frequency Data for Estimating Equity Return Models and Forecasting Risk,” *Working Paper*.
- DOORNIK, J. (2012): “Object-Oriented Matrix Programming Using Ox,” *3rd ed. London: Timberlake Consultants Press and Oxford*.
- DUFFEE, G. R. (2002): “Term Premia and Interest Rate Forecasts in Affine Models,” *The Journal of Finance*, 57(1), 405–443.
- DUFFIE, D., AND R. KAN (1996): “A Yield-Factor Model of Interest Rates,” *Mathematical Finance*, 6(4), 379–406.
- ENGLE, R. F. (2015): “Dynamic Conditional Beta,” *Working Paper*.
- ENGLE, R. F., AND K. F. KRONER (1995): “Multivariate simultaneous generalized ARCH,” *Econometric theory*, 11(01), 122–150.
- ENGLE, R. F., AND S. MANGANELLI (2004): “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles,” *Journal of Business & Economic Statistics*, 22(4), 367–381.
- ENGLE, R. F., N. SHEPHARD, AND K. SHEPPARD (2008): “Fitting Vast Dimensional Time-varying Covariance Models,” *Working Paper*.
- EPPE, T. W. (1979): “Comovements in Stock Prices in the Very Short Run,” *Journal of the American Statistical Association*, 74(366a), 291–298.
- ESCANCIANO, J., AND J. OLMO (2010): “Backtesting Parametric Value-at-Risk with Estimation risk,” *Journal of Business and Economic Statistics*, 28(1), 36–51.
- (2011): “Robust Backtesting Tests for Value-at-risk,” *Journal of Financial Econometrics*, 9(1), 132–161.
- FAMA, E. F., AND K. R. FRENCH (1993): “Common Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics*, 33(1), 3–56.
- FAN, J., Y. FAN, AND J. LV (2008): “High Dimensional Covariance Matrix Estimation using a Factor Model,” *Journal of Econometrics*, 147(1), 186–197.
- FAN, J., Y. LI, AND K. YU (2012): “Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection,” *Journal of the American Statistical Association*, 107(497), 412–428.
- FLEMING, J., C. KIRBY, AND B. OSTDIEK (2001): “The Economic Value of Volatility Timing,” *The Journal of Finance*, 56(1), 329–352.

## REFERENCES

---

- (2003): “The Economic Value of Volatility Timing using “Realized” Volatility,” *Journal of Financial Economics*, 67(3), 473–509.
- FRANCO, C., AND J. ZAKOÏAN (2015): “Risk-parameter Estimation in Volatility Models,” *Journal of Econometrics*, 184, 158–173.
- FROST, P. A., AND J. E. SAVARINO (1986): “An Empirical Bayes Approach to Efficient Portfolio Selection,” *Journal of Financial and Quantitative Analysis*, 21(3), 293–305.
- GHYSELS, E., AND A. SINKO (2011): “Volatility Forecasting and Microstructure Noise,” *Journal of Econometrics*, 160(1), 257–271.
- GIOT, P., AND S. LAURENT (2004): “Modelling Daily Value-at-Risk using Realized Volatility and ARCH Type Models,” *Journal of Empirical Finance*, 11(3), 379–398.
- GONÇALVES, S., AND N. MEDDAHI (2009): “Bootstrapping realized volatility,” *Econometrica*, 77(1), 283–306.
- GOURIÉROUX, C., AND J.-M. ZAKOÏAN (2013): “Estimation Adjusted VaR,” *Econometric Theory*, 29(4), 735–770.
- GUNTAY, L., AND P. KUPIEC (2015): “Testing for Systemic Risk using Stock Returns,” *Working paper*.
- HAN, Y. (2006): “Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model,” *Review of Financial Studies*, 19(1), 237–271.
- HANSEN, P. R., AND A. LUNDE (2006a): “Consistent Ranking of Volatility Models,” *Journal of Econometrics*, 131(1), 97–121.
- (2006b): “Realized variance and market microstructure noise,” *Journal of Business and Economic Statistics*, 24(2), 127–161.
- (2014a): “Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error,” *Econometric Theory*, 30(1), 60–93.
- (2014b): “MulCom 3.00, an Ox Software Package for Multiple Comparisons,” *Working Paper*.
- HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): “The Model Confidence Set,” *Econometrica*, 79(2), 453–497.
- HANSEN, P. R., A. LUNDE, AND V. VOEV (2014): “Realized Beta GARCH: a Multivariate GARCH Model with Realized Measures of Volatility,” *Journal of Applied Econometrics*, 29(5), 774–799.

- HARRIS, F., T. MCINISCH, G. SHOESMITH, AND R. WOOD (1995): “Cointegration, Error correction, and Price Discovery on Informationally Linked Security Markets,” *Journal of Financial and Quantitative Analysis*, 30, 563–581.
- HARTZ, C., S. MITTNIK, AND M. PAOLELLA (2006): “Accurate Value-at-Risk Forecasting Based on the Normal-GARCH Model,” *Computational Statistics & Data Analysis*, 51(4), 2295–2312.
- HAUTSCH, N., L. KYJ, AND R. OOMEN (2012): “A blocking and regularization approach to high-dimensional realized covariance estimation,” *Journal of Applied Econometrics*, 27, 627–645.
- HAUTSCH, N., L. M. KYJ, AND P. MALEC (2015): “Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?,” *Journal of Applied Econometrics*, 30(2), 263–290.
- HAUTSCH, N., AND Y. OU (2012): “Analyzing Interest Rate Risk: Stochastic Volatility in the Term Structure of Government Bond Yields,” *Journal of Banking & Finance*, 36(11), 2988–3007.
- HAYASHI, T., AND N. YOSHIDA (2005): “On Covariance Estimation of Nonsynchronously Observed Diffusion Processes,” *Bernoulli*, 11, 359–379.
- HAYASHI, T., AND N. YOSHIDA (2008): “Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes,” *Annals of the Institute of Statistical Mathematics*, 60(2), 367–406.
- HIDALGO, J., AND P. ZAFFARONI (2007): “A Goodness-of-fit Test for ARCH( $\infty$ ) Models,” *Journal of Econometrics*, 141(2), 835–875.
- HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): “Estimating Vector Autoregressions with Panel Data,” *Econometrica*, 56(6), 1371–1395.
- HUANG, X., AND G. TAUCHEN (2005): “The Relative Contribution of Jumps to Total Price Variance,” *Journal of Financial Econometrics*, 3(4), 456–499.
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): “Microstructure Noise in the Continuous Case: the Pre-Averaging Approach,” *Stochastic Processes and their Applications*, 119(7), 2249–2276.
- JAGANNATHAN, R., AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *The Journal of Finance*, 58(4), 1651–1684.
- KOMUNJER, I., AND S. NG (2014): “Measurement Errors in Dynamic Models,” *Econometric Theory*, 30(1), 150–175.

## REFERENCES

---

- KOOPMAN, S. J., M. I. MALLEE, AND M. VAN DER WEL (2010): “Analyzing the Term Structure of Interest Rates using the Dynamic Nelson–Siegel Model with Time-Varying Parameters,” *Journal of Business & Economic Statistics*, 28(3), 329–343.
- LAURENT, S. (2013): “G@RCH 7.0: Estimating and Forecasting ARCH Models,” *London: Timberlake Consultants Ltd.*
- LAURENT, S., J. V. ROMBOUTS, AND F. VIOLANTE (2013): “On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models,” *Journal of Econometrics*, 173(1), 1–10.
- LEDOIT, O., AND M. WOLF (2003): “Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection,” *Journal of Empirical Finance*, 10(5), 603–621.
- (2004a): “Honey, I Shrunk the Sample Covariance Matrix,” *Journal of Portfolio Management*, 30(4), 110–119.
- (2004b): “A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices,” *Journal of Multivariate Analysis*, 88(2), 365–411.
- LI, J. (2015): “Sparse and Stable Portfolio Selection with Parameter Uncertainty,” *Journal of Business & Economic Statistics*, 33(3), 381–392.
- LI, J., AND D. XIU (2013): “Spot Variance Regressions,” *Working Paper*.
- LIU, L., A. J. PATTON, AND K. SHEPPARD (2015): “Does Anything Beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes,” *Journal of Econometrics*, forthcoming.
- LIU, Q. (2009): “On Portfolio Optimization: How and When do we Benefit from High-Frequency Data?,” *Journal of Applied Econometrics*, 24(4), 560–582.
- LUNDE, A., N. SHEPHARD, AND K. SHEPPARD (2015): “Econometric Analysis of Vast Covariance Matrices using Composite Realized Kernels and their Application to Portfolio Choice,” *Journal of Business & Economic Statistics*, *Forthcoming*.
- MAGNUS, J. R., AND H. NEUDECKER (1980): “The Elimination Matrix: Some Lemmas and Applications,” *SIAM Journal on Algebraic Discrete Methods*, 1(4), 422–449.
- MANCINI, C. (2009): “Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps,” *Scandinavian Journal of Statistics*, 36(2), 270–296.
- MYKLAND, P. A. (2010): “A Gaussian Calculus for Inference from High Frequency Data,” *Annals of Finance*, 8(2), 235–258.

- NELSON, C. R., AND A. F. SIEGEL (1987): “Parsimonious Modeling of Yield Curves,” *Journal of business*, pp. 473–489.
- NEWBY, W. K., AND K. D. WEST (1987): “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55(3), 703–08.
- NOURDIN, I., G. PECCATI, AND A. RÉVEILLAC (2010): “Multivariate Normal Approximation using Stein’s Method and Malliavin Calculus,” in *Annales de l’institut Henri Poincaré (B)*, vol. 46, pp. 45–58.
- NOURELDIN, D., N. SHEPHARD, AND K. SHEPPARD (2012): “Multivariate High-Frequency-Based Volatility (HEAVY) Models,” *Journal of Applied Econometrics*, 27(6), 907–933.
- NUALART, D. (2006): *The Malliavin Calculus and Related Topics*, Probability and its Applications. Springer Berlin Heidelberg, 2nd edn.
- NUALART, D., AND G. PECCATI (2005): “Central Limit Theorems for Sequences of Multiple Stochastic Integrals,” *Ann. Probab.*, 33(1), 177–193.
- OH, D. H., AND A. J. PATTON (2015): “Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads,” *Working Paper*.
- PAKEL, C., N. SHEPHARD, K. SHEPPARD, AND R. F. ENGLE (2014): “Fitting Vast Dimensional Time-Varying Covariance Models,” *Working Paper*.
- PALANDRI, A. (2009): “Sequential Conditional Correlations: Inference and Evaluation,” *Journal of Econometrics*, 153, 122–132.
- PASCUAL, L., J. ROMO, AND E. RUIZ (2006): “Bootstrap Prediction for Returns and Volatilities in GARCH Models,” *Computational Statistics & Data Analysis*, 50(9), 2293–2312.
- PATTON, A. J. (2011a): “Data-Based Ranking of Realised Volatility Estimators,” *Journal of Econometrics*, 161(2), 284–303.
- (2011b): “Volatility Forecast Comparison using Imperfect Volatility Proxies,” *Journal of Econometrics*, 160(1), 246–256.
- PATTON, A. J., AND K. SHEPPARD (2015): “Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility,” *Review of Economics and Statistics*, forthcoming.
- PATTON, A. J., AND M. VERARDO (2012): “Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability,” *Review of Financial Studies*, 25(9), 2789–2839.

## REFERENCES

---

- PECCATI, G., AND C. A. TUDOR (2005): “Gaussian Limits for Vector-Valued Multiple Stochastic Integrals,” in *Séminaire de Probabilités XXXVIII*, pp. 247–262. Springer.
- PELUSO, S., F. CORSI, AND A. MIRA (2014): “A Bayesian High-Frequency Estimator of the Multivariate Covariance of Noisy and Asynchronous returns,” *Journal of Financial Econometrics*, 13(3), 665–697.
- POLITIS, D. N., AND J. P. ROMANO (1994): “The Stationary Bootstrap,” *Journal of the American Statistical Association*, 89(428), 1303–1313.
- POOTER, M. D., M. MARTENS, AND D. V. DIJK (2008): “Predicting the Daily Covariance Matrix for S&P 100 Stocks using Intraday Data — But which Frequency to use?,” *Econometric Reviews*, 27(1-3), 199–229.
- REEVES, J. J. (2005): “Bootstrap Prediction Intervals for ARCH Models,” *International Journal of Forecasting*, 21(2), 237–248.
- ROBIO, P. O. (1999): “Forecast Intervals in ARCH Models: Bootstrap versus Parametric Methods,” *Applied Economics Letters*, 6(5), 323–327.
- ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2008a): “Control of the False Discovery Rate under Dependence using the Bootstrap and Subsampling,” *Test*, 17(3), 417–442.
- ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2008b): “Formalized Data Snooping Based on Generalized Error Rates,” *Econometric Theory*, 24, 404–447.
- ROMANO, J. P., AND M. WOLF (2005): “Stepwise Multiple Testing as Formalized Data Snooping,” *Econometrica*, 73, 1237–1282.
- SANTOS, A. A., F. J. NOGALES, AND E. RUIZ (2013): “Comparing Univariate and Multivariate Models to Forecast Portfolio Value-at-Risk,” *Journal of Financial Econometrics*, 11(2), 400–441.
- SCAILLET, O. (2004): “Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall,” *Mathematical Finance*, 14(1), 115–129.
- SCAILLET, O. (2005): “Nonparametric Estimation of Conditional Expected Shortfall,” *Insurance and Risk Management Journal*, 74, 382–406.
- SCHMELZER, T., AND R. HAUSER (2013): “Seven Sins in Portfolio Optimization,” *Working Paper*.
- SHEPHARD, N., AND D. XIU (2014): “Econometric Analysis of Multivariate Realised QML: Efficient Positive Semi-definite Estimators of the Covariation of Equity Prices.,” *Working Paper*.

- SHIMIZU, K. (2013): “The Bootstrap does not Always Work for Heteroscedastic Models,” *Statistics & Risk Modeling*, 30(3), 189–204.
- SIZOVA, N. (2011): “Integrated Variance Forecasting: Model-Based versus Reduced-Form,” *Journal of Econometrics*, 162(2), 294–311.
- STAUDENMAYER, J., AND J. P. BUONACCORSI (2005): “Measurement Error in Linear Autoregressive Models,” *Journal of the American Statistical Association*, 100(471), 841–852.
- SWANSON, N. R., AND H. WHITE (1997): “Forecasting Economic Time Series using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models,” *International journal of Forecasting*, 13(4), 439–461.
- TSAY, R. (2010): *Analysis of Financial Time Series*, Wiley Series in Probability and Statistics Series. Wiley.
- VARNESKOV, R., AND V. VOEV (2013): “The Role of Realized Ex-Post Covariance Measures and Dynamic Model Choice on the Quality of Covariance Forecasts,” *Journal of Empirical Finance*, 20, 83–95.
- WEST, K. D. (1996): “Asymptotic Inference about Predictive Ability,” *Econometrica*, pp. 1067–1084.
- (2006): “Forecast Evaluation,” in *Handbook of economic forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann. North Holland Press, Amsterdam.
- WHITE, H. (1980): “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica*, 48(4), 817–838.
- (2000): “A Reality Check for Data Snooping,” *Econometrica*, 68(5), 1097–1126.
- ZHANG, L. (2011): “Estimating Covariation: Epps Effect, Microstructure Noise,” *Journal of Econometrics*, 160, 33–47.
- ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100(472), 1394–1411.

# Research Impact

‘Ask five different economists a question, and you will get six different answers’. This joke is often used to put economists in a bad light, while if anything, the economist’s response is closest to the academic ideal. We deal with a limited amount of data, which is often poorly recorded and driven by the beliefs of an impossible number of decision makers. Even worse, many of the official figures are subject to political pressure, and are often revised. In order to obtain a perfect setting in which we can get accurate and proper data from which it is easy to infer relations, we have to resort to the controlled setting of experiments, where we can only answer small questions that are impossible to generalize.

It is therefore no wonder that economists cannot agree on what to make of the economy. No two datasets are truly comparable and each will lead to different conclusions. The conclusions drawn are in the hand of the researcher or econometrician. The results of any study are a result of both the model specification and the data used to test the model on. Careful thought therefore has to be put in both of these aspects, and each specification and dataset has its pros and cons. These are very tough decisions to make.

And it gets worse from there, as the process of quantifying meaningful economic relations is a matter of econometrics and statistics. Since we are dealing with probabilities, there is always the chance that we find an economic relationship where there is none, and similarly, that we are unable to disentangle anything meaningful while in reality one variable causes the other. A large part of the econometrics literature concerns itself with accurately measuring the uncertainty in our estimates. Accurate assessment of the precision is vital, as it controls the degree of evidence a dataset provides for a certain claim.

This thesis provides new methods to (i) better quantify this uncertainty in order to be able to make more accurate claims, (ii) reduce the uncertainty in order to be able to make claims we were not able to make with less precise estimates, and (iii) use this precision in order to assess the reliability of data over time, such that we can better react to new information by disregarding poor estimates. Incorporating these techniques, and many other techniques developed by others, can hopefully lead to the fifth economist at the very least agreeing with himself, and only giving one answer to the question.

The applications in this thesis mainly revolve around risk management, which has clear applicability outside of academia. The methods presented may be of interest to banks, asset managers, pension funds, but also for regulatory institutions. All these institutions face the exact same problem of interpreting data to decide on their operations and policies. These decisions may have a large impact on their clients' investments, retirement savings or the economy as a whole in the case of policy makers.

### **(i) Quantifying uncertainty**

In chapter 2 we propose a method to classify different financial institutions in terms of their 'systemic risk', or the degree of how 'too-big-to-fail' they are. For regulators this is a daily concern, and many ideas have been proposed to somehow curb the banks' behavior to prevent another global financial crisis. Their accurate assessment of these firms' risks is of great importance, first, to hopefully be effective in their policies, and second, as they will impose restrictions on these firms. Academics have also tried to shed light on this issue and have developed an enormous literature of all kinds of measures of this systemic risk, the ultimate goal being to rank these firms. We quantify the uncertainty in these measures (and potentially many others), in order to make more accurate claims. We find that the uncertainty in these measures is of considerable magnitude and one has to be very careful in making bold claims about their ability to distinguish firms in terms of their riskiness.

### **(ii) Reducing uncertainty**

Chapter 3 proposes a new method to more accurately estimate the common risks of stocks. An accurate estimate is of great importance for at least two reasons. First, it can be used to improve portfolio allocations in terms of their risk-return balance, and second, it provides more accurate estimates of the risks inherent for a given holding. A precise estimator, such as the one proposed in this thesis, can therefore be of benefit to any company who deals with stocks.

With a similar goal, but in a very different way, chapter 6 proposes a method to improve the hedging of long-term liabilities. The problem is clearly of great interest to financial institutions such as pension funds. Here, the method to reduce uncertainty is to more accurately capture the time-variation in the dynamics of interest rates. The chapter provides a simple method that can be used by any practitioner to think about ways to improve their hedging decisions, and reduce surprises in future costs.

### **(iii) Exploiting uncertainty**

Chapters 4 and 5 propose to explicitly use the amount of uncertainty that we know is present to make better decisions. By using this information different data points can get different weights. These papers use this in a dynamic setting, where intuitively, new

information is only incorporated when the observation is sufficiently precise. The research shows that by properly taking into account this uncertainty, more accurate decisions can be made. Chapter 4 demonstrates improved volatility forecasts and the illustrations in chapter 5 include portfolio formation and beta hedging. However, there are many more potential applications. The methods can be used in any time series where there we have a general idea on the precision of the estimates, such as time series of analyst forecasts.

All of these three elements allow the economist, academic or practitioner, to either better extract economic implications from data, or better forecast important variables. Apart from the immediate appeal to risk management presented herein, most of the presented methods offer general applicability in a wide variety of set-ups, which for now are left unexplored.

# Biography

Rogier Quaedvlieg was born on July 10th, 1990 in Nijmegen, the Netherlands. He earned a B.Sc. in Econometrics and a B.Sc. in International Business from Maastricht University in 2011. In 2012 he received a M.Sc. in Econometrics and a M.Sc. in Financial Economics, both with distinction (*cum laude*) from Maastricht University.

After graduating, Rogier joined the Department of Finance at Maastricht University to pursue a doctorate degree under the supervision of Prof. dr. Peter Schotman and Prof. dr. Sébastien Laurent. During his PhD he visited CREATES in Denmark and Duke University in the United States.

The result of this effort are collected in this thesis. Rogier has presented this research at several international conferences including the 2014 and 2015 Society for Financial Econometrics (SoFiE) meetings, the 2015 Econometric Society World Congress and the 2015 European Finance Association Annual Meeting. Chapters of this book have been published in the Journal of Econometrics and the Journal of Business & Economic Statistics.