Essays on Financial Time Series with a Focus on High-Frequency Data

Der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
— Doctor rerum politicarum —

genehmigte Dissertation

von

M.Sc. Janis Becker
geboren am 04.10.1992 in Münster

2020

Referent: Prof. Dr. Philipp Sibbertsen, Leibniz Universität Hannover

Korreferent: Prof. Dr. Marcel Prokopczuk, Leibniz Universität Hannover

Tag der Promotion: 09.03.2020

Acknowledgements

Many people provided support for this dissertation, and I owe all of them a debt of gratitude.

First, I want to thank my supervisor and co-author Prof. Dr. Philipp Sibbertsen for giving me the chance to discover my passion for statistics and data analysis. I am grateful for his support and guidance over the last years.

Further, I want to thank my co-authors Prof. Dr. Marcel Prokopczuk and Dr. Fabian Hollstein as well as Prof. Dr. Marina Schröder for taking the time to be on my examination board.

I am also grateful to my other co-authors, Dr. Christian Leschinski and Dr. Kai Wenger. Particularly, I want to thank Christian for all the support during the first year of my PhD, for sharing his enthusiasm for statistics, and for numerous inspiring discussions.

Thanks must also go to Dr. Fabian Hollstein, Simon Wingert, and participants of the Statistical Week (2017–2019) and the CFE (2017, 2019) as they provided critical remarks on the essays in this dissertation.

Many thanks are further due to all of my colleagues and former colleagues for creating such a productive and friendly working atmosphere. I especially want to thank Dr. Kai Wenger with whom I became close friends. I appreciate our conversations, our (sport) competitions, and the laughter we shared.

Finally, I want to thank my family and friends for their constant support and encouragement.

Abstract

This thesis contains six essays on financial time series. Special attention is paid to the opportunities that high-frequency data offers for modeling and forecasting the return and the risk, measured by the volatility or beta, of an asset.

After an introduction in the first chapter, Chapter 2 shows that, using a variety of high-frequency based explanatory variables, the sign of daily stock returns is predictable in an out-of-sample environment. This predictability is of a magnitude that is statistically significant and consistent over time. Even after accounting for transaction costs, a simple trading strategy based on directional forecasts yields a Sharpe ratio that is nearly double that of the market and an annualized alpha of more than eight percent in a multi-factor model. Consequently, standard risk based models are not able to explain the returns generated by this strategy.

Chapter 3 provides a simple approach to estimate the volatility of economy wide risk factors such as size or value. Models based on these factors are ubiquitous in asset pricing. Therefore, portfolio allocation and risk management require estimates of the volatility of these factors. While realized measures based on high-frequency observations, such as realized variance, have become the standard tools for the estimation of the volatility of liquid individual assets, these measures are difficult to obtain for economy wide risk factors that include smaller illiquid stocks that are not traded at a high frequency. The approach suggested in Chapter 3 improves on this issue as it yields an estimate that is close in precision to realized variance. The efficacy of this approach is demonstrated using Monte Carlo simulations and forecasts of the variance of the market factor.

Chapter 4 shows that realized variance underestimates the variance of daily stock index returns by an average of 14 percent. This is documented for a wide range of international stock indices, using the fact that the average of realized variance and that of squared returns should be the same over longer time horizons. It is shown that the magnitude of this bias cannot be explained by market microstructure noise. Instead, it can be attributed to correlation between the continuous components of intra-day returns.

Chapter 5 reveals that beta series show consistent long-memory properties. This result is based on the analysis of the realized beta series of over 800 stocks. Researchers and practitioners employ a variety of time-series processes to forecast beta series, using either short-memory models or implicitly imposing infinite memory. The results in Chapter 5 suggest that both approaches are inadequate. A pure long-memory model reliably provides superior beta forecasts compared to all alternatives.

Building on the result that beta series can be best described by long-memory processes, Chapter 6 suggests a new multivariate approach to estimate the long-memory parameter robust to low-frequency contaminations. This estimator requires a priori knowledge of the cointegration rank. Since low-frequency contaminations bias inference on the cointegration rank, a robust estimator of the cointegration rank is also provided. An extensive Monte Carlo exercise shows the applicability of the estimators in finite samples. Furthermore, the procedures are applied to the realized beta series of two American energy companies discovering that the series are fractionally cointegrated. As the series exhibit low-frequency contaminations, standard procedures are unable to detect this relation.

Finally, Chapter 7 presents the R package *memochange*. The package includes several change-in-mean tests that are applicable under long memory as standard change-in-mean tests are invalid in this case. Moreover, the package contains various tests for a break in persistence. These can be used to detect a change in the memory parameter.

Keywords: Asset Pricing · Beta · Directional Predictability · Factor Models · Forecasting · Fractional Cointegration · High-Frequency Data · Long Memory · Persistence · Return Predictability · Realized Variance · Squared Returns · Volatility

Contents

List of Figures **VI**

List of Tables **VII**

1 Introduction **1**

2 Directional Predictability of Daily Stock Returns **5**
2.1 Introduction 5
2.2 Data 7
2.2.1 Explanatory Variables 7
2.3 Methodology 9
2.3.1 Model Framework 9
2.3.2 Model Selection 10
2.3.3 Forecasting Procedure 12
2.4 Empirical Analysis 13
2.4.1 Forecast Results 13
2.4.2 Economic Significance 16
2.4.3 Performance of the Trading Strategy over Time 22
2.4.4 Economic Implications 24
2.5 Additional Analyses and Robustness 26
2.5.1 Further Results on the Behavior of the Trading Strategy 26
2.5.2 Impact of the Subprime Mortgage Crisis 28
2.5.3 Robustness 28
2.6 Conclusion 30
Appendix 31

3 Estimating the Volatility of Asset Pricing Factors **36**
3.1 Introduction 36
3.2 Estimating Factor Volatility 38
3.3 Monte Carlo Simulation 40
3.4 Empirical Analysis 42
3.4.1 In-Sample Volatility Estimates and Model Diagnostics 43
3.4.2 Out-of-Sample Forecasts of Market Factor Volatility 47
3.5 Conclusion 49

4 The Bias of Realized Variance **51**
4.1 Introduction and Main Finding 51
4.2 Explaining the Bias of Realized Variance 57
4.2.1 Market Microstructure Noise 59
4.2.2 Jumps and Continuous Returns 59
4.3 Conclusion 65
Appendix 66

5 The Memory of Beta 76
5.1 Introduction . . . 76
5.2 Data and Methodology . . . 79
5.2.1 Data . . . 79
5.2.2 Beta Estimation . . . 80
5.2.3 Long-Memory Estimation . . . 80
5.3 Long Memory in Beta . . . 82
5.3.1 Estimation Results . . . 82
5.3.2 Beta Decomposition . . . 84
5.4 Forecasting . . . 86
5.4.1 Forecasting Methodology . . . 86
5.4.2 Forecast Results . . . 88
5.4.3 Longer Forecast Horizons . . . 90
5.5 Economic Implications . . . 92
5.5.1 The Memory in Beta and Stock Characteristics . . . 92
5.5.2 The Determinants of Forecast Errors . . . 94
5.6 Additional Analyses and Robustness . . . 97
5.6.1 Hedging Errors . . . 97
5.6.2 Entire CRSP Dataset . . . 97
5.6.3 Alternative Models . . . 100
5.6.4 Alternative Long-Memory Estimator . . . 102
5.6.5 Alternative Sampling Frequencies . . . 102
5.6.6 Alternative Estimation Windows and Bandwidths . . . 103
5.7 Conclusion . . . 104
Appendix . . . 105

6 Robust Multivariate Local Whittle Estimation and Spurious Fractional Cointegration 116
6.1 Introduction . . . 116
6.2 The Periodogram of Spurious Long-Memory Processes . . . 117
6.3 Robust Fractional Cointegration Rank Estimator . . . 121
6.4 Robust Multivariate Local Whittle Estimator . . . 126
6.5 Monte Carlo Simulation . . . 129
6.5.1 Fractional Cointegration . . . 130
6.5.2 Order of Integration . . . 132
6.6 Empirical Example . . . 134
6.7 Conclusion . . . 137
Appendix . . . 138
Supplementary Appendix . . . 157

7 memochange: an R Package for Estimation and Tests in Persistent Time Series 171

Bibliography 172

List of Figures

2.1 Density and model confidence plots 17
2.2 Break-even transaction costs 21
2.3 Cumulative difference plot and cumulative trading return plot 22
2.4 Correlation plot for the estimated coefficients from the logistic regression models 25
2.5 Range of the predicted probabilities for a positive return 27
2.6 Correlation plot of the explanatory variables 32
2.7 Cumulative difference plots for a number of robustness checks 35

3.1 Graphical comparison of Ridge-RV and squared returns 44
3.2 Graphical comparison of Ridge-RV and true RV 47

4.1 Average variance scatter plot 54
4.2 Average variance time-series plots S&P 500 55
4.3 Correlation matrix S&P 500 62
4.4 Corrected average variance scatter plot 64
4.5 Average variance scatter plot — realized library 66
4.6 Average variance scatter plot — 5-minute data 67
4.7 Average variance scatter plot — 30-minute data 68
4.8 Average variance scatter plot — realized kernel variance 69
4.9 Corrected average variance time-series plots S&P 500 73
4.10 Corrected average variance time-series plots SSEC 73
4.11 Corrected average variance time-series plots BSESN 74
4.12 Corrected average variance time-series plots DAX 74

5.1 Density plots for memory parameter estimates 83

6.1 Realized beta plots 135

List of Tables

2.1 Forecast results 15
2.2 Annualized performance measures for the trading strategy 20
2.3 Summary statistics of the coefficients in the logistic regression models . . . 24
2.4 Summary statistics of the explanatory variables 31
2.5 Regression results 33
2.6 Annualized performance measures for the trading strategy excluding the period of the subprime mortgage crisis 33
2.7 Aggregated forecasting results for a number of robustness checks 34
2.8 Annualized performance measures for the trading strategy for a number of robustness checks 34

3.1 Simulation results 42
3.2 Auxiliary regression results 45
3.3 Fractional cointegration test results 46
3.4 Forecast results 48
3.5 Diebold–Mariano test statistics 49

4.1 Statistics on average squared return and average RV 56
4.2 HAC t-test statistics 61
4.3 Significant correlations S&P 500 63
4.4 MAC t-test statistics 70
4.5 MOM t-test statistics 71
4.6 Statistics on average squared return and average corrected RV 72
4.7 Statistics on average squared return and average RV — excluding first and last hour of the trading day 75

5.1 Average memory parameter estimates – realized beta 82
5.2 Average memory parameter estimates – realized correlation, realized volatility, inverse of realized market volatility 85
5.3 Forecast results 88
5.4 MSE decomposition 89
5.5 Forecast results – longer horizons 91
5.6 Portfolio sorts by estimated d 93
5.7 Average memory estimate by industry 94
5.8 Forecast error regressions – RW 95
5.9 Forecast error regressions – ARMA 96
5.10 Hedging errors 98
5.11 Average memory parameter estimates – entire CRSP sample 98
5.12 Forecast results – entire CRSP sample 99
5.13 Simulation results 107
5.14 Average memory parameter estimates – bandwidth $m = T^{0.65}$ and $m = T^{0.75}$ 108
5.15 Portfolio sorts by beta 109
5.16 Forecast results – alternative models 110

5.17 Average memory parameter estimates – log-periodogram estimator 110
5.18 Forecast results – log-periodogram estimator 111
5.19 Average memory parameter estimates – 15-minute and 75-minute data . . 112
5.20 Forecast results – 15-minute and 75-minute data 113
5.21 Forecast results – rolling window sizes of 75 and 125 observations 114
5.22 Forecast results – bandwidth $m = T^{0.65}$ and $m = T^{0.75}$ 115

6.1 Simulation results fractional cointegration rank 131
6.2 Simulation results order of integration — no fractional cointegration 132
6.3 Simulation results order of integration — fractional cointegration 133
6.4 Estimation results 136
6.5 Simulation results fractional cointegration rank — cross-sectionally correlated errors 157
6.6 Simulation results order of integration — no fractional cointegration but cross-sectionally correlated errors 158
6.7 Simulation results order of integration — fractional cointegration and cross-sectionally correlated errors 159
6.8 Simulation results fractional cointegration rank — stationary random level-shift process shifts 161
6.9 Simulation results order of integration — stationary random level-shift process and no fractional cointegration 162
6.10 Simulation results order of integration — stationary random level-shift process and fractional cointegration 163
6.11 Simulation results fractional cointegration rank — deterministic trend . . . 165
6.12 Simulation results order of integration — deterministic trend and no fractional cointegration 166
6.13 Simulation results order of integration — deterministic trend and fractional cointegration 167
6.14 Simulation results estimating d when rk is unknown — no fractional cointegration 169
6.15 Simulation results estimating d when rk is unknown — fractional cointegration 170

Chapter 1

Introduction

The return and the risk of an asset or portfolio are two of the most important variables in financial decision making. Modeling and forecasting them is therefore of vital importance for academics and practitioners alike.

While the definition of asset returns is rather undisputed, there exist various measures for the risk of an asset. Until the 1960s the variance or standard deviation, commonly referred to as volatility, of asset returns was the most popular risk measure (Markowitz, 1952). The Capital Asset Pricing Model (Sharpe, 1964; Lintner, 1965; Mossin, 1966) and the arbitrage pricing theory (Ross, 1976) extend this view. They relate the risk of an asset to its sensitivity to economy wide risk factors, often referred to as beta. This beta captures the systematic part of an assets risk while volatility also includes non-systematic risk, which can be eliminated by diversification. For asset pricing and capital allocation decisions, beta is therefore the key risk measure nowadays. For risk management and option pricing, however, volatility also plays an important role.

While the calculation of asset returns is straightforward, neither the volatility of an asset nor its beta can be observed. Instead, both need to be estimated based on past asset prices. Clearly, from a statistical point of view, the precision of these estimates increases with an increasing number of observations. On the other hand, it is well established that both measures vary over time so that considering long time spans likely results in biased estimates. This causes a bias-variance trade-off which has been circumvented for a long time by imposing restrictions using parametric models (e.g., Bollerslev, 1986).

In recent years the increased availability of high-frequency data gave rise to nonparametric approaches for the estimation of volatility and beta. Nowadays, in a liquid market, the number of observations within a day equals the amount of daily data collected in 20 to 30 years. Consequently, long time spans are not needed to obtain a large number of observations for estimation. This is the basis for the concept of realized variance (RV). Here, the variance of an asset during some interval is estimated by summing up all squared returns of the asset during this interval. Given certain assumptions on the price process, Barndorff-Nielsen and Shephard (2001) show that a consistent estimate of the true volatility can be constructed using this approach. Based on this idea it is also possible to calculate realized betas (Barndorff-Nielsen and Shephard, 2004a), which allow for a better estimation of the systematic risk component.

It is, however, not only the estimation of the risk of an asset that has been improved by the availability of high-frequency data. Several authors find that high-frequency data can improve forecasting asset returns. Chernov (2007) and Bollerslev et al. (2009), for example, introduce a measure of aggregate risk aversion based on high-frequency observations that is shown to be a significant predictor for the level of monthly stock returns.

Given the recency of its availability, the opportunities high-frequency data offers for modeling and forecasting the return and the risk of assets or portfolios are far from exhaustedly explored. This thesis presents several approaches and methods that can help to close this gap.

Chapter 2 presents an approach to forecast the sign of daily stock returns. For this purpose, we use a simple logistic regression model and a variety of explanatory variables from which the majority is based on high-frequency measures which were not available before 1996. We find statistically significant evidence of sign predictability for 19 out of 26 stocks that were part of the Dow Jones Industrial Average in 1996. This predictability is time consistent and not restricted to certain periods. It can further be exploited by a market-timing strategy to generate abnormally high returns which cannot be explained by standard risk based models. These findings are unexpected as there seems to be a consensus among researchers that, when measured at high frequencies such as days, the returns of liquid stocks are unpredictable. This is based on popular theories and models such as the random walk theory (Regnault, 1863; Malkiel, 1973), the efficient market hypothesis (Fama, 1970), and consumption based asset pricing models (e.g., Lucas, 1978; Cochrane, 2009), which all rule out the possibility of significant daily return predictability.

Chapter 3 and 4 deal with opportunities that high-frequency data offers for the estimation of volatility.

In Chapter 3, we propose an approach to improve the estimation of portfolio volatility, in particular the volatility of economy wide risk factors. While RV has become a standard tool for liquid individual stocks, this measure is difficult to obtain for economy wide risk factors such as the size and value factor of Fama and French (1993). These contain much more illiquid stocks that are simply not traded often enough to calculate RVs. Practitioners or researchers that need to estimate factor volatilities therefore still rely on estimates based on daily data so that they face the bias-variance trade-off mentioned above (e.g., He et al., 2015; Moreira and Muir, 2017). Our approach improves on this issue as it yields an estimate of factor volatility that is close in precision to RV. This is achieved by approximating high-frequency factor returns by a linear combination of the returns of over 1,000 liquid stocks. Due to the large number of parameters in the linear combination that have to be estimated, it is necessary to apply a regularized estimation method such as ridge regression. We correct for the bias induced by this method and then perform a RV-type estimation using the approximated high-frequency factor returns. The efficacy of this approach is demonstrated using Monte Carlo simulations and forecasts of the variance of the market factor, where we can use the realized variance of the S&P 500 to evaluate the accuracy of our forecasts.

Despite being nonparametric, RV requires certain assumptions on the asset's price process to provide a consistent estimate of volatility. These include that markets are frictionless and that asset prices behave as semimartingales. While the consequences of market frictions have been largely investigated (e.g., Zhou, 1996; Hansen and Lunde, 2006; Bandi and Russell, 2008), the validity of the semimartingale assumption has been widely accepted so far. In a recent paper, Gao et al. (2018) find contrary evidence by showing that significant dependencies between aggregates of returns within a trading day exist for the S&P 500. The analysis in Chapter 4 shows that this correlation negatively biases the RV estimate causing an underestimation of the variance of daily S&P 500 returns by an average of 15 percent. Furthermore, it is revealed that the effect is not only present for the S&P 500 but also for a wide range of other international stock indices, such as the SSEC, the BSESN, and the N225. On average, the variance of the daily stock index returns is underestimated by 14 percent. By providing a detailed investigation of the source of the bias, Chapter 4 further shows that an alternative unbiased RV estimator can be constructed in the spirit of Hansen and Lunde (2005).

Chapter 5 and 6 deal with the systematic risk component. In contrast to the essays on volatility discussed above, the focus in these chapters is not on estimating the risk measure but on modeling and forecasting the estimated risk measure series. For these purposes, it is of major importance to adequately capture the dependency structure of the series. In empirical applications it is often found that volatility series show long-memory properties, i.e., the impact of shocks dies out slowly which creates a hyperbolically decaying autocorrelation function that solely depends on the memory parameter d for large lags. It is natural to ask whether this is also true for beta series.

Chapter 5 seeks to answer this question. For this purpose, we consider the monthly realized beta series relative to the market factor of over 800 stocks. We find that the realized beta series of the vast majority of stocks show long-memory properties. Furthermore, it is revealed that by incorporating this characteristic into the model, we can improve the accuracy of forecasts of the realized beta series. This is done by showing superior performance of the forecasts of a long-memory model compared to the forecasts of short-memory and random walk models, which ignore the long-memory properties, but yet have been commonly considered in the literature for forecasting realized betas. We further document the relation of firm characteristics with the forecast error differentials that result from inadequately imposing short-memory or random walk instead of long-memory processes.

Having documented that the degree of dependence plays an important role for modeling and forecasting asset risk, we suggest a new approach for estimating the memory parameter d in Chapter 6. The risk of companies continuously changes over time as they develop, create new products, and enter new markets. Certain events such as, for example, selling parts of the company, however, cause abrupt changes in the risk series. As shown by Granger and Ding (1996) and Diebold and Inoue (2001), among others, such structural changes bias the estimate of d. Forecasts based on this estimate would then also be biased and inaccurate. Iacone (2010) suggests an approach to robustify the estimation concerning the presence of structural changes. This estimator is, however, univariate. It is well known that working in a multivariate framework can result in efficiency gains and is therefore preferable where suitable. In Chapter 6, we therefore extend the approach by Iacone (2010) to a multivariate setting. As our estimator requires a priori knowledge of the cointegration rank of the series, i.e., whether there exists a linear combination of the series whose memory parameter is smaller than that of the original series, and structural changes also bias inference on the cointegration rank, we also provide a robust estimator of the cointegration rank. The efficacy of both estimators is demonstrated by Monte Carlo simulations and an application to the daily realized beta series of two American energy companies. Here, we discover that the series are fractionally cointegrated which standard procedures are unable to detect.

Although many economic time series, such as volatilities and betas, show long-memory properties, long-memory models are often ignored in the literature. This is likely due to the fact that at first glance they seem rather complicated and hard to implement. To remedy this, we provide the R package *memochange* which is presented in Chapter 7. The package includes several change-in-mean tests that are applicable under long memory as standard change-in-mean tests are invalid in this case. These can be used to determine whether a beta series truly exhibits an abrupt change as mentioned above. Moreover, the package contains various tests for a break in persistence. These can be used to detect a change in the memory parameter.

Chapter 2

Directional Predictability of Daily Stock Returns

Co-authored with Christian Leschinski.

2.1 Introduction

While the debate about predictability on longer time horizons is still ongoing (e.g., Welch and Goyal, 2008; Rapach and Zhou, 2013), there is a consensus that daily returns of liquid stocks are unpredictable. In this paper, we provide contrary evidence and show in an out-of-sample environment that the sign of daily returns is predictable. This effect is so sizeable that trading strategies exploiting it yield an annualized alpha of more than eight percent after transaction costs.

Asset return prediction is of vital importance for practitioners and academics alike. In its semi-strong form, the efficient market hypothesis (EMH) requires that asset prices fully reflect all publicly available information at all times (Fama, 1970). Price changes can therefore only reflect the arrival of new information, which is unpredictable by definition. This gives rise to the random walk hypothesis for the level of (log-)prices. If we denote the continuously compounded return by r_t, then it can be decomposed as

$$r_t = \pi_t + \mu_t + \varepsilon_t,$$

where π_t denotes the risk-free interest rate, μ_t is the equity premium, and ε_t is a mean-zero innovation term that is serially uncorrelated. If investors are risk-neutral, we have $\mu_t = 0$ and the EMH implies that excess returns should be unpredictable. On the other hand, if investors are risk averse, a certain degree of predictability in the equity premium μ_t is possible in efficient markets as long as the predictability reflects time-varying aggregate risk (Rapach and Zhou, 2013). On a daily frequency, however, both π_t and μ_t are essentially zero since their scale is minuscule compared to the variation of ε_t. The daily return r_t should therefore be unpredictable. The same conclusion can be drawn using consumption based asset pricing models. Ross (2009) and Zhou (2010), for example, derive an upper bound for the potential predictability of stock returns. This bound increases in the variance of the stochastic discount factor. Since on a daily horizon the variance of the stochastic discount factor is small compared to the monthly case, there is no theoretical basis to expect significant predictability of daily stock returns.

In contrast to these considerations, we show empirically that the sign of daily stock returns is in fact predictable. For this purpose, we use a dataset consisting of all stocks that were part of the Dow Jones Industrial Average (DJIA) in 1996 and a simple logistic regression model. The relevant explanatory variables are selected in the subsample from 1996 to 2003 based on a forward selection procedure. Subsequently, the predictive performance of the selected models is evaluated in an out-of-sample environment for the period from 2004 to 2017, where each model is re-estimated in a rolling window to generate one-step-ahead forecasts. The setup therefore mimics the situation a forecaster would face in real time.

Using this approach, we are able to correctly predict 52 percent of the signs of the considered stock returns. This predictability is found to be statistically significant for 19 out of 26 stocks. We further show that the predictability is time consistent and not restricted to certain periods.

To determine the economic significance of these findings, we devise a market-timing strategy that invests in a value-neutral portfolio. The long and short positions are determined with help of the classification model by sorting the stocks according to their predicted probability of a positive return. In absence of transaction costs, this strategy generates an annualized alpha of 16.70 percent compared to the market factor and an alpha of 16.08 percent relative to the five-factor model of Fama and French (2015). These numbers heavily exceed those of previous predictability studies, such as Moreira and Muir (2017), who report an annualized alpha of 4.9 percent. After accounting for realistic transaction costs, we find that the annualized alpha relative to the five-factor model of Fama and French (2015) is still at 7.17 percent.

The magnitude of the alpha generated by our strategy is rooted in the fact that the logistic model itself signals the presence or absence of directional predictability through the dispersion of the predicted probabilities for positive returns in the cross section of stocks. Our strategy only invests if this signal is sufficiently large. Otherwise, no trades are initiated. The strategy therefore only trades on a subset of all predictions. For this subset the hitrate is as high as 54.17 percent, which is considerably higher than the aforementioned 52 percent that is obtained for all stocks and trading days.

While there is no previous contribution in the finance literature that successfully forecasts the level or the sign of daily stock returns in an out-of-sample environment, our finding is in line with those of Linton and Whang (2007) and Han et al. (2016), who provide statistical evidence of directional predictability in daily stock returns. These findings, however, are in-sample and based on nonparametric tests that do not allow to generate actual forecasts.

The remainder of this paper is organized as follows. The next section introduces the dataset and the explanatory variables. Afterwards, we discuss our model selection and forecasting procedure in Section 2.3. Section 2.4 then reports the empirical results showing the statistical as well as economic significance of our forecasts. Furthermore, the direction of the influence of the explanatory variables is explored. Section 2.5 contains additional analyses and robustness checks and Section 2.6 concludes.

2.2 Data

We consider 5-minute data for $N = 30$ stocks included in the DJIA on January 1, 1996 and several explanatory variables. The dataset is obtained from the Thomson Reuters Tick History database and ends on January 31, 2017.

Since high-frequency data is often subject to minor recording mistakes, it is common practice to apply some form of data cleaning. Here, we adopt the approach of Barndorff-Nielsen et al. (2009), which comprises, among other things, the removal of observations with negative stock prices and abnormal high or low entries in comparison to other observations on the same day. The resulting cleaned dataset is then used to calculate daily stock returns and more than 20 explanatory variables.[1]

For modeling and forecasting purposes, we calculate logarithmic returns that facilitate the computation of explanatory variables such as moving averages or realized variances. Even though the differences between discrete and logarithmic returns are small due to the short time horizon, the logarithmic returns are transformed to discrete returns when considering trading strategies in Section 2.4.2. This allows for the calculation of portfolio returns.

Our analysis is based on the companies that were components of the DJIA in the beginning of our sample in January, 1996. Over the course of time, several companies faced bankruptcy or were taken over so that the respective time series end before 2017.[2]

2.2.1 Explanatory Variables

Since predictability of daily stock returns has been ruled out in the literature, there is no established set of potential explanatory variables. We therefore focus on variables that (i) exhibit meaningful variation on a daily frequency, (ii) are easily available, and (iii) for which a plausible economic argument can be made or where predictability on a longer horizon has been found in previous studies.

---

[1] Since transaction costs tend to be higher at market closing time than they are five minutes before and we intend to use the resulting forecasts for trading purposes, daily returns are calculated using the closing prices at 3:55pm each day.

[2] This concerns Bethlehem Steel, Eastman Kodak, Sears Roebuck, Texaco, Union Carbide, and Westinghouse Electric. General Motors went bankrupt in 2009 but returned to the New York Stock Exchange only one year later.

This includes measures of moments of the return distribution, such as log-realized variances and realized skewness, since Christoffersen and Diebold (2006) argue that directional predictability can be generated by persistence in $|r_t|$ if $\mu_t \neq 0$, and Christoffersen et al. (2007) argue that variation in higher moments can generate directional predictability even if $\mu_t = 0$, as long as the return distribution is asymmetric.

Market measures, such as S&P 500 returns and realized market betas, are considered because the CAPM implies a strong relationship between market and stock returns. The log-realized variance of the S&P 500 and the VIX are included for the same reasons as the other risk measures mentioned before.

A number of recent studies such as Chernov (2007), Bollerslev et al. (2009), and Bollerslev et al. (2014) introduce the variance premium that is defined as the difference between the implied variance under the assumption of risk neutrality and the conditional expectation of the variance. This measure is related to the aggregate risk aversion and it is shown to be a significant predictor for the level of monthly stock returns. Therefore, it is included as well.[3]

The yield curve is found to be predictive for future macroeconomic activity (Ang et al., 2006b), which should also influence stock prices. Furthermore, Koijen et al. (2017) show that risk factors driving bond yields are also priced in the cross section of stock returns. We therefore include the first three principal components (PC) of the yield curve (along with their changes) that can be interpreted as level, slope, and curvature of the curve. These are calculated using U.S. bonds with over 40 different maturities.

Finally, we consider some technical indicators used in the machine learning literature. Using these indicators, Kara et al. (2011) and Qiu and Song (2016) report that they are able to correctly classify 70-80 percent of the returns. Even though this is theoretically implausible and points to an issue with overfitting, we take this as a reason to include them in the set of potential explanatory variables. This results in the following list of 24 explanatory variables.

- **(Realized) Measures of Moments:** log-realized variance (e.g., Amaya et al., 2015), high-low variance (Corrado and Truong, 2007), realized skewness (e.g., Amaya et al., 2015).
- **Financial Market Indicators:** S&P 500 return, realized betas calculated from S&P 500 5-minute returns, log-realized variance S&P 500, level VIX, VIX return, oil return.
- **Risk Aversion Indicators:** variance premium (Bollerslev et al., 2009).
- **Yield Curve Measures:** level and change of first PC (level of the yield curve), second PC (slope of the yield curve), and third PC (curvature of the yield curve).

---

[3] We estimate the conditional expectation of the variance using the HAR model of Corsi (2009).

- **Technical Indicators:** stock return, 5-day moving average stock return, on-balance volume, 12-day moving average of binary stock returns, momentum indicator, A/O oscillator, and rate-of-change indicator (Qiu and Song, 2016).[4]

Detailed explanations and discussions of these variables can be found in the referenced articles. Summary statistics and the correlation matrix of the variables are shown in Table 2.4 and Figure 2.6 in the Appendix.

2.3 Methodology

In this section, we describe our model selection and forecasting framework and discuss the econometric reasoning behind the respective modeling choices.

2.3.1 Model Framework

In contrast to regression problems, where the dependent variable can take values in $\mathbb{R}$, directional predictability is a classification problem, which means that the dependent variable $y_{i,t+1} = I(r_{i,t+1} > 0)$ takes the value one if the return of stock i at time $t+1$ is positive, and zero otherwise. Therefore, we use predictive logistic regressions for forecasting. The model can be represented in the form

$$y_{i,t+1} = G(x_{i,t}^{'}\beta_i) + \varepsilon_{i,t+1}, \tag{2.1}$$

with $G(v) = \frac{1}{1+\exp(-v)}$. Here, $\varepsilon_{i,t+1}$ is an error term, $x_{i,t}$ is a $(p+1)$-dimensional vector of explanatory variables with 1 as the first element, β_i is a vector of $(p+1)$ parameters, $i = 1, \dots, N$, and $t = 1, \dots, T$. Note that the explanatory variables in $x_{i,t}$ are lagged by one period relative to the dependent variable $y_{i,t+1}$ as the aim is to obtain one-step-ahead forecasts.

The signal-to-noise ratio in stock returns is extremely low, which is challenging for statistical analysis in several ways. First, in-sample results likely overestimate the amount of predictability as the model adjusts to noise components that do not resurface in other periods. Second, we consider a large number of potential explanatory variables so that it is likely that variables appear to be significant due to random covariation with the noise component.

To address the first issue, we evaluate the predictive performance in an out-of-sample environment. In the empirical analysis, the model selection and the forecasting period are strictly separated. While the model is chosen in the years from 1996 to 2003 ($T_M \approx 2,000$

---

[4] Let $C_{i,t}$ be the closing price, $L_{i,t}$ the lowest price, $H_{i,t}$ the highest price, and $V_{i,t}$ the volume of trade of stock i at day t. Moreover, let $\Theta_{i,t} = 1$ if $C_t \geq C_{i,t-1}$, and $\Theta_{i,t} = -1$ otherwise. Then, the on-balance volume (OBV) is given by $OBV_{i,t-1} + \Theta_{i,t} * V_{i,t}$, the momentum indicator by $C_{i,t} - C_{i,t-4}$, the A/O oscillator by $\frac{H_{i,t} - C_{i,t-1}}{H_{i,t} - L_{i,t}}$, and the rate-of-change indicator by $C_{i,t}/C_{i,t-14} \times 100$.

observations per stock), out-of-sample forecasts take place from 2004 to 2017 ($T_F \approx 3,300$ predictions per stock).[5] Obviously, it holds that $T_M + T_F = T$. It needs to be emphasized that this separation entails that our out-of-sample analysis simulates the situation an investor starting to invest in 2004 would have faced in real time. Clearly, the length of these two periods is selected rather arbitrarily. However, we show that the results are robust with respect to the choice of these periods in Section 2.5.3.

To deal with the second issue, we restrict the set of explanatory variables $\mathbb{M}$, which is included in the regressor vector $x_{i,t}$, to be the same across all stocks. This has the advantage that the number of observations for model selection increases drastically compared to the case where each stock is allowed to have a different set of regressors. Consequently, we obtain more stable results. This restriction neither means that the values of the explanatory variables are the same for all stocks nor that the coefficients β_i associated with the variables are identical.

2.3.2 Model Selection

With 24 possible explanatory variables and a low signal-to-noise ratio, we require a model selection procedure to obtain a more parsimonious model. In general, a model selection procedure consists of two components, a goodness-of-fit criterion to evaluate the performance of each candidate model and a rule that defines which models are considered as candidate models.

It is important to note that the objective of this study is to predict the direction of future returns and not to determine the true data generating process driving them. It is well known that these are conflicting objectives. For a discussion of this issue, often referred to as the AIC-BIC dilemma, cf. Arlot and Celisse (2010).

With respect to the choice of the goodness-of-fit criterion, we face several statistical complications. First, daily stock returns are non-Gaussian so that likelihood based information criteria lose their optimality properties. Second, since some of the potential predictor variables are highly persistent but returns are not, there is an unbalanced regression issue similar to those discussed by Stambaugh (1999). We therefore adopt a simple, yet robust approach and evaluate each candidate model based on its actual out-of-sample performance. To achieve this, we consider one-step-ahead forecasts for the latter part of the model selection period generated from the logistic model estimated in an expanding window. We then calculate the proportion of correctly predicted $y_{i,t}$ (the so-called hitrate) in this pseudo out-of-sample experiment. This is a direct estimate of the out-of-sample hitrate conditional on the model, which is exactly the objective of our forecasting exercise.

---

[5] Some companies went bankrupt which results in varying T_M and T_F per stock. Furthermore, days with recording errors (indicated by the procedure of Barndorff-Nielsen et al., 2009) were excluded, which also implies slightly different sample sizes per stock.

More formally, denote the logistic model estimated using the observations from time 1 to time t by $G_{[1,t]}(x_{i,t}^{\prime}\hat{\beta}_i)$. Then, for a given set of explanatory variables $\mathbb{M}$, the average out-of-sample hitrate ($OOSH$) is defined as

$$OOSH(\mathbb{M}) = N^{-1} \sum_{i=1}^{N} \frac{\# \left\{ I \left(G_{[1,t]}(x_{i,t}^{\prime}\hat{\beta}_i) > 0.5 \right) = y_{i,t+1} \right\}_{t \in \{S, \ldots, T_M - 1\}}}{(T_M - S)}. \qquad (2.2)$$

With respect to the set of candidate models, it would be optimal to consider all possible combinations of variables. This approach is referred to as best subset selection. Unfortunately, this would require to evaluate 2^{24} models, which is computationally infeasible. Therefore, we use stepwise forward selection as a computational surrogate.

If the set of all K possible explanatory variables is denoted by $\mathbb{P}$ and $\mathbb{M}_k$ is the set of explanatory variables that is already selected as part of the model in step k, then $\mathbb{P}_k = \mathbb{P} \setminus \mathbb{M}_{k-1}$ is the set of variables that could still be added to the model. Each of these variables is referred to as $\mathbb{P}_{k,j}$. The procedure then proceeds as follows:

0. Initialization:

Set $k = 1$ and $\mathbb{M}_0 = \emptyset$.

1. Forward Selection:

Set $\mathbb{M}_k = \arg\max_{j=1,\ldots,K-k+1} OOSH\,(\mathbb{M}_{k-1} \cup \mathbb{P}_{k,j})$.

2. Model Selection:

If $k < K$, increase k by one and go back to Step 1.
Otherwise set $\mathbb{M} = \arg\max_{k=1,\ldots,K} OOSH\,(\mathbb{M}_k)$ and terminate the procedure.

In simpler terms, we generate a sequence of models by iterating through a procedure that starts out with the empty model in Step 0 and sequentially adds variables to the model until the full set of regressors is used in the K-th model $\mathbb{M}_K$.

In each iteration k of Step 1, we sequentially add each of the remaining variables in $\mathbb{P}_k$ to the $k-1$-variable model from the previous iteration to generate candidate models with k regressors. For each candidate model we calculate the $OOSH$ and then select the model that generates the largest improvement. To reduce the computational effort, the expanding window estimate $G_{[1,t]}(x_{i,t}^{\prime}\hat{\beta}_i)$ is updated only every 40th observation.

Finally in Step 2, we consider the sequence of models of increasing size and select the set of regressors $\mathbb{M}$ that achieves the lowest $OOSH$.

It is important to note that since the findings presented in this paper are based on a pure out-of-sample experiment, their validity is not tied to that of the model selection procedure. In fact, a better model selection procedure could potentially select a better forecasting model and thus produce even stronger results in the out-of-sample environment. However, as discussed above, there are good reasons to assume that the procedure suggested here is the most suitable to select an appropriate forecasting model.

Applying the procedure results in the following list of seven explanatory variables that are selected for the final model.

- S&P 500 Return, VIX Return, Variance Premium, Level Third PC (Curvature), Stock Return, 5-Day Moving Average Return, A/O Oscillator.

Of course, it would be desirable to gain further insights into the form of the dependence between the selected explanatory variables and the sign of the next days return. Unfortunately, the parameters of the models are allowed to vary across stocks and the estimations are carried out in an expanding window so that the size and direction of effects may change over time.

Instead of considering the functional form here, we therefore conduct a full sample analysis later on in Section 2.4.4, once it is established that there actually is significant predictability.

2.3.3 Forecasting Procedure

After selecting the set of explanatory variables based on the model selection period from 1996 to 2003, the actual tests of directional predictability are carried out for the out-of-sample period from 2004 to 2017.

Based on (2.1) we generate one-step-ahead forecasts using

$$G_{[T_M-W+t+1,T_M+t]}(x_{i,t}^{'}\hat{\beta}_i) = \hat{P}(y_{i,t+1}=1|x_{i,t}) \tag{2.3}$$

for $t = 0, \dots, T_F - 1$, where $x_{i,t}$ contains the seven variables in $\mathbb{M}$ as presented in the previous section. This means the model is re-estimated in each period using a rolling window of the previous $W = 1,000$ observations.

As indicated by the equation, each value of $G_{[T_M-W+t+1,T_M+t]}(x_{i,t}^{'}\hat{\beta}_i)$ is an estimate of the probability that there is a positive return on the next day. To convert these probabilities into actual forecasts for $y_{i,t+1}$, we need to define a threshold above which we predict the next days return to be positive.

The simplest threshold is the Bayes classifier that assigns

$$\hat{y}_{i,t+1} = I\left(G_{[T_M-W+t+1,T_M+t]}(x_{i,t}^{'}\hat{\beta}_i) > 0.5\right),$$

with $I(\cdot)$ being the indicator function. This decision rule is optimal since the loss of falsely classifying positive and negative returns is equal. Note that this threshold has already been used in Equation (2.2).

2.4 Empirical Analysis

In the following, we evaluate the forecasting performance of the logistic model compared to two naive benchmarks in Section 2.4.1. Section 2.4.2 then addresses the economic significance of the results, 2.4.3 investigates the stability of the results over time, and Section 2.4.4 analyzes the influence of the selected variables on the predicted probability of a positive return on the next trading day.

2.4.1 Forecast Results

To evaluate the performance of the forecasts generated by the procedure described in Section 2.3.3, it is helpful to consider the performance of a naive benchmark forecast.

In the level predictability literature, the naive benchmark is typically the expanding mean of previous returns (Welch and Goyal, 2008). This is because on monthly or longer horizons the average return should be equal to the equity premium. Since the size of the equity premium is unknown, the expanding mean is the best available estimate of the unconditional expectation.

For directional predictions there is no such established benchmark. We therefore consider two possible approaches.

A first naive approach could be to assume that positive and negative returns are equally likely. This would correspond to a random walk model without drift and with a symmetric innovation distribution. Although the scale of the equity premium and the risk-free rate is minuscule compared to the variation of the returns, theoretically both should be slightly positive. This introduces a positive drift. Furthermore, daily stock returns are typically found to have slightly negative skewness (Christoffersen, 2003). Therefore, positive returns can occur more often than negative returns, even if the mean of the returns is zero. Taking these two arguments together, the hitrate obtained by predicting a positive return for each day should be higher than randomly predicting positive and negative returns with equal probability. We refer to this benchmark as the *optimist* forecast.

As the second benchmark, we consider an analogue to the expanding mean used in the level predictability literature. This forecast, referred to as the *historical majority* forecast, is obtained by predicting a positive or a negative return depending on whether the majority of the previous returns of the stock in an expanding window was positive or negative.

The quality of the predictions will be judged based on the hitrate (HR), the sensitivity (SE), and the specificity (SP). For each stock these are defined as follows:

$$\begin{aligned} HR_i &= \frac{\sum_{t=1}^{T_F} I(\hat{y}_{it} = y_{it})}{T_F}, \\ SE_i &= \frac{\sum_{t=1}^{T_F} I(\hat{y}_{it} = y_{it} = 1)}{\sum_{t=1}^{T_F} I(y_{it} = 1)}, \\ \text{and} \quad SP_i &= \frac{\sum_{t=1}^{T_F} I(\hat{y}_{it} = y_{it} = 0)}{\sum_{t=1}^{T_F} I(y_{it} = 0)}. \end{aligned}$$

The hitrate is therefore simply the proportion of correctly classified returns, the sensitivity is defined as the proportion of positive returns that are correctly classified, and the specificity is the corresponding proportion for negative returns. Consequently, hitrate, sensitivity, and specificity lie between zero and one with higher values indicating better classification performance.

The results of the forecasting procedure for each stock are shown in Table 2.1. It can be seen that the hitrate of the logistic model varies between 50.06 percent for Industrial Paper (IP) and 53.27 percent for General Electric (GE). The hitrate of the historical majority forecast ranges from 48.28 percent for Caterpillar (CAT) to 53.36 percent for Eastman Kodak (EK) and the hitrate of the optimist forecast ranges from 46.64 percent for EK to 53.61 percent for Altria (MO). On average, the logistic model achieves a hitrate of 51.99 percent while that of the historical majority benchmark is 50.03 percent and that of the optimist benchmark is 50.86 percent. With regard to the sensitivity and the specificity, it can be observed that the sensitivity typically outweighs the specificity, except for those stocks where the proportion of positive returns in the sample (the hitrate of the optimist forecast) is below 50 percent. The model therefore overpredicts the majority class. A similar observation is made by Nyberg (2011), who predicts monthly stock index movements using a dynamic probit model.

Overall, it can be stated that the hitrate achieved with the logistic model is higher than that of the benchmark models. This holds for the majority of stocks as well as for the average across all stocks.

Testing whether this difference is significant can be done using a Diebold and Mariano (1995) type test statistic. Denote the forecasts of two competing models by $\hat{y}_{i,t}^{(1)}$ and $\hat{y}_{i,t}^{(2)}$. Then, the loss differential between these forecasts at day t for stock i is given by

$$l_{i,t} = \left(y_{it} - \hat{y}_{i,t}^{(1)}\right)^2 - \left(y_{i,t} - \hat{y}_{i,t}^{(2)}\right)^2 .$$

				DM test		
RIC	HR	SE	SP	HM	Opt	PT test
AA	51.95	43.17	60.46	50.76	49.24**	2.11**
AXP	52.68	56.90	48.29	50.27**	51.00*	2.98***
BA	52.14	55.44	48.62	49.77**	51.66	2.33**
AT	51.69	57.45	45.64	48.28***	51.23	1.78**
CVX	52.66	75.70	27.50	50.32**	52.21	2.09**
D	50.75	57.75	43.27	48.37**	51.63	0.59
DIS	51.75	58.33	44.82	50.11*	51.29	1.82**
EK	53.17	23.75	78.87	53.36	46.64***	1.41*
FL	52.18	48.52	55.90	49.59**	50.41*	2.54***
GE	53.27	45.04	61.39	50.35***	49.65***	3.73***
GM	52.41	27.59	74.92	52.44	47.56***	1.53*
GT	50.78	44.24	57.45	49.47	50.53	0.98
HON	52.49	62.07	42.26	50.40**	51.64	2.54***
IBM	52.33	51.94	52.73	50.05**	50.90	2.68***
IP	50.06	52.54	47.50	49.18	50.82	0.02
JPM	52.39	49.76	55.05	49.74**	50.26**	2.76***
KO	53.09	58.93	47.07	48.92***	50.71**	3.47***
MCD	51.81	77.90	22.43	50.08	52.96	0.23
MMM	52.23	64.78	38.51	49.89**	52.23	1.96**
MO	53.27	78.60	24.00	53.09	53.61	1.78**
MRK	51.41	47.58	55.28	48.80***	50.23	1.65**
PG	52.20	60.66	43.53	50.62*	50.62*	2.44***
S	50.49	71.14	30.77	51.15	48.85	0.36
T	50.85	51.16	50.53	48.69**	51.31	0.97
UTX	51.06	61.74	40.22	49.30*	50.40	1.15
XOM	51.87	63.80	39.45	50.44*	50.99	1.93**
Average	51.99	55.99	47.86	50.03***	50.86**	-

Table 2.1: Forecast results.
RIC states the identification code of the stock used by Thomson Reuters. The left panel shows the hitrate (HR), sensitivity (SE), and specificity (SP) in percent for the logistic model. The panel in the middle shows the hitrate in percent for the historical majority forecast (HM) and the optimist forecast (Opt). *, **, and *** indicate whether the logistic forecasts are significantly better than the respective benchmark according to a one-sided Diebold–Mariano (DM) test at the ten percent, five percent, and one percent level, respectively. The right Panel show the PT test results for which it holds that under the null hypothesis the realizations are not predictable using the model under consideration and under the alternative they are predictable.

A simple Diebold–Mariano statistic that is asymptotically standard normal is then given by

$$DM_i = \sqrt{T_F}\frac{\sum_{t=1}^{T_F} l_{i,t}}{\sqrt{Var(l_{i,t})}}.$$

Another test specifically designed to test the null of no directional predictability is the test by Pesaran and Timmermann (1992). Under the null hypothesis the forecasts $\hat{y}_{i,t}$ and the realizations $y_{i,t}$ are independent, which implies that the realizations are not predictable using the model under consideration. Under the alternative hypothesis there is a positive relationship between $\hat{y}_{i,t}$ and $y_{i,t}$. This would imply that the direction of the returns is, to some extent, predictable. The test statistic is given by

$$PT_i = \frac{\sqrt{T_F}(SE_i + SP_i - 1)}{\left(\frac{\bar{Z}_{y_i}(1-\bar{Z}_{y_i})}{\bar{y}_i(1-\bar{y}_i)}\right)^{1/2}},$$

where sensitivity and specificity are defined as above, $\bar{y}_i$ is the average class of the stock, and $\bar{Z}_{y_i}$ evolves as $\bar{Z}_{y_i} = \bar{y}_i \cdot SE_i + (1 - \bar{y}_i) \cdot (1 - SP_i)$.

The test is consistent and asymptotically standard normal distributed under the null hypothesis. In case of no predictability the values of sensitivity and specificity will sum up to one meaning the numerator and consequently the test statistic will be close to zero. If a positive relationship between the forecasts and the realizations is present, it holds that $SE_i + SP_i > 1$ and hence $PT_i \to \infty$ for $T_F \to \infty$.

It needs to be noted that the Diebold–Mariano test can be applied for both, the cross section as well as individual stocks. The test by Pesaran and Timmermann (1992), on the other hand, can only be applied for univariate time series.

With regard to the Diebold–Mariano test, the null of equal predictive accuracy is rejected at the 10 percent level in 18 cases for the historical majority benchmark and in 9 cases for the optimist benchmark. Moreover, the hypothesis of equal predictive accuracy relative to both benchmarks is rejected if tested jointly for the whole cross section, which is shown in the last row of Table 2.1. For the PT test, we find that the statistic is positive for all stocks indicating that there is some extent of predictability. This predictability is reported to be significant at the 10 percent level for 19 out of 26 stocks.

Overall, it can therefore be concluded that there is strong evidence for the statistical significance of directional predictability. We need to emphasize that finding significant predictability for only one stock might already allow for economically meaningful trading strategies as long as this stock can be identified ahead of time.

2.4.2 Economic Significance

The previous section showed that signs of daily stock returns are, to some extent, predictable. The obvious question raised by this finding is whether this predictability is of a magnitude that is economically meaningful so that it can be exploited to generate abnormal returns. This section therefore addresses the design and performance of trading strategies based on sign predictability.

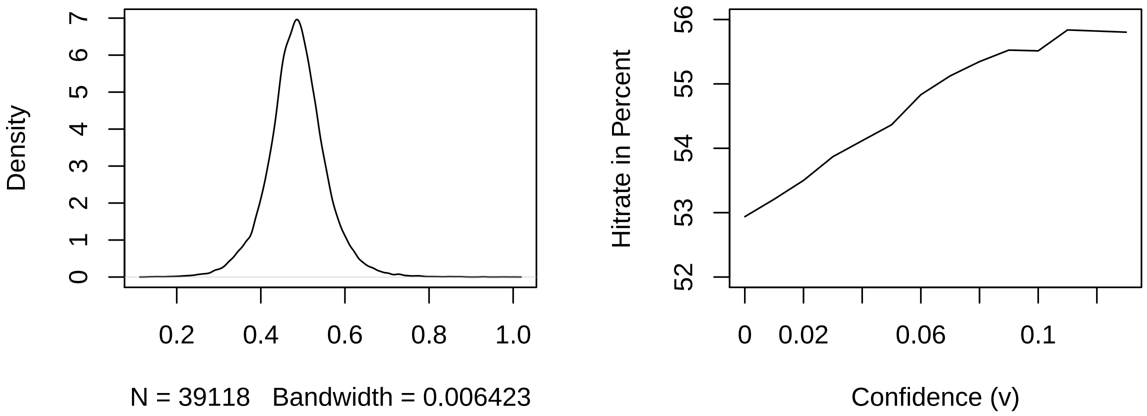

Figure 2.1: Density and model confidence plots.
The left graph shows the density of the predicted probabilities in the model selection period from 1996 until 2003 across all stocks. The right graph shows the hitrate in percent during this period conditional on the difference of the forecasts from the threshold 0.5.

Directional forecasting is as an attempt to time the market. It is therefore obvious to buy stocks which are expected to have positive returns and to sell stocks which are expected to have negative returns. This is the basis for a long-short equity strategy that trades a value neutral portfolio with zero net-investment and where the market risk is hedged. There are, however, several a priori considerations that have to be addressed in the design of the trading strategy.

As mentioned before, the logistic model does not only produce a binary forecast $\hat{y}_{it}$ that specifies whether the stock is expected to have a positive or negative return. Instead, Equation (2.3) gives an estimate of the probabilities $\hat{P}(y_{i,t+1} = 1|x_{i,t})$ that the stock return will be positive.

The plot on the left hand side of Figure 2.1 shows the density of these predicted probabilities in the pseudo out-of-sample part of the model selection period. It can be seen that the majority of the probabilities are in the neighborhood of 50 percent. Obviously, the hitrate that can be expected from directional forecasts based on these probabilities will also be close to 50 percent. Conversely, if we only consider predictions for which the probability to be positive is further away from 50 percent, then the hitrate will be higher. This can be seen on the right hand side of Figure 2.1 that shows the hitrate in the pseudo out-of-sample part of the model selection period for those stocks where the probability to be positive was at least v percent higher or lower than 50 percent. We refer to this distance v of the predicted probabilities from the 50 percent threshold as the *confidence* of the prediction. It is clear to see that the hitrate for those stocks that are predicted to have a probability to move up of at least 60 percent or at most 40 percent is nearly 56 percent. This implies that accounting for the probabilities leads to higher hitrates indicating a larger degree of predictability than that reported in the previous section. In other terms, the graph gives strong evidence that the model is able to identify stocks and periods for which directional predictability exists ahead of time.

Now, with regard to the design of the trading strategy, there is a trade-off between trading few stocks for which the model generates a strong signal so that the hitrate and therefore the average return per trade are high, and trading many stocks to reduce the variability of the traded portfolio.

Furthermore, it should be noted that a strategy that requires daily trading will generate much higher transaction costs than a strategy with monthly portfolio rebalancing. It is therefore crucial to keep the portfolio turnover low and to carefully consider the effect of transaction costs on the performance of the portfolio.

It would be possible to consider a strategy where all stocks for which the probability of a positive return is higher than $50 + v$ percent are bought and where all stocks for which the probability is lower than $50 - v$ percent are sold. This could, however, cause situations where trading is suspended, because all stocks are predicted to have positive returns with a probability of more than $50 - v$ percent so that the short portfolio is empty. The same would be possible for the long portfolio. These effects can arise even though the dispersion among the predicted probabilities is high so that we would have a strong signal to trade on. If, for example, stock A is predicted to increase with probability 95 percent and stock B is predicted to increase with probability 51 percent, then buying A and selling B will on average generate positive returns. To capture this effect, the strategy is formulated in terms of the difference w in the probability of up-movements between pairs of stocks.

Assume, without loss of generality, that there is an even number of stocks N. The strategy is then implemented as follows. For each day $t = 1, \dots, T_F$:

1) Sort the stocks in ascending order according to the predicted probabilities $\hat{P}(y_{i,t+1} = 1|x_{i,t})$ from (2.3) and denote the rank by $r = 1, \dots, N$.

2) Form pairs of stocks so that the stock with the lowest probability to have a positive return ($r = 1$) and the stock with the highest probability ($r = N$) are matched together, the stock with the second lowest and highest probability ($r = 2$ and $r = N - 1$) are matched together, and so on.

3) For all m pairs where the difference between the probabilities to have a positive return is at least w percent, buy the stock that is more likely to go up and sell the stock that is less likely to go up.

This strategy allows us to trade on the cross-sectional dispersion in the predicted probabilities for positive returns. If the probabilities are close to 50 percent for all stocks, trading is suspended. If there is a strong signal for a low number of stocks, the number of pairs that is traded is kept low, and if the signal is strong for a large number of stocks, the traded portfolio contains a large number of stocks.

To avoid trading stocks that are distressed, trading is suspended for those stocks where the log-realized variance at the previous day was more than 5 standard deviations larger than the cross-sectional average.[6]

To consider the effect of transaction costs on the performance of the directional trading strategy, we follow Bajgrowicz and Scaillet (2012) and Moreira and Muir (2017) and deduct a fee to account for the commission, spread, and market impact of each trade. Historically, transaction costs have been steadily declining (Hasbrouck, 2009). Recent estimates of transaction costs by De Groot et al. (2012) range from 5bps for large cap stocks to 50bps for small cap stocks. Since the components of the DJIA are liquid large cap stocks where the market impact is typically low, the 5bps can be expected to be the most appropriate estimate for our setup.

Table 2.2 reports the trading returns of the strategy outlined above with $w = (0.15, 0.17, 0.19, 0.21)$. In addition to the mean and standard deviation of the trading returns, the table presents several commonly used performance measures. This includes the Sharpe ratio and CAPM-alphas as well as alphas for the five-factor model considered by Fama and French (2015).[7]

As a benchmark, we report performance measures for buying and holding the S&P 500 and the DJIA. Both of these benchmarks almost doubled between 2004 and 2017 and are closely related to the optimist forecast, since an investor that predicts a positive return for every stock and every trading day could simply buy and hold the index portfolio.

In absence of transaction costs, the table reveals superior performance of the trading strategy compared to the benchmarks for all considered values of w. For $w = 0.15$, for example, the average annualized return of the trading strategy is 15.73 percent, whereas that of the S&P 500, which outperforms the DJIA for the considered time period, is only 7.58 percent. At the same time, the standard deviation of 12.94 percent is lower than that of the S&P 500, which is 18.70 percent. Consequently, the Sharpe ratio of the trading strategy is 3.3 times that of the best performing benchmark.

Cumulated over the trading period, the trading strategy yielded a return of 597 percent, whereas the benchmark only increased by 205 percent during the same period.

The magnitude of these results shines a different light on the extent of directional predictability. We discussed above that the hitrate can be expected to be significantly higher for those days where the model predicts a probability for a positive return that is further away from 50 percent. In fact, the hitrate for all stocks traded by the strategy

---

[6] This rule would have been suitable in the model selection part of our sample to identify the period during which Bethlehem Steel went bankrupt.

[7] Returns of the factors are obtained from the website of Kenneth French.

	Mean	SD	SR	α (CAPM)	α (five factors)	Max TC
Benchmark (S&P 500)	7.58	18.70	0.34	-	-	-
Benchmark (DJIA)	6.72	17.59	0.31	-	-	-
				$w = 0.15$		
0bps	15.73	12.94	1.11	16.70***	16.08***	9bps
5bps	6.77	12.92	0.43	7.74**	7.17**	
10bps	-1.50	12.92	-0.21	-0.54	-1.05	
				$w = 0.17$		
0bps	14.86	12.48	1.08	15.88***	15.44***	11bps
5bps	8.12	12.45	0.55	9.08***	8.68***	
10bps	1.78	12.45	0.05	2.68	2.31	
				$w = 0.19$		
0bps	13.14	11.76	1.00	13.86***	13.30***	14bps
5bps	8.40	11.73	0.61	8.95***	8.43***	
10bps	3.59	11.72	0.20	4.25	3.76	
				$w = 0.21$		
0bps	10.60	10.72	0.87	11.07***	10.62***	16bps
5bps	7.04	10.68	0.54	7.61***	7.19***	
10bps	3.85	10.67	0.25	4.26*	3.86*	

Table 2.2: Annualized performance measures for the trading strategy. SD states the standard deviation, SR the Sharpe ratio, and w states the amount of dispersion across forecasts that is required to initiate a trade. Moreover, α states the intercept of regressing the daily trading return of our strategy on the market factor, respectively the five factors proposed by Fama and French (2015). *, **, and *** indicate that this alpha is significantly larger zero at the ten percent, five percent, and one percent level, respectively. MaxTC is the break-even transaction cost level that would reduce the α in the five-factor model to zero. All measures reported are annualized and all values, except for the Sharpe ratio, are given in percent. As benchmarks, we state the return of buying and holding the S&P 500 respectively the DJIA in the trading period from 2004 until 2017. Furthermore, the returns are also stated after adjusting for transaction costs of 5bps and 10bps per trade, respectively.

with $w = 0.15$ is as high as 54.17 percent, which is higher than that for any single stock and explains the magnitude of the mean returns shown in Table 2.2. This underlines that the model itself indicates the degree of predictability through the magnitude of the probability of a positive return. If this is close to 50 percent, the predictability is low, otherwise it is stronger.

Most importantly, however, a positive alpha is indicated by the CAPM as well as the five-factor model for all considered values of w and it is found to be statistically significant. The alphas relative to the five-factor model are as high as 16.08 percent. This implies that the trading returns cannot be explained by any of the established risk factors. As a comparison, the alpha reported by Moreira and Muir (2017) for volatility-managed

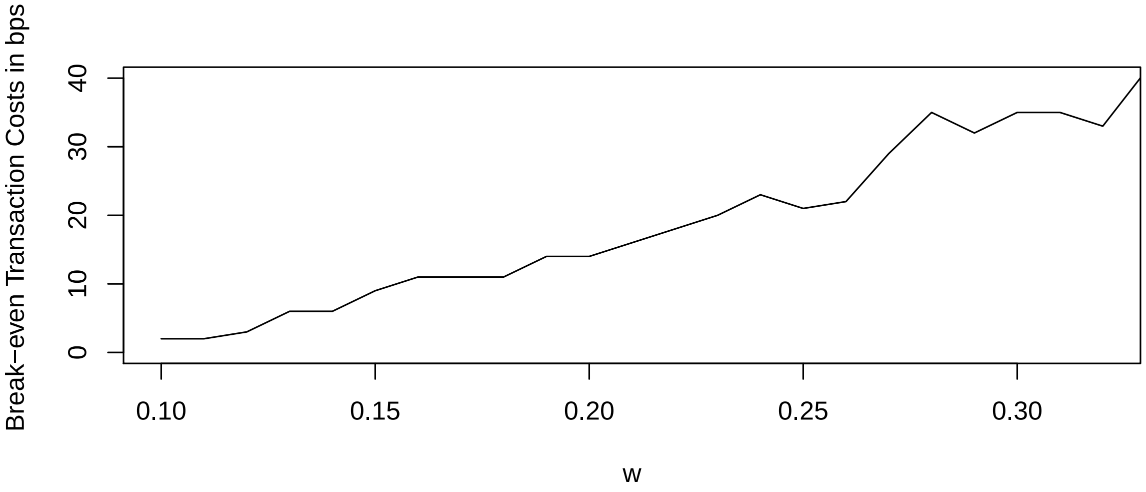

Figure 2.2: Break-even transaction costs.
For different values of w, the break-even transaction costs per trade are plotted so that they set the alpha relative to the five-factor model of Fama and French (2015) to zero.

portfolios is only 4.9 percent before transaction costs, which underlines the magnitude of predictability uncovered here.

With transaction costs of 5bps significant alphas between 7 and 9 percent are obtained for all w. If transaction costs are higher than that, one can observe that the choice of w has an important effect. The higher w, the stronger the signal required to initiate the trade. Therefore, the strategy will trade less often and can be expected to have a high hitrate on those trades that are actually initiated. Therefore, for $w = 0.21$, we obtain significant alpha of about 4 percent even with transaction costs of 10bps. Further insights into this mechanism can be gained from Figure 2.2 that shows the break-even transaction costs that would set the alpha relative to the five-factor model of Fama and French (2015) to zero for different values of w. As one can see, if w is sufficiently large, the strategy generates excess returns even with transaction costs of more than 30bps.

While Section 2.4.1 establishes the fact that directional predictability of daily stock returns is statistically significant in an out-of-sample environment, the findings presented in this section clearly show that this predictability is also economically significant so that it can be used to construct market-timing strategies. Even after accounting for transaction costs, we obtain higher Sharpe ratios than the S&P 500 and significant alphas measured relative to the five-factor model of Fama and French (2015). The positive alphas imply that standard risk-based arguments fail to explain the generated returns. It should be emphasized that the results presented here are based on a relatively small cross section of 30 stocks or less. A trading strategy based on a larger asset universe can be expected

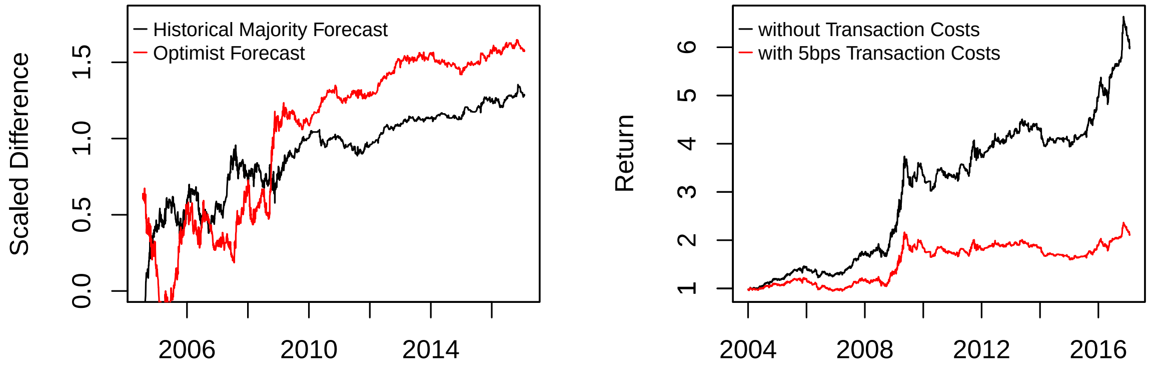

Figure 2.3: Cumulative difference plot and cumulative trading return plot. The left plot shows the rescaled cumulative difference between the hitrate of the logistic model and the hitrate of the respective benchmark for $w = 0.15$. To ensure that the variance of the curve is stable over time, this difference is scaled by the square root of the time index. We also excluded the first 100 observations as these were too variable leading to scales that distort the true relationship. Therefore, the curves do not necessarily start at zero. The right plot shows the cumulative trading return of the trading strategy with and without transaction costs.

to produce even more dispersion among the predicted probabilities and therefore higher trading returns.

2.4.3 Performance of the Trading Strategy over Time

In the level predictability literature, it is often found that predictability is concentrated in recession periods. Welch and Goyal (2008), for example, find a short period of predictability after the oil price shock in 1973. This is also backed by theoretical arguments, for example those of Timmermann and Granger (2004) and Lo (2004), who argue that rational investors will pick up on emerging patterns and exploit them so that it cannot be expected that any forecasting patterns will persist for long periods of time. Therefore, this section investigates the forecasting and trading performance over time.

The left plot of Figure 2.3 presents a cumulated scaled difference plot as a graphical illustration. This can be considered as the classification equivalent to the cumulative plots given in Welch and Goyal (2008). The plot shows the cumulated difference between the hitrate of the logistic model and the benchmarks for all traded stocks for $w = 0.15$. To ensure that the variance of the curve is stable over time, this difference is scaled by the square root of the time index.[8]

As argued in Welch and Goyal (2008), the level of these curves cannot be interpreted but the slope. For periods with positive slopes, it holds that the directional forecasts outperform the benchmark and for periods with negative slopes the opposite holds. Therefore, the plots help to identify whether a model is superior compared to its benchmark for any chosen period by simply comparing the height of the curve at the beginning of the period with the height of the curve at the end of the period.

For both benchmarks, the curves exhibit a predominantly positive slope indicating consistent superior performance of the forecasts made by the logistic model. Consequently, the directional predictability of daily stock returns is not restricted to certain periods. Instead, it constitutes a phenomenon that can be observed consistently over time.

The plot on the right hand side of Figure 2.3 shows the cumulative return of the trading strategy over time for $w = 0.15$. In contrast to the plots on the left hand side, the curves in Figure 2.3 are not scaled since they do not represent means and therefore exhibit equal variance over time. Therefore, we can interpret both, the slope and the level of these curves.

Again, the curve that does not take transaction costs into account has a predominantly positive slope, which shows that the model picks up on a signal that is consistently present over time.

If we take transaction costs into account, the curve is mostly flat, which indicates that the returns from successful directional trades are offset by the transaction costs. This implies that directional predictability is still present but not of a magnitude that is economically significant. However, there are brief periods with extremely high slopes. This concerns mostly the period from 2008 to 2009 and that from 2014 onwards. We therefore find that the economic significance of directional predictability is concentrated in these periods.

Overall, the analysis in this section demonstrates that directional predictability is a permanent feature, at least for the time period investigated here. This is in clear contradiction to the random walk hypothesis and to the best of our knowledge the first approach with the ability to predict *daily* stock returns consistently over such a long time period.

Concerning the trading strategy, it is found that a large proportion of the returns is concentrated in the periods from 2008 until 2009 and from 2014 onwards. While it is tempting to argue that a risk based explanation for this phenomenon may exist in the context of the subprime mortgage crisis, such an explanation seems far fetched for the second period from 2014 onwards. Furthermore, the analysis in Section 2.4.2 already

---

[8] We also excluded the first 100 observations as these were too variable leading to scales that distort the true relationship. Therefore, the curves do not necessarily start at zero.

	Mean	Minimum	Median	Maximum	Positive
S&P 500 Return	-0.03	-0.13	-0.04	0.13	0.35
VIX Return	-0.01	-0.08	-0.01	0.07	0.46
Variance Premium	0.03	-0.01	0.05	0.10	0.96
Level Third PC (Curvature)	-0.03	-0.09	-0.03	0.01	0.08
Stock Return	-0.06	-0.21	-0.06	0.08	0.27
5-Day Moving Average Return	-0.05	-0.15	-0.04	0.04	0.08
A/O Oscillator	0.08	-0.04	0.08	0.22	0.77

Table 2.3: Summary statistics of the coefficients in the logistic regression models.
For all 26 stocks the logistic regression model is estimated separately for the full sample. The column *Positive* then states the percentage of stocks for which the variable is estimated to have a positive impact on the probability of a positive return. As the variables are standardized to have zero mean and unit variance prior to the estimation of the models, the magnitude of the effects is comparable across variables.

controls for the known priced risk factors that are subsumed in the five-factor model of Fama and French (2015).

2.4.4 Economic Implications

From an economic perspective, it is interesting to gain insights into the sources of predictability. In the following, we therefore analyze the influence of the selected variables on the predicted probability of a positive return on the next trading day.

The logistic regression model can be interpreted as a linear model for the log-odds and the sign of the estimated coefficients indicates whether an increase of the respective explanatory variable leads to an increase or decrease in the probability to observe a positive return on the next day.

Due to the rolling window procedure and the fact that the estimated coefficients are allowed to vary across stocks, it is difficult to give a comprehensive overview of the direction of the effects. We therefore report the results of a full sample estimation of the selected model over the period from 1996 to 2017. This gives 26 different models — one for each stock. Due to issues with the nonstationarity of some of the regressors, standard inference is not applicable so that we cannot test whether the estimated effect for a specific regressor is significantly different from zero. Instead, we focus on the interpretation of the signs.

Table 2.3 shows summary statistics for the estimated coefficients. Since the variables are standardized to have zero mean and unit variance prior to the estimation of the models, the magnitude of the effects is comparable across variables. The last column on the right shows the proportion of the estimated coefficients for the respective variable that is positive. It is found that the effects of the A/O oscillator and the variance premium

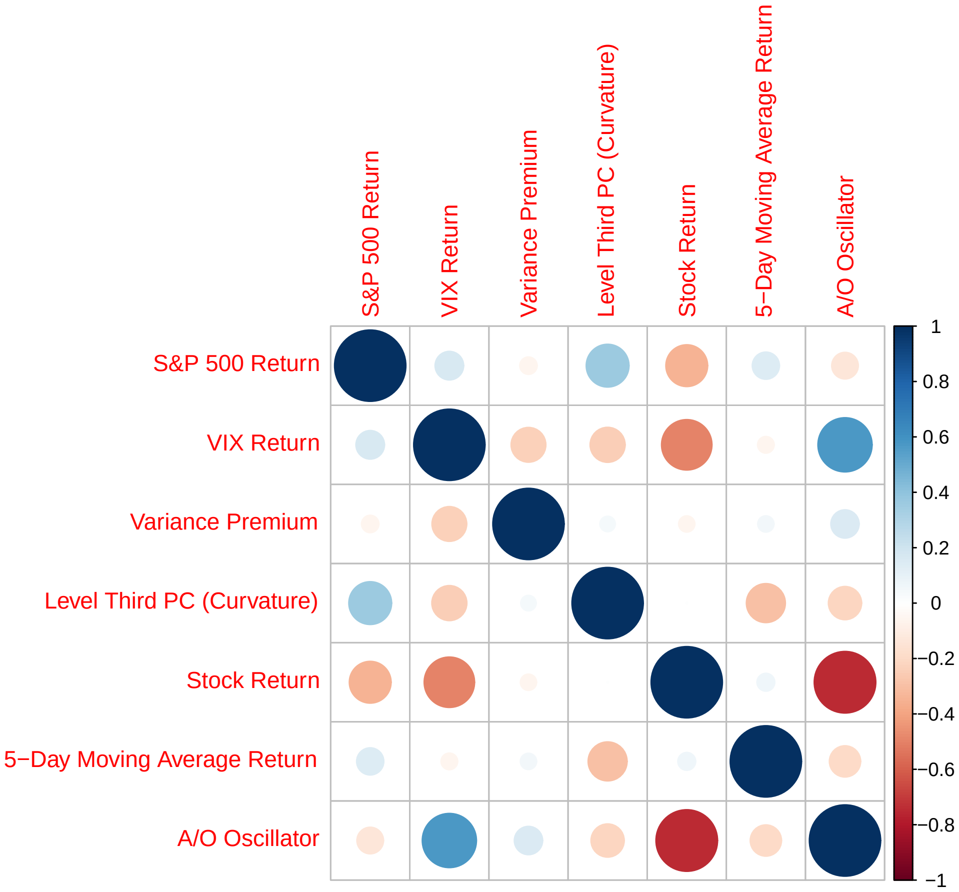

Figure 2.4: Correlation plot for the estimated coefficients from the logistic regression models.
Size and color of the circles correspond to degree and direction of the correlation.

are predominantly positive. Conversely, both the 5-day moving average return and the curvature of the yield curve have a negative influence for nearly all stocks. Both, the lagged return of the S&P 500 and the lagged return of the stock itself have a positive impact on the probability to observe a positive return on the following day for approximately a third of the stocks and a negative impact for two thirds of them. The percentage change of the VIX has a positive impact for approximately half of the stocks.

We therefore find that on average positive returns are more likely if the trading range on the previous day and the variance premium are high, and if the previous returns and the curvature of the yield curve are low. If one restricts the analysis to the model selection period, the results remain qualitatively similar.

The correlation of the estimated coefficients for the different stocks is given in Figure 2.4. It can be seen that there is a negative correlation among the coefficient on the lagged stock return and that of the A/O oscillator and the change of the VIX. This means that stocks that have a higher probability of a positive return after a negative return also tend to have higher returns if the previous day's trading range and the change of expected risk are high. In line with this, there is a positive relationship between the coefficient of the A/O oscillator and that of the return of the VIX.

One finding that strikes out is that the coefficients for Stock Return, 5-day moving average, and for the S&P 500 are negative. This is evidence for individual return reversals and reversals of market-wide returns.

This appears to be in line with previous findings on short-term reversals such as those of Brown and Harlow (1988) and Atkins and Dyl (1990), who find significant reversals for stocks that experience one-day price declines. Cox and Peterson (1994) attribute these reversals to the bid-ask bounce and liquidity effects which implies that short term reversals cannot be exploited by trading strategies. In a more recent contribution, De Groot et al. (2012) show that short term reversal effects are mostly observed for small cap stocks and not for large cap stocks such as the components of the DJIA considered here.

To establish that daily directional predictability and short term reversal are separate phenomena, we construct a short term reversal factor by buying the 20 percent of stocks that performed worst on the previous day and selling the 20 percent of stocks that performed best. Adding this factor to the five-factor model yields annualized alphas of 19.13, 17.38, 15.58, and 10.00 for $w = (0.15, 0.17, 0.19, 0.21)$. As the alphas are similar to those presented in the last section, the abnormal trading returns cannot be explained by short term reversal effects.

2.5 Additional Analyses and Robustness

2.5.1 Further Results on the Behavior of the Trading Strategy

To gain further insights into the behavior of the trading strategy, the first two rows of Figure 2.5 show the predicted probabilities from Equation (2.3) exemplary for four stocks. It can be seen that the model tends to predict a probability for a positive return that is close to 50 percent for most of the time. However, the variation in the predicted probabilities changes significantly over time and there is some degree of persistence in the deviations from 50 percent. It is interesting to note that the variation in the predicted probabilities seems consistently higher for General Electric (GE) and Coca-Cola (KO) than for Goodyear Tire (GT) and Du Pont (DD). When considering the results on the significance of the observed predictability in Table 2.1, it is found that GE and KO, for which the model generates stronger variation in the predicted probabilities, are indicated to be significantly predictable by all tests, whereas GT and DD are not.

The overall dispersion of the forecasts across all stocks is considered in the bottom left plot of Figure 2.5. The figure reveals that the model differentiates especially strong between potential winner- and loser stocks during the subprime mortgage crisis. An effect

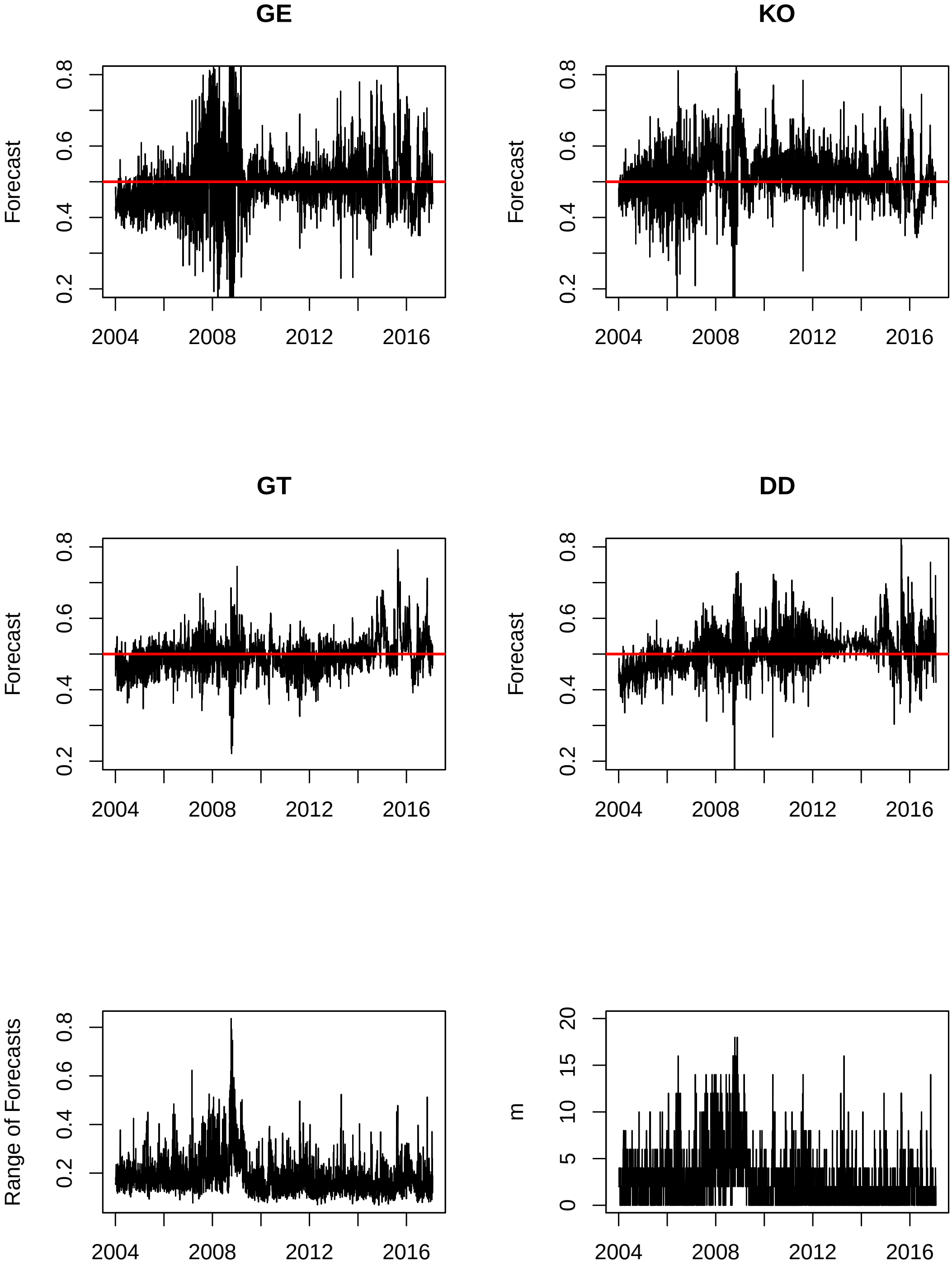

Figure 2.5: Range of the predicted probabilities for a positive return. The graphs show individual forecasts for four stocks (top), the range of the predicted probabilities for positive returns across all stocks (bottom left), and the number of stocks m traded by the strategy with $w = 0.15$ (bottom right).

that can also be seen, although less obvious, in the four plots above. Consequently, the trading strategy invests in a larger number of stocks during this period as the dispersion across stocks is larger. This is confirmed in the bottom right graph of Figure 2.5, where the number of stocks m in which the strategy invests is plotted over time. This behavior corresponds to the more pronounced slopes in the period from 2008 to 2009 in Figure 2.3.

2.5.2 Impact of the Subprime Mortgage Crisis

Section 2.4.3 revealed that a large proportion of the returns of the strategy is generated during the subprime mortgage crisis. This is in line with the known finding that level predictability appears to be concentrated in recessions (Welch and Goyal, 2008). Since our analysis relies on high-frequency data, there are not enough recession periods in the sample to test this hypothesis. Instead, Table 2.5 in the Appendix presents the result of a regression of the daily trading return of our strategy on the probability of a bull market obtained through a Markov switching mean-variance model with two regimes corresponding to a bull and a bear market. The coefficient of the bull market probability is negative, which implies that the trading returns are indeed higher in bear markets. It should be noted, however, that the intercept is still significantly larger than zero so that the market sentiment alone is not able to explain the average returns generated by the trading strategy.

To further investigate the impact of the subprime mortgage crisis period on our overall results, we repeat the analysis from Table 2.2 excluding the period from August 2008 until March 2009 (where the market dropped by 43 percent) completely from the sample. As already demonstrated in Figure 2.3, the statistical evidence for directional predictability is remarkably stable over time. The results in Table 2.6 in the Appendix underline that even though the economic significance of directional predictability was particularly pronounced during the subprime mortgage crisis, it is not limited to it. Instead, Table 2.6 shows that significant alpha is generated even if the subprime mortgage crisis is disregarded and we account for transactions costs.

2.5.3 Robustness

The methodology established in Section 2.3 involves a number of ad-hoc choices, most notably the relative size of the model selection and the forecasting period, the size S of the initial window in the model selection period in (2.2), and the size W of the rolling estimation window for the forecasting model in (2.3).

In the following, we therefore consider the robustness of the forecasting performance with respect to these modeling choices. The results of this exercise are reported in the Appendix in Table 2.7 that shows aggregated average forecast results analogous to the last row of Table 2.1, in Table 2.8 that reports trading returns repeating the analysis of Table

2.2 for $w = 0.15$ and 0bps transaction costs, and in Figure 2.7 that shows cumulative difference plots in analogy to the left plot of Figure 2.3.

First, we consider changing the length of the model selection period so that the window ends in 2002 or 2004 instead of 2003. In theory, a longer model selection period should be beneficial for the model selection procedure as more observations are available and consequently the results get more stable. However, a longer model selection period implies a shorter forecasting period so that less observations are available to evaluate the actual out-of-sample performance of the models.

For a model selection period from 1996 until the end of 2002, the procedure selected exactly the same model as reported in Section 2.3.2. Consequently, all results presented in Section 4 also hold if the model selection period is one year shorter. The hitrate and trading return change slightly, however, due to the additional year in the forecasting period.

Adding another year, i.e., performing model selection from 1996 until the end of 2004, results in a slightly different model. The variables oil return and the 12-day moving average of the y_{it} are now included in the model with all other variables being the same. As the top right graph in Figure 2.7 shows, this has only a slight impact on the forecasting performance. The curves are still predominantly upward sloped indicating superior performance of the forecasts. Furthermore, the hitrate of 52.00 percent reported in Table 2.7 is almost exactly the same as in the main analysis. The average trading return slightly decreases to 11.44 percent as the trading strategy performs above average in the year 2004, which is now excluded.

Second, we change the length S of the initial window in the model selection procedure from 500 to 250, respectively 750. In general, smaller values of S lead to a lower stability of the initial estimates, but there is also a larger number of pseudo out-of-sample observations available to select the variables.

The right graph in the second row of Figure 2.7 reveals that increasing this number to 750 has a moderate effect on the forecasts. The selected model stays the same with the exception that the variable VIX return is replaced by the change of the third PC (curvature). Consequently, graph, hitrate (51.93 percent), and average trading return (15.44 percent) are almost identical to the ones in the initial setup.

This, however, does not hold when decreasing the size of the initial window S to 250. The forecasts for this specification remain slightly better than the optimist benchmark, but between 2010 and 2017 a predominantly negative slope is observed. Nevertheless, as the trading strategy trades on the dispersion of probabilities and therefore does not invest in all forecasts, it still generates an average trading return of 13.42 percent. This is due to the fact that the hitrate of all stocks traded is at 54.08 percent which is only marginally smaller than in the main analysis.

Third, we changed the length W of the estimation window for out-of-sample forecasting in (2.3) from 1,000 to 750, respectively 1,250. As before, graphs and hitrates remain qualitatively similar despite these changes. For the average trading return, we observe a decline for the larger window size, which is again due to the smaller forecasting period excluding the year 2004.

Fourth, as argued in the introduction, returns and excess returns are essentially the same on a daily horizon since the magnitude of the risk-free rate is marginal. The analysis in this paper is therefore conducted directly for the log-returns. To judge the impact of this modeling choice, we repeat the analysis using excess returns. As can be seen from Table 2.7 and Table 2.8, this delivers results that are very similar to those presented in Section 2.4.

2.6 Conclusion

Daily stock returns are generally regarded as unpredictable. However, the results presented here show that the direction of daily stock returns is, to some extent, predictable. This predictability is shown to be statistically significant in an out-of-sample environment, of a magnitude that is economically meaningful so that it can be exploited by suitable trading strategies, and consistent over time.

The logistic regression model used to generate these predictions is not designed to address the properties of stock returns in an efficient manner and it is likely that a more suitable statistical approach can generate even better forecast results. Furthermore, the trading strategies proposed in Section 2.4.2 trade on the dispersion between the predicted probabilities for expected returns, but the cross section of the DJIA dataset considered is relatively small. It is therefore likely that the trading performance can be further improved by considering a larger asset universe.

While the degree of predictability is much lower than that found in previous studies in the machine learning literature, it is still clear that there is some form of nonlinear dependence in daily stock returns. These results may be unexpected from a theoretical perspective, but they are clearly in line with those of Linton and Whang (2007) and Han et al. (2016), who find evidence for directional predictability of daily returns based on nonparametric tests. We therefore conclude that directional forecasts are a promising field for future research on the predictability of stock returns.

Appendix

	Mean	SD	1%	25%	50%	75%	99%
Log-Realized Variance	-8.592	0.911	-10.476	-9.344	-8.708	-8.036	-6.370
High-Low Variance in Percent	0.033	0.163	0.001	0.006	0.013	0.029	0.309
Realized Skewness	0.033	1.009	-2.670	-0.544	0.019	0.603	2.758
S&P 500 Return in Percent	0.025	1.210	-3.412	-0.522	0.056	0.612	3.386
Realized Beta	0.800	0.396	-0.049	0.654	0.914	1.151	1.868
Log-Realized Variance S&P 500	-9.715	1.015	-11.784	-10.443	-9.794	-9.109	-6.888
Level VIX	21.562	8.632	10.680	15.160	20.300	25.240	50.930
VIX Return in Percent	-0.001	6.690	-16.040	-3.818	-0.411	3.308	20.197
Variance Premium	6.868	4.217	0.000	4.092	6.318	9.257	19.507
Oil Return in Percent	0.028	2.380	-6.305	-1.260	0.036	1.379	6.029
Level First PC (Level)	-1.301	6.389	-12.334	-6.492	-0.684	5.907	8.162
Level Second PC (Slope)	-0.138	1.120	-2.800	-0.798	-0.057	0.761	1.750
Level Third PC (Curvature)	0.004	0.310	-0.807	-0.197	0.023	0.224	0.574
Change First PC (Level)	0.003	0.171	-0.480	-0.082	0.000	0.087	0.460
Change Second PC (Slope)	0.000	0.090	-0.240	-0.043	0.001	0.046	0.222
Change Third PC (Curvature)	0.000	0.049	-0.122	-0.019	0.000	0.018	0.151
Stock Return in Percent	0.013	2.167	-5.905	-0.926	0.000	0.966	5.811
5-Day Moving Average Return in Percent	0.014	0.944	-2.683	-0.407	0.040	0.465	2.439
On-Balance Volume ($\times 10^{-4}$)	-0.426	3.146	-10.268	-1.262	-0.098	0.606	12.443
12-Day Binary Moving Average	0.497	0.140	0.167	0.417	0.500	0.583	0.833
Momentum Indicator	0.047	1.544	-4.559	-0.539	0.053	0.665	4.312
A/O Oscillator	0.515	0.538	-0.680	0.167	0.500	0.861	1.753
Rate-of-Change Indicator	1.006	0.077	0.791	0.969	1.007	1.044	1.211

Table 2.4: Summary statistics of the explanatory variables.
Mean corresponds to the mean of the variables, SD to the standard deviation and the remaining columns state the respective quantiles of the variables. All statistics are based on the whole period, i.e., 1996 until 2017.

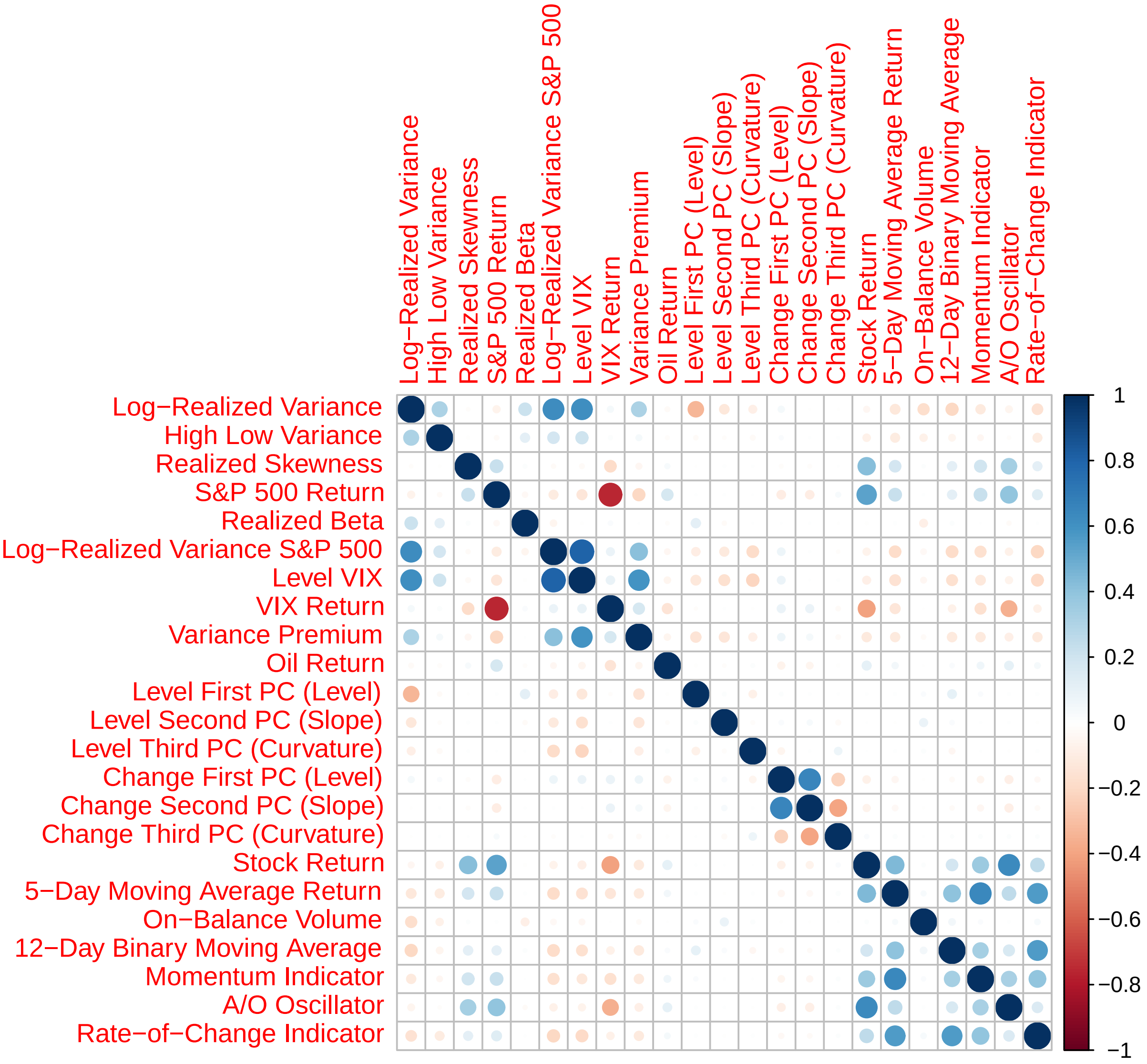

Figure 2.6: Correlation plot of the explanatory variables.
Size and color of the circles correspond to the degree and direction of the correlation. All correlations are based on the whole period, i.e., 1996 until 2017.

	Level	p-value
Intercept	0.146	0.000
Probability of Bull Market	-0.189	0.000

Table 2.5: Regression results.
The table shows the results of regressing the daily trading return on the probability of a bull market. The bull-market probability is measured by the filtered state probability from a Markov switching mean-variance model fitted to the S&P 500 returns. The Markov switching model has two regimes corresponding to a bull and a bear market.

	Mean	SD	SR	α (CAPM)	α (five factors)	Max TC
				$w = 0.15$		
0bps	11.72	12.10	0.86	13.22***	12.42***	8bps
5bps	3.33	12.10	0.17	4.73*	4.00	
10bps	-4.44	12.10	-0.46	-3.11	-3.78	
				$w = 0.17$		
0bps	10.60	11.33	0.82	11.68***	11.16***	9bps
5bps	4.38	11.32	0.28	5.43**	4.95*	
10bps	-1.50	11.32	-0.24	-0.47	-0.91	
				$w = 0.19$		
0bps	8.67	10.24	0.72	9.37***	8.85***	11bps
5bps	4.38	10.21	0.31	5.01**	4.53*	
10bps	0.25	10.21	-0.09	0.83	0.37	
				$w = 0.21$		
0bps	7.58	9.11	0.69	8.03***	7.68***	13bps
5bps	4.64	9.10	0.37	5.05**	4.72**	
10bps	1.78	9.08	0.06	2.14	1.84	

Table 2.6: Annualized performance measures for the trading strategy excluding the period of the subprime mortgage crisis.
In analogy to Table 2.2, the table shows performance measures for the trading strategy when excluding the period from August 2008 until March 2009 from the sample.

				DM test	
	HR	SE	SP	HM	Opt
Modelselection until 2002	51.77	53.61	49.85	49.91***	50.88**
Modelselection until 2004	52.00	54.07	49.86	50.10***	50.88**
Initial Model Selection Window 250	50.98	52.89	49.02	50.03***	50.86
Initial Model Selection Window 750	51.93	56.41	47.29	50.03***	50.86**
Forecasting Window 750	51.76	56.34	47.02	50.03***	50.86**
Forecasting Window 1,250	51.70	53.05	50.30	50.03***	50.86**
Excess Return Forecasting	51.85	54.85	48.76	50.00***	50.82**

Table 2.7: Aggregated forecasting results for a number of robustness checks. In analogy to the last row of Table 2.1, the table reports average hitrate, sensitivity, and specificity in percent for the robustness checks. Again, the right panel shows the hitrate in percent for the historical majority forecast and the optimist forecast and the asterisks indicate whether the logistic forecasts are significantly better than the respective benchmark according to a one-sided Diebold–Mariano test.

	Mean	SD	SR	α (CAPM)	α (five factors)	Max TC
Modelselection until 2002	13.71	12.43	1.00	14.66***	14.45***	11bps
Modelselection until 2004	11.44	13.30	0.76	12.44***	11.75***	10bps
Initial Model Selection Window 250	13.42	12.76	0.95	13.98***	13.49***	11bps
Initial Model Selection Window 750	15.44	13.18	1.07	15.98***	15.71***	12bps
Forecasting Window 750	18.09	13.22	1.26	18.97***	18.24***	13bps
Forecasting Window 1,250	10.33	12.21	0.74	11.31***	10.79***	11bps
Excess Return Forecasting	12.29	13.06	0.84	13.10***	12.22***	11bps

Table 2.8: Annualized performance measures for the trading strategy for a number of robustness checks. In analogy to Table 2.2, the table shows performance measures for the trading strategy for a number of robustness checks for $w = 0.15$ and 0bps transaction costs. All measures reported are annualized and all values, except for the Sharpe ratio, are given in percent.

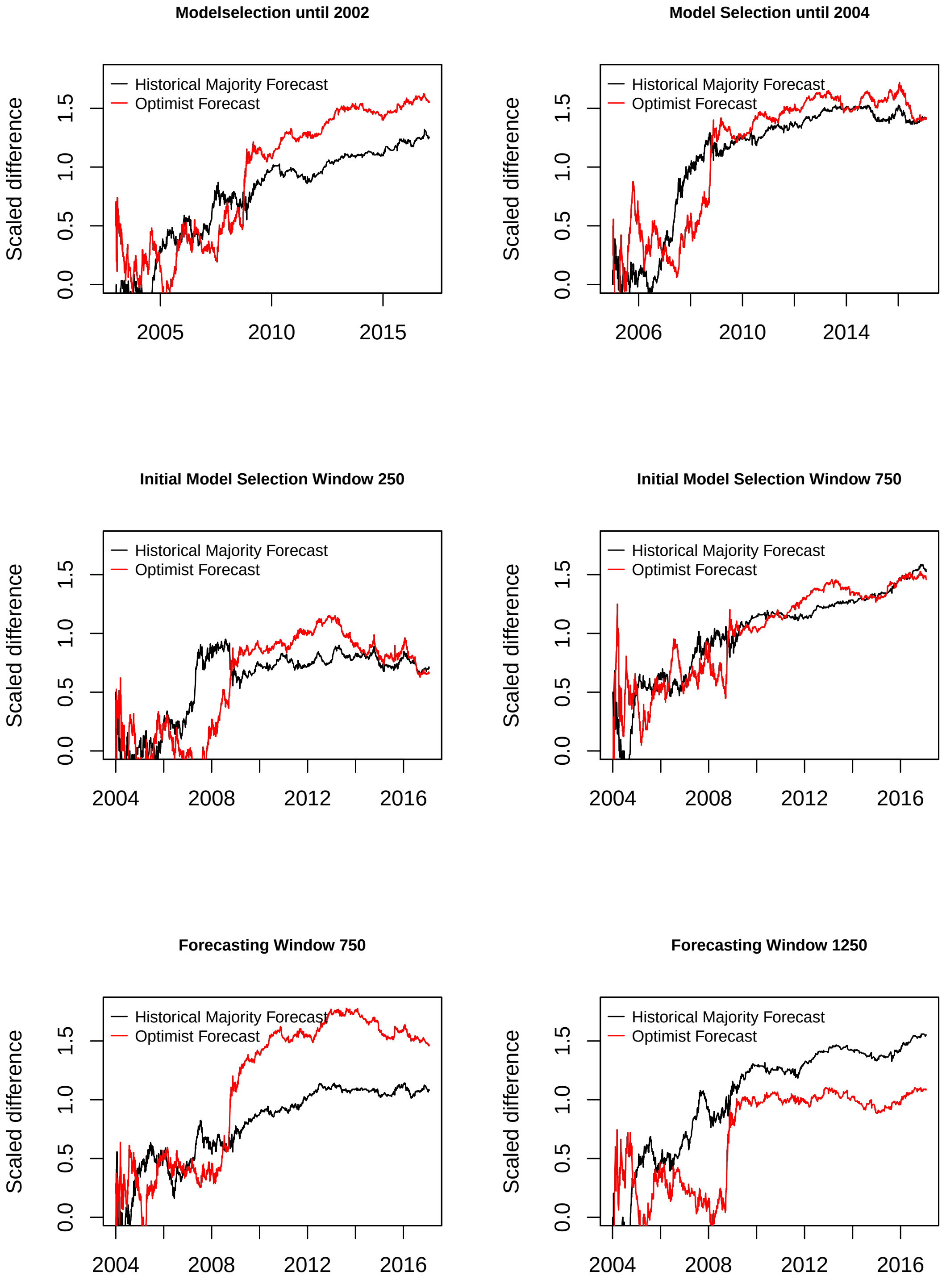

Figure 2.7: Cumulative difference plots for a number of robustness checks. The plots are in analogy to the left plot in Figure 2.3. The first row shows the effect of changing the length of the model selection window, the second row corresponds to changing the window size S in the model selection period, and the last row considers changes in the window size W of the forecasting period.

Chapter 3

Estimating the Volatility of Asset Pricing Factors

Co-authored with Christian Leschinski.

3.1 Introduction

There is a wide consensus that the cross section of asset returns is best described by factor models that proxy for economy wide risk factors. In addition to the established market, size, and value factors of Fama and French (1993), and the momentum factor of Carhart (1997), a plethora of anomalies has been uncovered in the literature that largely failed to attain the status of additional factors (Stambaugh and Yuan, 2016). Recently, Fama and French (2015), Hou et al. (2015), and Stambaugh and Yuan (2016) suggest investment, profitability, and mispricing factors that subsume a large proportion of these anomalies.

In a simplified form, these factors are constructed as follows. First, all stocks in the asset universe are sorted according to some firm characteristic. Second, two value weighted portfolios are formed from those stocks whose firm characteristics fall into the highest and lowest $x\%$-quantile. The factor return is then obtained as the return from buying one of these portfolios and selling the other.

For risk management and portfolio formation purposes, it is, however, not only the return but also the volatility of these factors that is of interest. Return volatility is a key variable for the pricing of options, speaks directly to the risk-return trade-off central to portfolio allocation, and even finds its way into government regulations.

For liquid individual assets, the unobservability of volatility has been alleviated through the increased availability of high-frequency data and the advent of realized variance. Given that returns of the asset can be observed frictionless in arbitrarily small time intervals, realized variance provides a consistent estimate of the quadratic variation of the asset return. For a review of these concepts cf. Andersen and Benzoni (2009).

While this approach is straightforward for individual assets, the calculation of realized variances for empirical asset pricing factors is challenging. This is because the COMPUSTAT and CRSP databases that are typically used to construct the factor returns do not provide high-frequency data. To calculate realized factor variances, it would therefore be necessary to match the stocks in these databases with those from a high-frequency data provider.

This is the approach considered by Ait-Sahalia et al. (2019). It is, however, not straightforward. High-frequency data is typically only available for the most liquid stocks that are traded regularly in short time intervals. The CRSP portfolios that are used to construct empirical factor models, on the other hand, contain much more illiquid stocks that are simply not traded often enough to calculate realized variances. Furthermore, high-frequency databases are not necessarily free of survivorship bias, and finally — even if these hindrances would not exist — the matching of databases typically constitutes a large effort and there tend to be non-negligible matching errors.

Practitioners or researchers that need to estimate factor volatilities therefore either rely on squared daily returns as, for example, in Moreira and Muir (2017), or estimate the underlying volatility process through a GARCH model as, for example, in He et al. (2015). Both approaches have major drawbacks. Squared returns provide an unbiased but inconsistent estimate of the true volatility and were the standard measure considered in the GARCH literature prior to the emergence of realized variance. It is, however, well known that squared returns are extremely noisy. Andersen and Bollerslev (1998a) show that, despite the high degree of persistence in stock return volatility, even the true model is only able to explain five to ten percent of the daily fluctuation in squared returns. Volatility estimates based on GARCH models, on the other hand, have a lower variance, but they are biased and inconsistent if the model is misspecified.

The main contribution of this paper is to propose an estimation method for factor volatility that is close in precision to realized variance. Our approach is applicable whenever the researcher has access to daily factor return series and some high-frequency database. The idea is to approximate the factor return using a linear combination of the returns in the database. In the first step, an appropriate linear combination is estimated using ridge regression. In the second step, the bias of the approximate factor is corrected, before the realized variance of this approximate factor is calculated and used as an estimate for the volatility of the actual factor.

The details of this procedure are discussed in Section 3.2. We demonstrate the validity of our approach in a Monte Carlo study in Section 3.3. The empirical validity and usefulness of this approach for the estimation and prediction of factor volatility is demonstrated in Section 3.4. First, we analyze the relationship between our estimate and the squared returns for the factors considered by Carhart (1997) and Fama and French (2015) and show that both are estimates of the same underlying volatility process. Second, we consider the example of the market factor where we can use the realized variance of the S&P 500 to evaluate the accuracy of volatility forecasts. Here, we find that using our measure improves forecasts of the factor volatility considerably compared to squared returns and GARCH-type models. Conclusions are discussed in Section 3.5.

3.2 Estimating Factor Volatility

If stock returns are driven by a given factor model, then it holds true that the return of each stock is a linear combination of the returns of these factors and an idiosyncratic error term. As our procedure is based on daily and high-frequency data, we assume that expected stock and factor returns are zero so that they do not contain risk premia. Let there be K factors and denote the return of factor $k = 1, \ldots, K$ at time t by f_{kt}. Then, the return of stock i at time t according to this model is given by

$$r_{it} = \sum_{k=1}^{K} \lambda_{ik} f_{kt} + \varepsilon_{it}, \tag{3.1}$$

where $\varepsilon_{it} \sim (0, \sigma_{\varepsilon}^2)$, λ_{ik} is the loading of the i-th stock on the k-th factor, and $i = 1, \ldots, N$. It is assumed that the ε_{it} have limited cross-sectional and serial dependence and that they are independent of all λ_{ik} and f_{kt}.

Conversely, it follows that the return of each factor can be approximated by a linear combination of the stock returns. For suitable β_{ik}, we therefore have

$$f_{kt} = \sum_{i=1}^{N} \beta_{ik} r_{it} + \nu_{kt}, \tag{3.2}$$

where ν_{kt} represents the approximation error, which can be expected to be small for large N since the idiosyncratic errors ε_{it} in (3.1) average out.

The rationale behind this approach becomes clear if we rewrite model (3.1) for a vector of N stocks. With $R_t = (r_{1t}, \ldots, r_{Nt})'$, $F_t = (f_{1t}, \ldots, f_{Kt})'$, $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{iK})'$, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, and $\Lambda = (\lambda_1, \ldots, \lambda_N)'$, we obtain $R_t = \Lambda F_t + \varepsilon_t$. If Λ was known (and $\Lambda'\Lambda$ invertible), we could estimate F_t by $(\Lambda'\Lambda)^{-1}\Lambda' R_t = F_t + (\Lambda'\Lambda)^{-1}\Lambda'\varepsilon_t = F_t + \varepsilon_t^*$.

Since Λ is $N \times K$ and ε_t is $N \times 1$, ε_t^* is $K \times 1$. Therefore, every element of ε_t^* is a weighted average of the innovation terms $\varepsilon_{1t}, \ldots, \varepsilon_{Nt}$ and the vector ε_t^* converges to zero by a suitable law of large numbers (cf. Stock and Watson (2011) for a related discussion of cross-sectional averaging and statistical factor models).

The coefficient vector $\beta_k = (\beta_{1k}, \ldots, \beta_{Nk})'$ in (3.2) corresponds to the k-th row of the matrix $(\Lambda'\Lambda)^{-1}\Lambda'$. Since the returns f_{kt} of the observed factors are readily available, the problem in estimating β_k is that it is N-dimensional and therefore potentially very variable if the time dimension T is not large enough. In fact, it is likely that $N > T$ in empirical applications so that standard estimation methods cannot be applied.

We therefore resort to regularization and estimate β_k using ridge regression. The estimator is given by

$$\hat{\beta}_k = \underset{\beta_{1k},\ldots,\beta_{Nk}}{\arg\min} \left\{ \sum_{t=1}^{T} \left(f_{kt} - \sum_{i=1}^{N} \beta_{ik} r_{it} \right)^2 + \gamma \sum_{i=1}^{N} \beta_{ik}^2 \right\}, \tag{3.3}$$

with $\gamma > 0$. This is a least squares estimator with an additional penalty term that shrinks the coefficients towards zero. The size of the penalty term depends on the parameter γ that can be selected using cross validation. Here, we select γ so that the out-of-sample mean squared error between the observed and estimated factor is minimized in ten-fold cross-validation. While the introduction of the penalty term introduces some bias, the rationale behind ridge regression is that for suitable γ, the reduction in variance outweighs the size of the bias so that $\hat{\beta}_k$ is more accurate than the OLS estimator in terms of the mean squared error. Moreover, γ lowers the effective degrees of freedom so that $N > T$ is permitted if γ is sufficiently large.

Holding the weights $\hat{\beta}_k$ in the linear combination constant then allows to obtain approximate high-frequency factor returns. Denote the m-th of M intra-day returns of stock i on day t by $r_{it}^{(m)}$, then the m-th intra-day return of factor k on day t is given by

$$\hat{f}_{kt}^{(m)} = \sum_{i=1}^{N} \hat{\beta}_{ik} r_{it}^{(m)}. \tag{3.4}$$

This allows for a RV-type estimation of the daily factor volatility V_{kt}.

The approach is subject to two sources of bias. On the one hand, regularization shrinks the coefficients towards zero so that the volatility is underestimated. On the other hand, the variance of the coefficient estimates can be translated to the volatility estimate, which causes a positive bias. To correct for these biases, we include an auxiliary regression step. We calculate the predicted daily values of the factors $\hat{f}_{kt} = \hat{\beta}_k' R_t$ based on (3.2) and then use ordinary least squares to estimate $f_{kt} = \delta \hat{f}_{kt} + \eta_{kt}$, where η_{kt} is assumed to be a mean-zero martingale difference sequence. Since $V_{kt} = \delta^2 Var(\hat{f}_{kt}) + \sigma_{\eta_k}^2$, we can use the estimated coefficient $\hat{\delta}$ and the residual variance estimate $\hat{\sigma}_{\eta_k}^2$ to correct for the bias.

Consequently, an unbiased estimator for V_{kt} analogous to realized variance is given by

$$\hat{V}_{kt} = \hat{\delta}^2 \sum_{m=1}^{M} \left(\hat{f}_{kt}^{(m)} \right)^2 + \hat{\sigma}_{\eta_k}^2. \tag{3.5}$$

We refer to $\hat{V}_{kt}$ as the Ridge-RV estimator. Note that when speaking of volatility, some researchers refer to the variance and others to the standard deviation of asset returns. By defining $\hat{V}_{kt}$ as in (3.5), we implicitly follow Andersen and Benzoni (2009) and Aït-Sahalia et al. (2011) and refer to volatility as the variance of asset returns. Performing the analyses in Section 3.3 and 3.4 for $\sqrt{\hat{V}_{kt}}$ leads to qualitatively similar results.

To summarize, our method proceeds as follows.

1) Perform a ridge regression of the daily factor return f_{kt} on the daily returns of the N stocks to obtain $\hat{\beta}_k$ and $\hat{f}_{kt} = \hat{\beta}_k' R_t$ from (3.2).

2) Obtain estimates $\hat{f}_{kt}^{(m)}$ of the intra-day returns of the factors using (3.4).

3) Estimate the auxiliary regression model $f_{kt} = \delta \hat{f}_{kt} + \eta_{kt}$.

4) Estimate the volatility of the factor from the estimated intra-day returns $\hat{f}_{kt}^{(m)}$ and the estimated coefficients $\hat{\delta}$ and $\hat{\sigma}^2_{\eta_k}$ using the Ridge-RV estimator in (3.5).

It should be noted that it is not necessary to have high-frequency returns of all stocks that are part of the original portfolios used to construct the asset pricing factors. As long as the assumed empirical asset pricing model is a linear factor model and it is a good approximation of the true underlying process, a large number of stocks should have non-zero loadings on the factor. For example, the return of the size factor can be estimated from large stocks that have negative loadings on the size factor. High-frequency observations of small illiquid stocks are not required.

3.3 Monte Carlo Simulation

To demonstrate the validity of the Ridge-RV estimator, we conduct a simulation study that is tailored to resemble the setup in the empirical applications in Section 3.4.

It is well known that stock volatilities tend to have long memory and are well described by fractionally integrated processes (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996; Ding and Granger, 1996). A fractionally integrated process X_t is given by

$$(1 - B)^d X_t = v_t, \tag{3.6}$$

where B defined by $BX_t = X_{t-1}$ is the lag operator, v_t is a short-memory process, and $-1/2 < d \leq 1$. The fractional difference operator $(1-B)^d$ is defined in terms of generalized binomial coefficients. For details cf. the original contributions of Granger and Joyeux (1980) or Hosking (1981). A process that fulfills (3.6), such as the well known ARFIMA model, is referred to as $I(d)$. Standard short-memory processes are included for $d = 0$ and unit root processes are obtained for $d = 1$.

To resemble these long-memory patterns in the daily volatility V_{kt} of the K factors, we use the long-memory stochastic volatility framework of Breidt et al. (1998) and simulate T daily observations (with 250 burn-in observations) for each factor using

$$V_{kt} = \exp(X_{kt}) \text{ with } X_{kt} \sim \text{ARFIMA(0,d,0)}.$$

The log-volatilities therefore follow a fractionally integrated model. Applying the exponential function guarantees that all volatilities are positive. The V_{kt} obtained this way are used as the true daily volatilities.

Based on these, we subsequently draw M intra-day factor returns $f_{kt}^{(m)} \overset{iid}{\sim} N\left(0, V_{kt}/M\right)$ for each day and factor. The daily factor returns are obtained as $\sum_{m=1}^{M} f_{kt}^{(m)}$ so that they have volatility V_{kt}. Using the intra-day factor returns, we can simulate intra-day returns of N stocks. In analogy to Equation (3.1), the m-th return of stock i at day t evolves as

$$r_{it}^{(m)} = \sum_{k=1}^{K} \lambda_{ik} f_{kt}^{(m)} + \varepsilon_{it}^{(m)},$$

where $\varepsilon_{it}^{(m)} \overset{iid}{\sim} N\left(0, \sigma_{\varepsilon}^{2}/M\right)$ is a noise component. As for the daily factor returns, daily stock returns are obtained as the sum over the M intra-day returns so that $r_{it} = \sum_{m=1}^{M} r_{it}^{(m)}$.

All parameters are chosen such that the situation in our empirical application in Section 3.4 is replicated as closely as possible. This means we consider $K = 6$ factors whose correlation matrix matches the correlation matrix of the six factors considered there, we chose the memory parameter d to be 0.6 for all factors as the literature suggests the memory parameter of return volatility to be in this region (Wenger et al., 2018), we simulate $M = 78$ intra-day returns, which corresponds to 5-minute stock data, the factor loadings λ_{ik} used for the simulation of stock returns are given by regression estimates of the factor loadings of $N = 500$ randomly chosen stocks that were in the S&P 500 at some point in the last 20 years, and σ_{ε}^{2} evolves as the residual variance of this regression. Moreover, we set $T = 750$.

Based on this simulated data, we then apply the procedure described in Section 3.2 based on Equations (3.3) to (3.5). Using the intra-day factor returns $f_{kt}^{(m)}$, we can also compute the actual realized variance. As a comparison, we further fit a GARCH(1,1) (Bollerslev, 1986) and a FIGARCH(1,d,1) (Baillie et al., 1996) model and we consider the squared daily factor returns as an estimate of V_{kt} too.

Results from 1,000 Monte Carlo repetitions can be found in the upper panel of Table 3.1 that shows the bias compared to the true volatility and the RMSE of all the procedures considered. The results are qualitatively similar for all factors and indicate realized variance and Ridge-RV to be the best estimators. They are both unbiased and exhibit a similar degree of variance resulting in comparable RMSEs.

The squared returns are unbiased, but their large variance leads to an RMSE that is several times larger than that of the Ridge-RV estimator. The GARCH(1,1) model cannot remedy the noise problem and is biased since it does not allow for long memory but the data generating process is $I(d)$. The FIGARCH(1,d,1) model achieves an improvement since it allows for long memory, but it is still too noisy resulting in a RMSE six times that of the Ridge-RV estimator.

		F1	F2	F3	F4	F5	F6
ARFIMA(0,d,0)							
Realized Variance	RMSE	1.042	1.195	1.088	1.278	1.612	1.238
	Bias	0.002	-0.005	0.007	0.000	-0.011	-0.006
Ridge-RV	RMSE	1.046	1.202	1.082	1.322	1.611	1.235
	Bias	0.002	-0.011	0.001	0.016	-0.016	-0.014
Squared Return	RMSE	8.383	9.787	8.708	12.188	12.486	10.163
	Bias	0.000	0.014	-0.041	0.073	0.103	-0.010
GARCH(1,1)	RMSE	8.337	10.401	8.717	13.266	14.237	10.738
	Bias	0.998	1.234	1.042	1.440	1.708	1.234
FIGARCH(1,d,1)	RMSE	6.257	7.424	6.460	8.207	9.079	7.811
	Bias	0.020	0.019	-0.039	0.053	0.113	0.002
ARMA(1,1)							
Realized Variance	RMSE	0.440	0.440	0.435	0.440	0.439	0.439
	Bias	0.001	-0.000	-0.001	0.000	0.001	0.001
Ridge-RV	RMSE	0.435	0.463	0.440	0.477	0.452	0.450
	Bias	0.005	-0.022	-0.021	-0.013	0.021	-0.020
Squared Return	RMSE	3.792	3.838	3.735	3.760	3.709	3.752
	Bias	0.002	0.012	-0.005	0.002	-0.008	-0.002
GARCH(1,1)	RMSE	2.215	2.212	2.201	2.225	2.212	2.216
	Bias	0.007	0.016	-0.001	0.006	-0.005	0.000
FIGARCH(1,d,1)	RMSE	2.221	2.224	2.207	2.232	2.217	2.220
	Bias	0.010	0.021	0.002	0.011	-0.002	0.004

Table 3.1: Simulation results.
Reported are RMSE $\times 10^3$ and Bias $\times 10^3$ for different volatility estimation approaches. The true volatility processes of the six factors (F1, F2,..) evolve as $V_{kt} = \exp(X_{kt})$ with $X_{kt} \sim$ ARFIMA(0,d,0) respectively $X_{kt} \sim$ARMA(1,1). Moreover, the correlation matrix of the simulated processes matches the correlation matrix of the six factors considered in the empirical application.

As a robustness check, we repeat the same simulation but with a stochastic volatility ARMA(1,1) process since it is still often assumed that short-memory GARCH-type models allow for an accurate description of the volatility process. The parameter values used for the simulation are obtained via estimation of an ARMA(1,1) for the respective Ridge-RV series.

The results are shown in the lower panel of Table 3.1. It can be seen that Ridge-RV still performs comparable to the infeasible RV estimate and that it is considerably better than the competitors. In situations where the intra-day returns of a portfolio cannot be observed, the Ridge-RV estimator is therefore the best available choice.

3.4 Empirical Analysis

In the following, we consider the market (MKT), size (SMB), and value (HML) factors included in the three-factor model of Fama and French (1993), the profitability (RMW) and investment (CMA) factors added in the five-factor model of Fama and French (2015), and the momentum factor (MOM) included by Carhart (1997). These factors are commonly used in the asset pricing literature and their validity is widely accepted. Daily returns of these factors are freely available on the homepage of Kenneth French.

In addition to the daily factor returns we require daily stock returns r_{it} and high-frequency returns $r_{it}^{(m)}$ for the estimation of (3.2) and the calculation of approximate high-frequency factor returns $f_{kt}^{(m)}$ from (3.4).

Since it is common to calculate realized variances from 5-minute returns, we extract 5-minute prices of all stocks that were part of the S&P 500 at some point between 1996 and 2017 from the Thomson Reuters Tick History database. This results in a total amount of 1,367 stocks that are considered. Since high-frequency data is often subject to minor recording mistakes, it is common practice to apply some form of data cleaning. Here, we adopt the approach of Barndorff-Nielsen et al. (2009), which comprises, among other things, the removal of observations with negative stock prices and abnormal high or low entries in comparison to other observations on the same day.

Due to the long time span, it cannot be expected that the coefficients β_{ik} stay constant over time. The loading of individual stocks on factors can change as competitors are acquired that have a different exposure to market risk, small firms grow into large firms, and growth stocks turn into value stocks as companies mature. We therefore conduct the estimation of the coefficient vector $\hat{\beta}_k$ according to (3.3) in a rolling window of size W. For the factors MKT, SMB, HML, and RMW, which are based on firm characteristics that are relatively stable over time, we set $W = 750$. The factors MOM and CMA that are based on more dynamic features are estimated in a window of size $W = 125$. Results for other values of W are qualitatively similar and available upon request.

To demonstrate the empirical validity of our factor volatility estimates, the next section shows a number of model diagnostics. Afterwards, Section 3.4.2 demonstrates that volatility forecasts can be improved by using our measure.

3.4.1 In-Sample Volatility Estimates and Model Diagnostics

When trying to evaluate the performance of the Ridge-RV estimator, we face the problem that the true volatility process is unobserved and realized variances are not available for the factors. Only squared returns can be observed. We therefore consider a number of model diagnostics that demonstrate the satisfactory performance of our procedure before turning to the forecasts in Section 3.4.2.

Figure 3.1 plots the natural logarithms of squared returns and Ridge-RV for the six factors over time. Two main observations can be made. First, our measure is comoving with the squared factor returns, which is a first indication that both measures estimate the same underlying volatility process. Larger values of the squared factor returns are

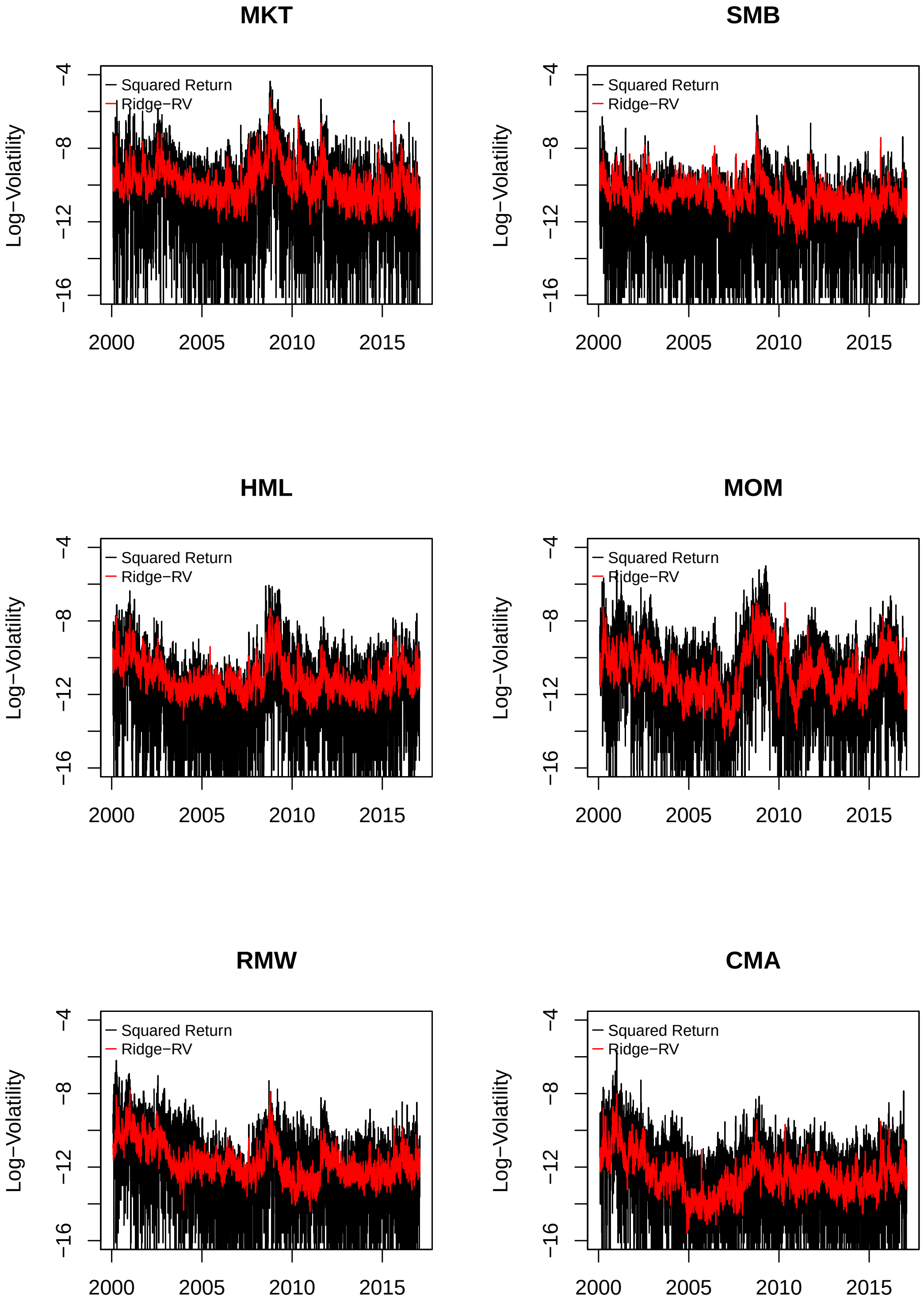

Figure 3.1: Graphical comparison of Ridge-RV and squared returns. The figure displays time-series plots of the natural logarithms of Ridge-RV and squared returns for market (MKT), size (SMB), value (HML), momentum (MOM), profitability (RMW), and investment (CMA) factor.

	MKT	SMB	HML	MOM	RMW	CWA
R^2	98.63	88.68	88.78	90.10	88.42	89.71

Table 3.2: Auxiliary regression results.
Reported is the coefficient of determination R^2 in percent for the auxiliary regression model $f_{kt} = \delta \hat{f}_{kt} + \eta_{kt}$.

associated with larger values of the Ridge-RV and vice versa. This holds for all factors and all time periods. Second, the Ridge-RV appears to be far less perturbed than the squared returns.

The Ridge-RV estimate is based on the approximation of the factor of interest by a linear combination of stock returns. If this approximation in (3.2) is sufficiently accurate, so are those in (3.4) and (3.5). A first indication of the quality of the estimate can therefore be obtained from the coefficients of determination R^2 in the auxiliary regression of f_{kt} on $\delta \hat{f}_{kt}$. Table 3.2 shows that the measure is above 88 percent for all of the six considered factors indicating a high precision of the estimates.

Since squared returns and Ridge-RV are both estimates of the same unobserved volatility process, they can both be understood as differently perturbed versions of it. An approach to test the validity of the Ridge-RV estimator in this empirical setup is therefore to test for fractional cointegration between the squared returns and $\hat{V}_{kt}$. Fractional cointegration is a natural generalization of cointegration to fractionally integrated series. Two time series X_t and Y_t are said to be fractionally cointegrated if both are $I(d)$ and there exists a linear combination $X_t - \alpha - \beta Y_t = u_t$ so that u_t is $I(d-b)$ for some $0 < b \leq d$. As in standard cointegration, both series must be highly persistent and they are (fractionally) cointegrated if a linear combination of them has reduced persistence. The extension lies in the fact that the reduction of persistence does not have to be from $I(1)$ to $I(0)$ but can be from $I(d)$ to $I(d-b)$.

When modeling volatility time series it is common practice to work with the natural logarithm of the series since it is better approximated by the normal distribution (Andersen et al., 2001). If $\log V_{kt}$ denotes the true volatility process, then $\log f_{kt}^2 = \log V_{kt} + \omega_{kt}$ and $\log \hat{V}_{kt} = \log V_{kt} + \eta_{kt}$, where ω_{kt} and η_{kt} are the respective estimation errors. Therefore, if $\log V_{kt}$ is $I(d)$, then $\hat{V}_{kt}$ can only be a reasonable estimator of $\log V_{kt}$, if it is fractionally cointegrated with $\log f_{kt}^2$ so that $\log \hat{V}_{kt} - \log f_{kt}^2 = \eta_{kt} - \omega_{kt}$ is $I(d-b)$.

To formally test the hypothesis of fractional cointegration between $\log \hat{V}_{kt}$ and $\log f_{kt}^2$, we apply the tests of Chen and Hurvich (2006) and Souza et al. (2018) for the null hypothesis of no fractional cointegration. Under the alternative a fractional cointegration relationship exists.

Table 3.3 reports the results of the tests. As expected from Figure 3.1, the test by Chen and Hurvich (2006) rejects the null of no fractional cointegration for all factors and the test by Souza et al. (2018) rejects the null for all factors, except for the size

	MKT	SMB	HML	MOM	RMW	CMA	
CH	4.438	2.542	3.673	4.923	1.897	4.780	(1.697)
SRF	3.807	1.381	3.020	3.263	2.751	3.770	(1.960)

Table 3.3: Fractional cointegration test results. Reported are test statistics and critical values for the tests by Chen and Hurvich (2006) (CH) and Souza et al. (2018) (SRF). Here, the null of no fractional cointegration between log-squared returns and log-Ridge-RVs is tested against the alternative of fractional cointegration. The values in brackets are critical values at the five percent level.

factor. Therefore, we can conclude that squared returns and Ridge-RV are fractionally cointegrated.

All of the statistics presented so far show that our Ridge-RV estimator works well. However, as discussed above, the evidence provided is indirect since the actual volatility process is unobserved. For the market factor, however, we can conduct one experiment that provides insight into the actual accuracy of the Ridge-RV estimate. Even though we do not observe realized variances for the market factor, it is well known that the value weighted CRSP return, which is generally regarded as the best available market proxy, is highly correlated with the return of the S&P 500. The correlation coefficient is about 99 percent meaning that the direction of the variation and its scaling over time is essentially the same. For the S&P 500, it is possible to obtain intra-day prices so that we can calculate realized variances. Consequently, we can compare our estimate of market factor volatility with the realized variance of the S&P 500. As Andersen and Benzoni (2009) stress, the realized variance is the natural ex-post measure of the underlying volatility process to consider.

Figure 3.2 shows that the two measures are close to identical. In fact, they have a correlation of 91.4 percent, are fractionally cointegrated, and regressing our volatility estimate on the realized variance yields a slope of 0.99 that is not significantly different from 1.

We therefore conclude that our estimate is appropriate for describing the volatility of the market factor. Even though the results in Tables 3.2 and 3.3 indicate that the procedure works slightly better for the market factor than for the other factors, the degree of precision obtained for the market implies that the Ridge-RV should still be a good estimate for the volatility of the other factors.

It should be noted, however, that the procedure is based on the assumption that the factors under consideration are actually relevant for the cross section of stock returns. This may be an issue if one wishes to apply the procedure to any of the many weak factors discussed in the literature.

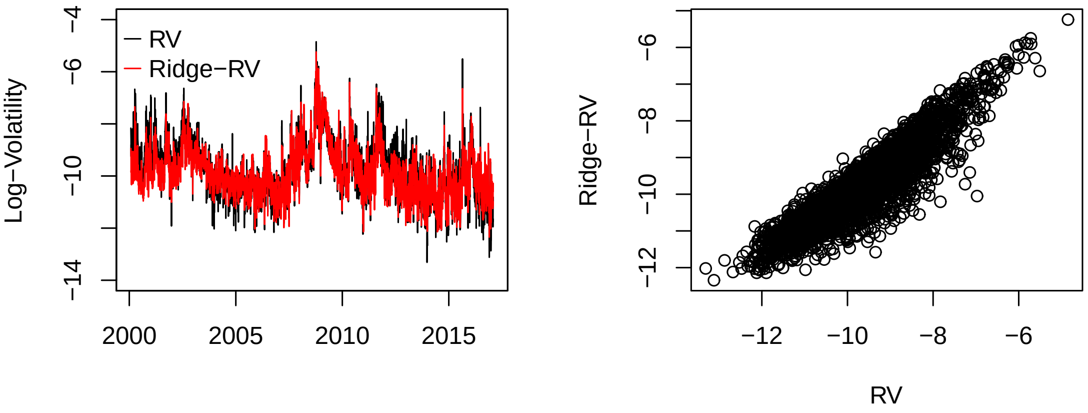

Figure 3.2: Graphical comparison of Ridge-RV and true RV. Both plots display the Ridge-RV estimate of market factor volatility and the true market factor volatility approximated by the realized variance of the S&P 500. While the left plot shows the two measures over time, the right plot displays a scatter plot.

3.4.2 Out-of-Sample Forecasts of Market Factor Volatility

For portfolio allocation and risk management purposes, accurate forecasts are needed in addition to ex-post and on-line estimates of the factor volatility. In this section, we therefore compare the performance of forecasts using squared returns and GARCH-type models with those using Ridge-RV.

When trying to evaluate these forecasts, we again face the problem that the true factor volatility is unobserved. As shown by Andersen and Bollerslev (1998a), considering squared returns as a proxy for the true factor volatility when evaluating volatility forecasts is not suitable since the tremendous amount of noise in the return generating process inevitably causes a poor performance of the forecasting models. On the other hand, it seems tautological to show superior performance of our Ridge-RV measure when considering it as the true factor volatility. We therefore proceed as in the previous section and conduct a forecast comparison for the volatility of the market factor, where we can use realized variances of the S&P 500 to proxy for the true factor volatility. This makes for a fair comparison since both types of models (Ridge-RV and models based on squared returns) do not use the realized variance of the S&P 500 in any way.

The Ridge-RV is predicted using the HAR model of Corsi (2009). We refer to this forecast as the HAR-Ridge-RV model. As a benchmark, we also consider the standard HAR-RV model, which is possible for the market but not for the other factors. It can thus be interpreted as the "infeasible" model that we try to approximate when predicting factors such as SMB, HML, or others. As feasible benchmark models, we include a

		1-Step			5-Step			22-Step	
	RMSE	QLIKE	R^2	RMSE	QLIKE	R^2	RMSE	QLIKE	R^2
GARCH(1,1)	0.211	0.298	0.492	0.171	0.238	0.596	0.181	0.259	0.513
FIGARCH(1,d,1)	0.222	0.299	0.444	0.168	0.230	0.606	0.175	0.250	0.535
HAR-RV	0.175	0.215	0.603	0.120	0.195	0.745	0.113	0.248	0.710
HAR-Ridge-RV	0.175	0.220	0.608	0.123	0.196	0.731	0.120	0.240	0.664

Table 3.4: Forecast results.
Reported are RMSE $\times 10^3$, QLIKE, and R^2 from Mincer-Zarnowitz regressions for the competing models and different forecast horizons. GARCH and FIGARCH use squared returns to forecast the market factor volatility, HAR-RV uses the true volatility approximated by the realized variance of the S&P 500, and HAR-Ridge-RV uses the Ridge-RV estimate.

GARCH(1,1) and due to the long range dependence in factor volatility, we also use a FIGARCH(1,d,1) model fitted to the squared returns. All estimations are carried out in a rolling window of 750 observations and for multistep-ahead forecasts we predict the mean volatility over the multiperiod horizon.

For the evaluation of the forecasts, we consider the RMSE and the QLIKE loss function, since Patton (2011) shows that these are the only commonly used measures that preserve the true ordering of the forecasts if they are evaluated on a perturbed volatility proxy. Furthermore, we report the coefficient of determination R^2 from Mincer-Zarnowitz (Mincer and Zarnowitz, 1969) regressions given by

$$\frac{1}{h}\sum_{j=1}^{h} RV_{t+j} = b_0 + b_1 \frac{1}{h}\sum_{j=1}^{h} \hat{V}_{t+j} + u_{kt}.$$

Here, RV_{t+j} is the observed volatility approximated by the realized variance of the S&P 500, $\hat{V}_{t+j}$ is the predicted volatility based on all information available in t, h is the forecast horizon, and u_{kt} is an error term. Consequently, larger values of R^2 imply that the forecasts are performing better in predicting the true volatility.

Table 3.4 shows the results of this forecasting exercise for 1-step, 5-step, and 22-step forecasts. It can be seen that for all forecasting horizons and for all evaluation measures, the HAR-Ridge-RV model performs better than all of the models based on squared daily returns. For 1-step forecasts, for example, the RMSE $\times 10^3$ of the HAR-Ridge-RV model is 0.175, QLIKE is 0.220, and the R^2 is 0.608, while for the GARCH(1,1) model, which is the best model using squared returns, the RMSE $\times 10^3$ is 0.211, QLIKE is 0.298, and the R^2 is 0.492. Due to the averaging, the forecasting performance of the models becomes slightly better on longer horizons. The ranking of the models, however, stays the same.

When comparing the forecasts based on our volatility estimate with the HAR-RV forecasts based on the realized variance of the S&P 500, it can be seen that the two models deliver qualitatively similar results.

	1-Step	5-Step	22-Step
GARCH(1,1)	3.632***	2.526**	1.654*
FIGARCH(1,d,1)	3.840***	2.596***	1.680*
HAR-RV	-0.013	-0.260	-0.538

Table 3.5: Diebold–Mariano test statistics.
Reported are the test statistics of the modified Diebold–Mariano test (Harvey et al., 1997) when testing for equal accuracy of the forecasts made by the HAR-Ridge-RV model and the competing models. Positive statistics imply that the forecasts of the HAR-Ridge-RV model are better and *, **, and *** indicate that this difference is significant at the ten percent, five percent, and one percent level, respectively.

Both of these findings are confirmed by results of modified Diebold–Mariano tests (Harvey et al., 1997) reported in Table 3.5. Here, we report the test statistics for equal accuracy of the forecasts by the HAR-Ridge-RV model and the forecasts by the competing models. Positive statistics imply that the forecasts of the HAR-Ridge-RV model are more accurate. It can be seen that for all considered forecast horizons the forecasts of the HAR-Ridge-RV model are significantly better than those of the GARCH(1,1) and FIGARCH(1,d,1) model at the ten percent level. In contrast, when comparing to the HAR-RV forecasts, the null hypothesis of equal forecast accuracy cannot be rejected at any commonly considered significance level.

Consequently, forecasts based on Ridge-RV achieve their objective to approximate those that are obtained if realized variances are available and they significantly outperform forecasts of models that use squared returns. For factors other than the market, where realized variances are not available, they can therefore be expected to provide results that are far better than standard approaches.

3.5 Conclusion

Although the volatility of economy wide risk factors such as the size and value factors of Fama and French (1993) are of importance for risk management and portfolio allocation purposes, the development of methods for their estimation has lagged behind that for liquid individual assets or indices, where intra-day returns are available.

The Ridge-RV approach suggested in this paper circumvents the lack of high-frequency data for factor returns and provides a volatility measure that is closely related to realized variance. This is achieved by approximating the daily factor returns by a linear combination of the returns of stocks for which intra-day returns are available. Holding the weights in the linear combination constant then allows to obtain approximate high-frequency factor returns that are the basis for the estimation of the factor volatility. Due to the large number of parameters in the linear combination that have to be estimated, it is necessary to apply a regularized estimation method such as ridge regression. This introduces a bias which is corrected by including an auxiliary regression step.

The subsequent applications to the market, size, value, momentum, investment, and profitability factors demonstrate that the proposed measure performs well in practice and outperforms competing approaches such as GARCH-type models. We therefore find that adopting the proposed approach has the potential for significant improvements in asset allocation decisions and risk management.

Chapter 4

The Bias of Realized Variance

Co-authored with Christian Leschinski.

4.1 Introduction and Main Finding

Volatility is at the heart of everything from risk management to derivative pricing and asset management. While estimates of the unobserved volatility process were originally obtained using GARCH and stochastic volatility models, today high-frequency data has become widely available and realized variance (RV) has been adopted as the standard measure. Due to its nature as a nonparametric estimate that is consistent for the quadratic variation in continuous price processes that behave as semimartingales, RV is often even treated as a direct observation of the underlying volatility process. This drastic improvements in the quality of volatility estimates has led to major advances in volatility forecasting and risk management.

Previous contributions on the shortcomings of RV have mostly focused on the effect of market microstructure noise and violations of the assumption that the price process can be observed frictionless at arbitrarily small time intervals (e.g., Zhou, 1996; Hansen and Lunde, 2006; Bandi and Russell, 2008).

Here, we focus on the semimartingale assumption and show that RV is a biased estimator for the variance of stock index returns on daily and longer horizons. The bias is negative so that the stock market risk is systematically underestimated. This effect is demonstrated for a wide range of international stock market indices and the average magnitude of the bias is 14 percent. The RV of the S&P 500, for example, underestimates the mean level of daily return variance by 15 percent.

We further provide a detailed investigation of the bias, which reveals that it is caused by dependencies between aggregates of returns within a trading day as recently documented by Gao et al. (2018). This invalidates not only the assumptions of the standard RV estimator but also the assumptions of any alternative high-frequency estimator suggested in the literature so far, such as the BV estimator (Barndorff-Nielsen and Shephard, 2004b), the medRV estimator (Andersen et al., 2012), and sub-sampled RV estimators. Our analysis shows that even realized kernel variance estimators (Barndorff-Nielsen et al.,

2008) are biased, although these are specifically designed to be robust against serial dependencies in the returns.

To make our point, we consider the following jump-diffusion model, which is customarily used in the literature to model the log-price process $p(\tau)$ of a financial asset,

$$dp(\tau) = \mu(\tau)d\tau + \sigma(\tau)dB(\tau) + \xi(\tau)dq(\tau). \tag{4.1}$$

Here, $\sigma(\tau)$ is the instantaneous or spot volatility, strictly positive and (almost) surely square integrable. Furthermore, $q(\tau)$ is a Poisson process uncorrelated with $B(\tau)$ and governed by the jump intensity $\lambda(\tau)$ so that $P(dq(\tau) = 1) = \lambda(\tau)d\tau$, which implies a finite number of jumps in the price path per time period. The scaling factor $\xi(\tau)$ denotes the magnitude of the jump in the return process if a jump occurs at time τ.

For the sake of the argument presented here, we will assume that $\mu(\tau) = 0$ for all τ. This means that the equity premium is zero, which is a reasonable assumption on short time horizons such as days because it is so small. Nevertheless, this is purely for expositional purposes and the arguments could easily be extended to allow for $\mu(\tau) \neq 0$.

Denote the continuously compounded return at day t by $r_t = p(t) - p(t-1)$, for $t = 1, \ldots, T$. Since $\mu(\tau) = 0$, we have

$$r_t = \int_{t-1}^{t} dp(\tau)d\tau = \int_{t-1}^{t} \sigma(\tau)dB(\tau) + \sum_{t-1 \leq \tau \leq t} J(\tau).$$

$$\text{Therefore,} \quad E[r_t^2] = Var[r_t] = IV_t + E\left[\sum_{t-1 \leq \tau \leq t} J^2(\tau)\right], \tag{4.2}$$

$$\text{where} \quad IV_t = \int_{t-1}^{t} \sigma^2(\tau)d\tau$$

and $J(\tau) = \xi(\tau)dq(\tau)$ is non-zero only if there is a jump at time τ. This is due to the assumed independence between the continuous components and jump components, the independence of the increments of the Brownian motion, and the independence between successive jumps.

Equation (4.2) shows that squared returns are an unbiased estimator for the daily return variance. It is, however, well known that squared returns are extremely noisy and inconsistent, since there is only a single daily return per trading day (Andersen and Bollerslev, 1998a).

RV, on the other hand, makes use of the availability of high-frequency data. If M intra-day returns are observed, then the RV is given by

$$RV_t = \sum_{i=1}^{M} r_{it}^2,$$

where r_{it} is the i-th intra-day return. Under the assumption that the log-price process follows a jump-diffusion, such as (4.1), with $\mu(\tau) = 0$, RV is an unbiased measure for the quadratic variation QV_t of the price process. For QV_t it holds that

$$QV_t = IV_t + \sum_{t-1 \leq \tau \leq t} J^2(\tau). \tag{4.3}$$

This means the quadratic variation equals the integrated variance plus the sum of squared jumps. For a review of these concepts cf. Andersen and Benzoni (2009).

Equations (4.2) and (4.3) therefore imply that

$$E[r_t^2] = Var[r_t] = E[RV_t]. \tag{4.4}$$

This equality is the basis for the arguments made in this paper. It implies the convergence of long term averages of r_t^2 and RV_t so that for $\overline{r_t^2} = T^{-1} \sum_{t=1}^{T} r_t^2$ and $\overline{RV_t} = T^{-1} \sum_{t=1}^{T} RV_t$, we have

$$\overline{r_t^2} - \overline{RV_t} \xrightarrow{p} 0,$$

as $T \to \infty$. Furthermore, we have

$$\sqrt{T} \Delta \sigma^2 = \sqrt{T} \left(\overline{r_t^2} - \overline{RV_t} \right) \xrightarrow{d} N(0, V), \tag{4.5}$$

where the long run variance V of the differential $r_t^2 - RV_t$ can be estimated with HAC estimators so that we can test the hypothesis that (4.4) is true using (4.5).

Note that the first equality in (4.2) and (4.4) holds generally as long as $\mu(\tau) = 0$. The equality $E[RV_t] = Var[r_t]$, on the other hand, only holds under the assumption that the log-price process $p(t)$ follows a jump-diffusion, since this implies the independence within and between continuous changes and jumps. A rejection of (4.4) is therefore indicative of a bias in RV_t and not in r_t^2.

To summarize, both RVs and squared returns approximate the same underlying variance process with the difference being that squared returns are noisier. Therefore, deviations between the two measures should be random and cancel each other out over time. Consequently, average squared return and average RV should be equal given a long enough horizon.

Figure 4.1 shows that this is not the case for a wide cross section of stock indices. Here, and in the following, we use 15-minute data for the years 1996 until 2017 from the Thomson Reuters Tick History database. The annualized average RV is plotted against the annualized average squared return for 22 commonly considered indices, such as the

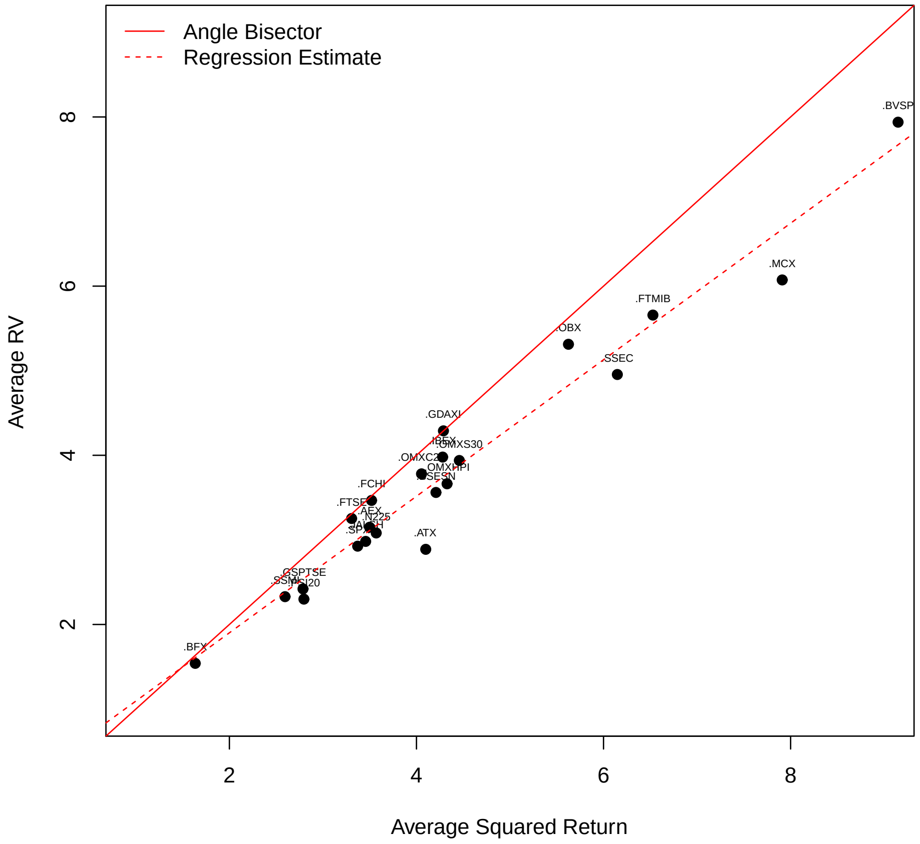

Figure 4.1: Average variance scatter plot.
The plot shows average annualized variance estimates for 22 stock indices using squared returns respectively RVs. RVs are calculated from 15-minute data and squared returns are adjusted for overnight returns so that both estimates are based on the same time horizon.

S&P 500, the DAX, and the SSEC. It can be seen that for all indices, except for the DAX, the average squared return is larger than the average RV in the same time period.[9]

Figure 4.2 sheds light on the relation of average RV and average squared return over time. In the left plot, the average variance estimate of the two estimators in a rolling window of 750 observations is displayed for the S&P 500. Again, it can be observed that the average RV is systematically smaller than the average squared return. This is not only true for isolated periods but holds all the time. However, the difference between the two time series seems to be larger in times of high variance, such as the subprime mortgage

---

[9] It should be noted that the squared returns are calculated from open-to-close returns so that overnight returns are excluded and the time horizon is the same as that for the RVs. As a robustness check, Figure 4.5 in the Appendix shows the results of the same analysis using RVs for 31 stock indices from *Oxford-Man Institute's realised library* compiled by Heber et al. (2009) and considered by Shephard and Sheppard (2010) and Han and Kristensen (2014), among others. These RVs are calculated from 10-minute data. It can be seen that the analysis yields very similar results, which underlines the robustness of our finding.

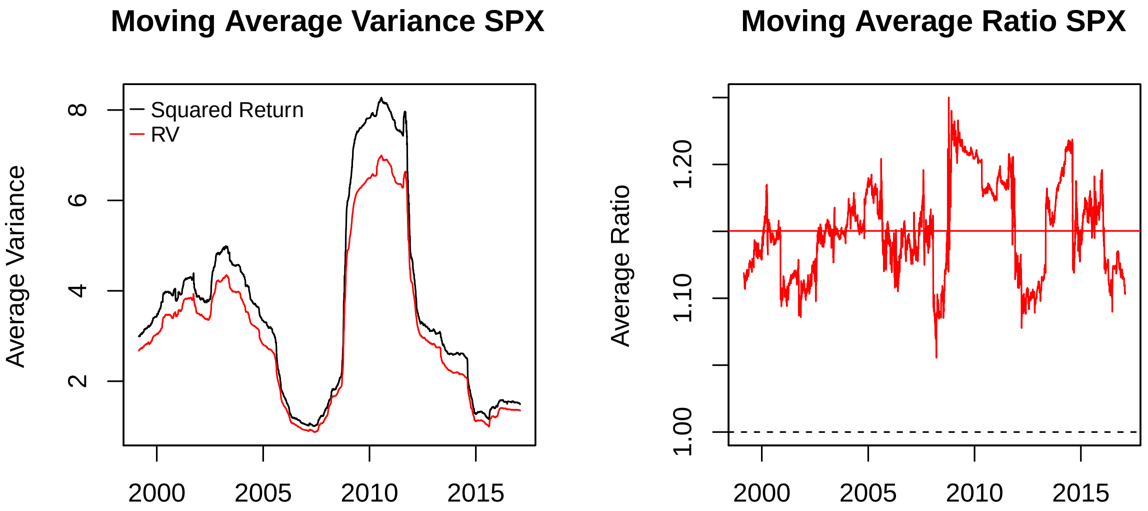

Figure 4.2: Average variance time-series plots S&P 500. Left: average annualized S&P 500 variance estimate of the last 750 observations using squared returns respectively RVs extracted from 15-minute returns. Right: ratio between rolling averages of squared return and RV.

crisis, indicating that average RV and average squared return differ by a factor rather than a constant. This factor is plotted over time on the right hand side of Figure 4.2. It seems relatively stable over time with an average of 1.15, which means that the RV underestimates the variance of the S&P 500 by 15 percent.[10]

Table 4.1 presents more detailed results for all 22 indices. It is apparent that the effect is especially pronounced for ATX, MCX, and SSEC with a factor between average squared return and average RV of 1.42, 1.30, and 1.24. This results in standard deviations that are larger by 3.25, 3.5, respectively 2.53 percentage points per annum than indicated by the RV estimator. For DAX, FCHI, and FTSE with factors of 1.00, 1.01, and 1.02, on the other hand, the effect is negligible. Averaged over all indices, the mean RV is 14 percent smaller than the mean squared return. This amounts to an annualized underestimation of the standard deviation by 1.34 percentage points.

The table further reports the test statistics $t_{HAC} = \sqrt{T}(\overline{r_t^2} - \overline{RV_t})/\sqrt{V_{HAC}}$ for autocorrelation and heteroscedasticity robust t-tests of the null hypothesis that $E[RV_t] = E[r_t^2]$. Here, V_{HAC} is the long run variance of $r_t^2 - RV_t$, which is estimated using the method of Andrews (1991). As a robustification, we also report the t_{MAC} statistic of Robinson (2005) and Abadir et al. (2009), which accounts for the possibility of long memory in

---

[10] Here and hereafter, we focus our analysis mostly on the S&P 500. Plots for other indices show that investigating any of the indices for which the effect is present would have yielded similar results. Some of these are displayed in Figure 4.10 to 4.12 in the Appendix. The others are available from the authors upon request.

RIC	Country	$\overline{r^2}$	$\overline{RV}$	$\overline{r^2}/\overline{RV}$	$\sqrt{\overline{r^2}} - \sqrt{\overline{RV}}$	t_{HAC}	t_{MAC}	t_{MOM}	T
.AEX	Netherlands	3.50	3.15	1.11	0.96	2.91***	2.48**	3.59***	4,567
.ATX	Austria	4.10	2.89	1.42	3.25	8.56***	4.31***	6.91***	4,179
.BFX	Belgium	1.63	1.54	1.06	0.38	1.91*	1.41	1.37	5,275
.BSESN	India	4.21	3.56	1.18	1.65	6.63***	3.86***	8.03***	4,948
.BVSP	Brazil	9.15	7.94	1.15	2.07	4.08***	2.58***	5.13***	4,728
.GDAXI	Germany	4.29	4.29	1.00	-0.00	-0.01	-0.01	0.91	5,264
.FCHI	France	3.52	3.47	1.02	0.14	0.58	0.49	1.24	5,272
.FTMIB	Italy	6.53	5.66	1.15	1.76	3.23***	3.43***	3.77***	1,942
.FTSE	Great Britain	3.31	3.25	1.02	0.15	0.53	0.87	2.77***	5,218
.GSPTSE	Canada	2.79	2.42	1.15	1.14	2.77***	3.73***	7.35***	3,654
.IBEX	Spain	4.28	3.98	1.08	0.74	2.74***	2.31**	3.40***	5,188
.JALSH	South Africa	3.46	2.98	1.16	1.32	4.72***	4.72***	6.27***	3,216
.MCX	Russia	7.91	6.07	1.30	3.48	6.19***	2.36**	7.78***	3,892
.N225	Japan	3.57	3.08	1.16	1.34	3.83***	0.76	4.59***	5,080
.OBX	Norway	5.62	5.31	1.06	0.67	1.44	4.87***	4.40***	2,679
.OMXC20	Denmark	4.05	3.78	1.07	0.69	1.78*	4.42***	4.48***	2,821
.OMXHPI	Finland	4.33	3.66	1.18	1.66	4.12***	4.04***	6.59***	2,835
.OMXS30	Sweden	4.46	3.94	1.13	1.27	3.13***	2.88***	5.39***	3,005
.PSI20	Portugal	2.80	2.30	1.22	1.56	4.89***	4.24***	5.08***	4,866
.SPX	United States	3.37	2.93	1.15	1.26	4.96***	3.47***	5.77***	5,183
.SSEC	China	6.15	4.96	1.24	2.53	7.24***	5.37***	9.10***	5,019
.SSMI	Switzerland	2.59	2.33	1.11	0.85	3.39***	1.90*	3.90***	4,770

Table 4.1: Statistics on average squared return and average RV. Reported are average squared return $\overline{r^2}$ per annum in percent, average RV $\overline{RV}$ per annum in percent, and ratio between the two measures for all of the 22 considered indices. For better assessing the degree of the bias, $\sqrt{\overline{r^2}} - \sqrt{\overline{RV}}$ is stated, which gives the average percentage points that the standard deviation implied by the two measures deviates per annum. Moreover, the table reports results of autocorrelation and heteroscedasticity robust t-tests t_{HAC} for the null hypothesis that the two estimates are equal. As it is commonly found in the literature that squared returns and RVs are highly persistent, t_{MAC} (Robinson, 2005; Abadir et al., 2009), which accounts for this degree of persistence, is also stated. Moreover, we report t_{MOM}, which yields valid inference if the return distribution does not exhibit unconditional finite fourth moments. For all tests *, **, and *** indicate that the null hypothesis $E[r_t^2] = E[RV_t]$ is rejected at the ten percent, five percent, and one percent level, respectively. Positive test statistics thereby imply that squared returns are larger on average. Finally, T gives the number of days considered for estimation.

$r_t^2 - RV_t$. This might be present since both, r_t^2 and RV_t, are commonly found to be highly persistent (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996; Ding and Granger, 1996). To account for the fact that the unconditional fourth moment of the return distribution might not exist, we conduct an additional test t_{MOM} for which the difference between RVs and squared returns is standardized by an estimate of the conditional standard deviation of the series. The test results suggest that for 16 (t_{HAC}), 13 (t_{MAC}), respectively 19 (t_{MOM}) indices the average squared return is significantly larger than the average RV at the one percent level.

To summarize, average squared returns and average RVs are not identical in expectations. Instead, mean squared returns are larger by a factor of 1.14. This observation is time consistent and can be found for all of the 22 considered indices except DAX, FCHI, and FTSE.

Recalling Equation (4.4) and the considerations stated thereafter, this implies that RV is a biased estimator for the variance of daily index returns. The next section provides a detailed investigation of possible explanations for this bias. Here, we provide evidence that the deviation between average squared return and average RV is caused by dependencies in intra-day returns that violate the assumptions required for consistency of RV as an estimator for the daily variance. Section 4.3 then concludes.

4.2 Explaining the Bias of Realized Variance

To determine the source of the difference between average squared return and average RV, it is useful to decompose the observed continuously compounded return r_{it} into its components.

It is now broadly accepted that stock prices can be represented by a jump-diffusion model such as (4.1) (e.g., Aït-Sahalia, 2004; Barndorff-Nielsen and Shephard, 2007; Corsi et al., 2010). Consequently, we can decompose the continuously compounded stock return at time i on day t r_{it} into jump component J_{it}, continuous component C_{it}, and equity premium. As mentioned before, we assume that the equity premium is zero to ease the presentation.[11]

Another important component of the observed return at high frequencies is market microstructure noise due to price discreteness, bid-ask spreads, trades taking place at different markets and networks, gradual response of prices to a block trade, difference in information contained in orders of different sizes, strategic order flows, and recording errors. Starting with Zhou (1996), numerous contributions find these effects to significantly influence the observed intra-day return at high frequencies such as 1-second data. At low frequencies, however, the effect is often found to be negligible. To capture this characteristic, market microstructure noise is denoted by $\eta_{it,M}$ in the following with M defined as the number of intra-day observations and $E[\eta_{it,M}] = 0$. We then assume that market microstructure effects are not present on a daily frequency, i.e., $Var(\eta_{it,1}) = 0$, and that $Var(\eta_{it,M})$ is monotonically increasing with the sampling frequency M.

The observed continuously compounded return can then be written as

$$r_{it} = C_{it} + J_{it} + \eta_{it,M}$$

---

[11] Our results would not be altered by a non-zero equity premium unless it would exhibit sizable intra-day variation (which is theoretically implausible).

so that for the two estimators it holds that

$$\begin{aligned} RV_t &= \sum_{i=1}^{M} (C_{it} + J_{it} + \eta_{it,M})^2 \\ \text{and} \quad r_t^2 &= \left(\sum_{i=1}^{M} C_{it} + \sum_{i=1}^{M} J_{it} \right)^2 . \end{aligned}$$

Calculating the difference between the two estimators yields

$$\begin{aligned} r_t^2 - RV_t = & \sum_{i,j=1, i \neq j}^{M} C_{it} C_{jt} + \sum_{i,j=1, i \neq j}^{M} J_{it} J_{jt} + 2 \sum_{i,j=1, i \neq j}^{M} C_{it} J_{jt} \\ & - \sum_{i=1}^{M} \eta_{it,M} \eta_{it,M} - 2 \sum_{i=1}^{M} C_{it} \eta_{it,M} - 2 \sum_{i=1}^{M} J_{it} \eta_{it,M} . \end{aligned}$$

To simplify notation, let $AB_t = \sum_{i,j=1, i \neq j}^{M} A_{it} B_{jt}$ and $AB_t^* = \sum_{i=1}^{M} A_{it} B_{it}$, such that

$$r_t^2 - RV_t = CC_t + JJ_t + 2CJ_t - \eta\eta_{t,M}^* - 2C\eta_{t,M}^* - 2J\eta_{t,M}^* . \tag{4.6}$$

The first two terms capture the intra-day dependencies in continuous and jump component. If, for example, $E[CC_t]$ is positive, then positive and negative intra-day continuous returns would tend to occur in clusters. It would therefore be more likely that C_{it} is a large positive return if $(C_{1t}, \ldots, C_{i-1t}, C_{i+1t}, \ldots, C_{Mt})$ are large positive returns. The third term captures the dependencies between the leads and lags of jump and continuous component. If $E[CJ_t]$ is positive, then it would be more likely to observe positive continuous returns at days where a positive jump occurs. The fourth term captures the variance of the market microstructure noise and the last two terms capture intra-day dependencies between the noise component and the continuous and jump components. If, for example, $E[C\eta_{t,M}^*]$ is positive, then it is more likely to observe positive microstructure noise $\eta_{it,M}$ if C_{it} is large and positive.

As mentioned before, when calculating the RV estimator it is commonly assumed that the log-price process follows a jump-diffusion such as (4.1) and that markets are frictionless. This implies that all of the terms in Equation (4.6) are zero in expectations and the expected values of squared return and RV are identical. However, Figures 4.1 and 4.2 together with Table 4.1 show that the average squared return is systematically larger than the average RV for a wide cross section of stock indices. Hence, at least one of the terms in Equation (4.6) has to be significantly larger than zero to explain the negative bias of RV as an estimator for the daily variance. In the following, we therefore analyze each of the terms in Equation (4.6) separately to determine the source of the bias.

4.2.1 Market Microstructure Noise

It is well established that market microstructure effects cause biased RV estimates (e.g., Hansen and Lunde, 2006; Bandi and Russell, 2006). Therefore, it is tempting to conjecture that the difference between average squared return and average RV can be attributed to the presence of microstructure noise. Since microstructure noise is not observable and cannot be estimated without access to tick data, this conjecture can only be refuted on the basis of plausibility arguments. These, however, are quite compelling.

First, the results presented here are obtained using 15-minute data to mitigate the impact of microstructure noise right from the start. Microstructure effects are commonly found to be relevant on high frequencies such as 1-second data. For sampling frequency lower than five minutes as considered here, Bandi and Russell (2008) argue that the effect of market microstructure noise is negligible. This is also confirmed by Figures 4.6 and 4.7 in the Appendix, which show that repeating our analysis for 5-minute and 30-minute data yields qualitatively similar results.

Second, market microstructure noise only generates a negative bias in the RV if $\eta\eta^*_{t,M} < -2C\eta^*_{t,M} - 2J\eta^*_{t,M}$. This would imply negative correlation between noise and continuous component respectively between noise and jump component which outweighs the variance of the market microstructure noise. This is typically not the case as can be seen in the volatility signature plots of, for example, Hansen and Lunde (2006), Bandi and Russell (2006), and Aït-Sahalia et al. (2011). Here, market microstructure effects generate a positive bias in the RV. In contrast to that, the bias observed here is negative.

Third, as a final robustness check, Figure 4.8 in the Appendix repeats the analysis of Figure 4.1 for the realized kernel variance estimator of Barndorff-Nielsen et al. (2008) that is constructed to be robust against market microstructure noise. Again, it can be observed that the average squared returns are significantly larger.

We therefore conclude that market microstructure effects cannot explain the difference between average RV and average squared return documented in Section 4.1. Since we use 15-minute data, it seems reasonable to assume that the magnitude of η_{it} is negligible and Equation (4.6) can be further simplified such that

$$r_t^2 - RV_t \;\approx\; CC_t + JJ_t + 2CJ_t. \tag{4.7}$$

4.2.2 Jumps and Continuous Returns

To determine the relative magnitude of the remaining terms in Equation (4.7), we need to decompose the intra-day returns r_{it} into continuous and jump components. While numerous model-free estimators and tests have been proposed to disentangle the contribution of jumps and continuous components to the daily RV (e.g., Barndorff-Nielsen and Shephard, 2004b; Aït-Sahalia and Jacod, 2009; Corsi et al., 2010), only the methodology of Lee and

Mykland (2007) is able to determine jump and continuous components for every intra-day return. The idea of Lee and Mykland (2007) is to compare each 15-minute return to an estimate of the volatility using the previous K observations. If the 15-minute return is large in comparison to the volatility of the previous observations, then it is concluded that a jump occurred. Since $Var(C_{it}) = O(M^{-1})$, the jump asymptotically dominates the continuous component as $M \to \infty$ and r_{it} is a suitable estimator for J_{it}. Consequently, if the test rejects the null of no jump at time i on day t, we conclude that a jump of size r_{it} has occurred.

Lee and Mykland (2007) suggest to estimate the volatility of the previous K observations using the bipower variation introduced by Barndorff-Nielsen and Shephard (2004b). However, Corsi et al. (2010) show that this estimator is substantially biased in finite samples leading to a large underestimation of the jump component. To circumvent this problem, we consider their threshold bipower variation estimator. This is less affected by small sample bias and has the same limit as bipower variation in probability.[12]

The test statistic then evolves as

$$\mathcal{L}_{it} = \frac{r_{it}}{\sqrt{\hat{\sigma}_{it}^2}}, \qquad \text{where} \qquad \hat{\sigma}_{it}^2 = \frac{\pi}{2} \frac{1}{K-2} \sum_{j=i-M+2}^{i-1} |r_j||r_{j-1}| I(r_j^2 \leq \theta) I(r_{j-1}^2 \leq \theta).$$

Here, $I(\cdot)$ is the indicator function and θ is a threshold, which is estimated using the approach suggested by Corsi et al. (2010) that ensures that jumps do not influence the estimation of $\hat{\sigma}_{it}^2$.

In the absence of jumps, a single test is standard normally distributed. For multiple testing as it is performed here, critical values are derived based on Gaussian extreme value theory.

Since the variance estimator $\hat{\sigma}_{it}^2$ is the threshold bipower variation by Corsi et al. (2010), the test is still consistent if a jump has already occurred in one of the previous K observations. As suggested by Lee and Mykland (2007), we set $K = 156$ and employ a significance level of one percent to decrease the likelihood of spuriously detecting jumps.

After decomposing every return into jump and continuous component, we are able to calculate daily values of CC_t, CJ_t, and JJ_t. Panel A of Table 4.2 provides t-statistics robust to autocorrelation and heteroscedasticity for the null hypothesis that $E[CC_t] = 0$, respectively $E[CJ_t] = 0$ or $E[JJ_t] = 0$.[13] For all indices with a significant bias, it can be seen that the component CC_t is positive and significantly different from zero at the one percent level. For the S&P 500, for example, the value of the test statistic is 4.83. The components CJ_t and JJ_t, on the other hand, are not indicated to be significantly

---

[12] The results using the bipower variation estimator are qualitatively similar and available upon request.

[13] In analogy, Tables 4.4 and 4.5 in the Appendix state t_{MAC} and t_{MOM}, i.e., the t-statistics when accounting for persistence respectively infinite unconditional fourth moments of the return distribution. The results are qualitatively similar.

		A: 15-minute Data			B: Seasonally Adjusted			C: 5-minute Data		
RIC	Country	CC	CJ	JJ	CC	CJ	JJ	CC	CJ	JJ
.AEX	Netherlands	3.21***	0.64	0.38	3.18***	0.17	0.43	2.23**	1.81*	0.38
.ATX	Austria	10.47***	0.64	-0.51	10.09***	-0.21	-0.57	11.26***	1.57	2.67***
.BFX	Belgium	1.78*	0.94	-0.54	1.66*	1.20	-0.74	1.10	0.80	1.04
.BSESN	India	7.17***	-0.46	-0.59	7.10***	-1.51	1.42	6.79***	3.49***	-0.72
.BVSP	Brazil	6.10***	1.62	-1.76*	4.60***	0.51	-1.55	6.82***	4.12***	-0.71
.GDAXI	Germany	-0.02	0.38	-0.66	0.12	-0.31	0.07	-0.68	1.51	0.82
.FCHI	France	0.97	-0.13	-0.69	0.51	-0.05	0.55	2.08**	2.10**	0.77
.FTMIB	Italy	3.99***	1.89*	-0.40	3.46***	1.94*	-0.53	3.28***	1.56	0.47
.FTSE	Great Britain	2.04**	0.18	-2.07**	2.16**	-0.98	-1.92*	-3.53***	1.06	1.90*
.GSPTSE	Canada	5.05***	0.18	-0.48	5.21***	-0.87	-1.65*	6.60***	-0.24	-1.78*
.IBEX	Spain	3.53***	0.60	-0.35	3.07***	0.73	0.41	2.95***	1.67*	0.27
.JALSH	South Africa	5.79***	1.91*	-1.07	5.31***	1.35	-0.17	4.97***	3.81***	-0.66
.MCX	Russia	7.31***	1.81*	-0.69	6.10***	3.03***	-0.73	7.62***	1.84*	0.73
.N225	Japan	3.34***	2.17**	1.45	3.71***	1.49	1.69*	3.58***	4.98***	2.76***
.OBX	Norway	2.01**	0.43	-1.26	1.91*	-0.26	0.53	2.87***	-1.00	-2.2**
.OMXC20	Denmark	4.11***	-0.77	-1.71*	2.00**	1.44	-2.22**	4.92***	0.55	-2.01**
.OMXHPI	Finland	5.18***	1.65*	-1.13	4.43***	0.50	0.88	6.65***	1.65*	0.41
.OMXS30	Sweden	2.93***	1.86*	-1.11	2.52**	2.11**	-1.19	2.95***	1.03	-0.77
.PSI20	Portugal	6.31***	-0.42	0.99	5.61***	0.47	1.04	4.22***	0.47	0.95
.SPX	United States	4.83***	1.12	-1.98**	4.79***	0.68	-2.33**	5.39***	3.27***	-0.74
.SSEC	China	7.68***	2.43**	1.03	7.78***	2.59***	0.99	8.49***	3.91***	1.38
.SSMI	Switzerland	2.79***	0.78	1.49	2.37**	3.17***	1.48	1.54	2.36**	1.46

Table 4.2: HAC t-test statistics.
Reported are HAC t-test statistics for the null hypothesis that $E[CC_t] = 0$, $E[CJ_t] = 0$, and $E[JJ_t] = 0$. *, **, and *** indicate that the null hypothesis is rejected at the ten percent, five percent, and one percent level, respectively. In Panel A, CC_t, CJ_t, and JJ_t are determined using 15-minute data and the methodology by Lee and Mykland (2007) with the threshold bipower variation instead of the bipower variation to estimate the instantaneous volatility. In analogy, Panel B shows the test results when the estimates of σ^2_{it} are adjusted for intra-day seasonality as found by Andersen and Bollerslev (1997) and Andersen and Bollerslev (1998b), and in Panel C, the three components are estimated using 5-minute data.

different from zero at the one percent level for any of the indices. It can further be seen that for the DAX and the FCHI for which Table 4.1 reports no significant bias of the RV estimator, CC_t is not significantly different from zero. For the FTSE, Table 4.2 reports significant dependence at the five percent level in the continuous component, although Table 4.1 states that there is no significant bias. The reason for this observation is that for the FTSE there is significant negative dependence in the jump component JJ_t at the five percent level which compensates for the bias caused by CC_t.

To show the robustness of the results, Panel B of Table 4.2 repeats the analysis adjusting the estimate of $\hat{\sigma}^2_{it}$ for intra-day seasonality as documented by Andersen and Bollerslev (1997) and Andersen and Bollerslev (1998b), among others, and Panel C reports the test statistics using 5-minute instead of 15-minute data. The results are qualitatively similar with CC_t being positive and significantly different from zero for most indices and CJ_t and JJ_t being insignificant for most indices.

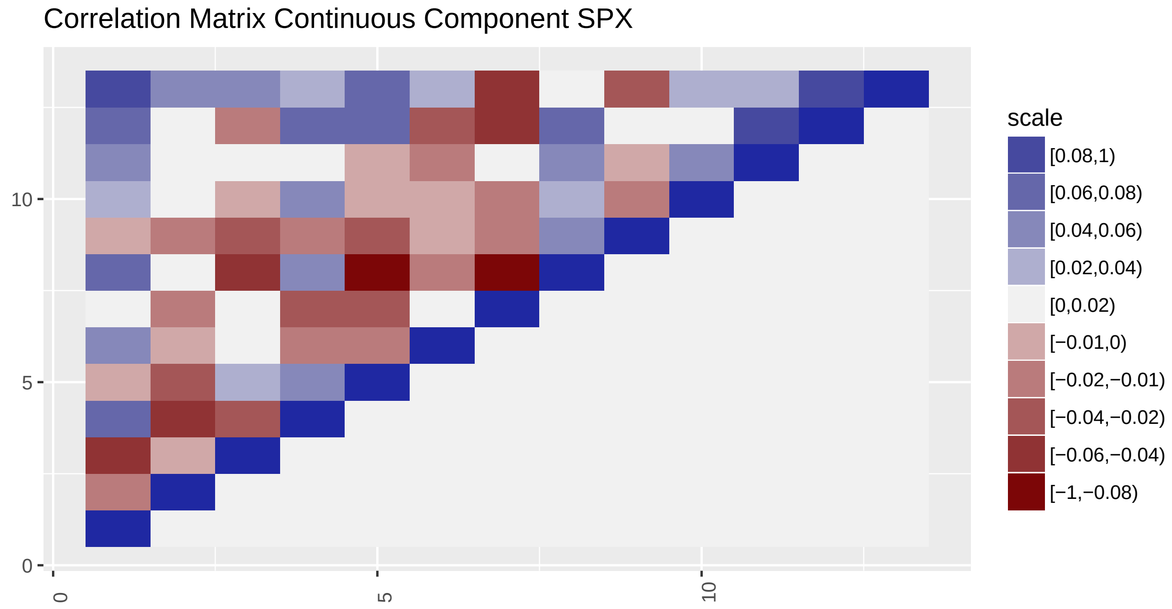

Figure 4.3: Correlation matrix S&P 500.
Depicted is the correlation matrix for the continuous return component (CC_t) of the S&P 500 using 30-minute Returns.

The term CC_t captures the dependence structure in C_{it}. Let $c_t = (C_{1t}, \ldots, C_{Mt})'$ denote the vector of continuous returns on day t and let ι be an $M \times 1$ vector of ones. Then,

$$CC_t = \iota'(c_t c_t')\iota - c_t' c_t.$$

It is well established, that the autocorrelation function of r_{it} is essentially zero at all leads and lags. This justifies the semimartingale assumptions imposed in (4.1) that implies that all off-diagonal elements of $E[c_t c_t']$ are zero so that $E[CC_t] = 0$.

An unfortunate property of the autocorrelation function, however, is the fact that it masks dependencies that do not depend on the lag but on the location of the returns within a trading day. In a recent paper, Gao et al. (2018) find significant correlation between 30-minute intra-day returns of the S&P 500. This concerns in particular the last two returns in a trading day and the first and last returns of the day. They argue that these correlation patterns stem from investors infrequent rebalancing of their portfolios and late-informed investors who trade early morning information in the last hour, where liquidity is larger.

To shed further light into the dependence structure of the continuous components of the intra-day returns of the S&P 500, Figure 4.3 shows their average correlation matrix. In line with Gao et al. (2018), the correlation matrix shows a momentum effect between first and last half hour of the trading day and in the last hour of the trading day. However, the plot also reveals positive as well as negative correlation between the other half-hour

Return locations within a trading day		Correlation	p-value	Critical p-value
$\rho_{5,8}$	*(11:30 - 12:00, 13:00 - 13:30)*	-0.0888	0.0004	0.0006
$\rho_{1,13}$	*(09:30 - 10:00, 15:30 - 16:00)*	0.1314	0.0010	0.0013
$\rho_{1,4}$	*(09:30 - 10:00, 11:00 - 11:30)*	0.0740	0.0010	0.0019
$\rho_{12,13}$	*(15:00 - 15:30, 15:30 - 16:00)*	0.1379	0.0016	0.0026
$\rho_{11,12}$	*(14:30 - 15:00, 15:00 - 15:30)*	0.0815	0.0025	0.0032

Table 4.3: Significant correlations S&P 500.
The table shows all correlations between half-hour returns of the S&P 500 that are significantly different from zero after accounting for the multiple testing problem. The last column states the corresponding critical p-values for an alpha of five percent when applying Simes' correction (Simes, 1986).

returns of the S&P 500. When testing for the joint significance of all pairwise correlation coefficients, we obtain a chi-square statistic of 168, which vastly exceeds the critical value of 110 at the 1 percent level.

When testing for the significance of individual correlations, there is a multiple testing problem. We account for this by applying Simes correction (Simes, 1986), which consists of ordering all p-values in ascending order and then comparing them with $\frac{\alpha}{N}, \frac{2\alpha}{N}, \ldots, \frac{N\alpha}{N}$ with $N = 78$ being the number of performed tests. If any of the ordered p-values exceeds its respective threshold, the null hypothesis that the corresponding correlation between the two returns equals zero is rejected. Table 4.3 reports all combinations of returns for which this is the case at an alpha level of five percent. The table states that there is significant positive correlation between first and fourth, first and thirteenth (last), fourth and twelfth, eleventh and twelfth and the last two half-hour returns. In contrast, negative correlation that is significantly different from zero only exists between the fifth and the eighth half-hour return. With regard to the strength of the dependency, the correlation between first and last and second last and last half-hour return are found to be the largest.

To summarize, there is significant positive correlation in continuous index returns. This does not only hold when considering the correlations separately but also when taking them all together. This violates the semimartingale assumption that is necessary for the consistency of the RV as an estimator for the variance of daily stock returns and causes a significant negative bias as observed in Figure 4.1.[14] Since the correlation is in the continuous component, jump robust estimators such as the BV estimator (Barndorff-Nielsen and Shephard, 2004b) or the medRV estimator (Andersen et al., 2012) are biased

---

[14] There is a growing body of literature which provides evidence that jumps are often erroneously identified when estimating them from 5 or 15-minute data. When considering tick data, where estimation precision is higher, the jump component is found to account for only a small fraction of the total price variation making it almost negligible (Christensen et al., 2014; Bajgrowicz et al., 2015). Additionally, it is theoretically unclear whether the approach by Lee and Mykland (2007) is still consistent under this kind of dependence. Consequently, we might have misidentified jump and continuous component, at least for some returns. We should therefore note that performing our analysis without differentiating between jump and continuous component yields qualitatively the same results, i.e., there is correlation in index returns which causes biased RV estimates.

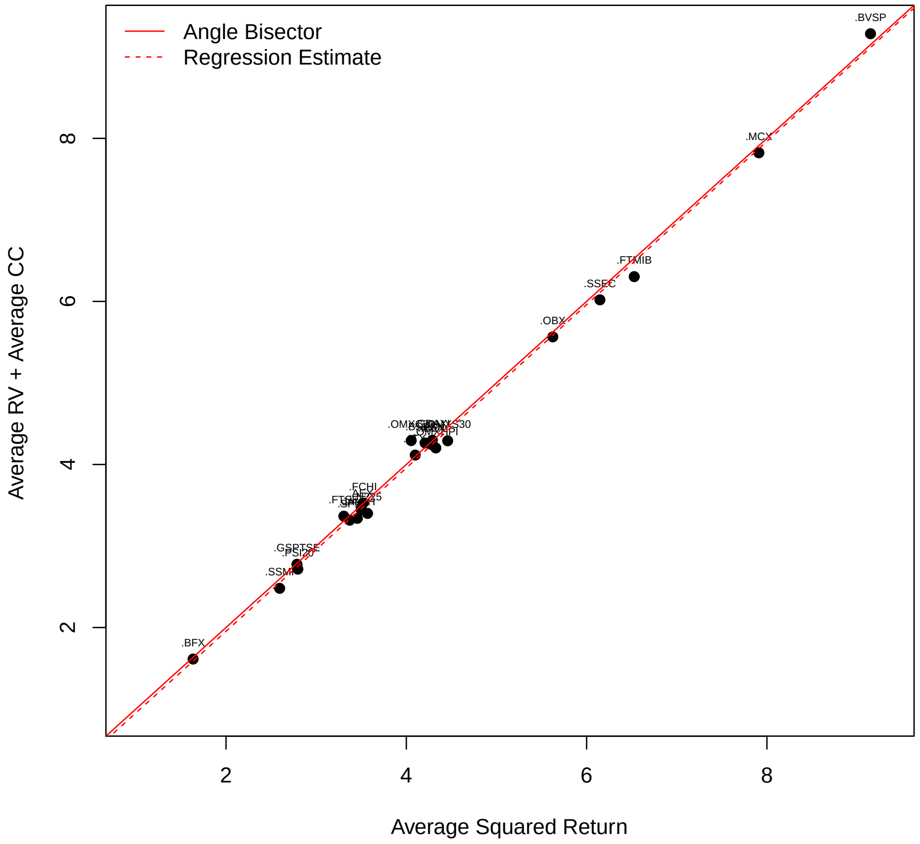

Figure 4.4: Corrected average variance scatter plot.
In analogy to Figure 4.1 with the difference that the RVs are now corrected by CC_t. Again, the RV estimates are calculated from 15-minute data and the squared returns are adjusted for overnight returns so that both estimates are based on the same time horizon.

as well. This also holds for sub-sampled RV estimators and realized kernel variance estimators (Barndorff-Nielsen et al., 2008). These can handle frictions in event time which vanish with decreasing sampling frequency. They are, however, not robust to correlations in calendar time as found here.

To further illustrate that CC_t causes the bias, Figure 4.4 repeats the analysis of Figure 4.1 and plots the average squared return against the sample average of the corrected RV measure $\widetilde{RV}_t = RV_t + CC_t$. From Equation (4.7), adding CJ_t and JJ_t to $\widetilde{RV}_t$ would give exactly r_t^2. It can be seen that accounting for the CC_t component almost completely eliminates the difference observed in Figure 4.1. Moreover, robust t-tests reject $E[r_t^2 - \widetilde{RV}_t] = 0$ not once for any of the indices at the one percent level and moving average plots indicate the factor between r_t^2 and $\widetilde{RV}$ to be close to one all the time. More detailed results can be found in the Appendix in Table 4.6 and Figure 4.9 which repeat the analysis of Table 4.1 and Figure 4.2 for the corrected RV measure.

As a final check whether for the S&P 500 the correlation in the first and last hour of the trading day is the reason for the observed bias, we repeat our analysis excluding the first and last hour of the trading day for the calculation of RVs and squared returns. As expected, average squared return and average RV now have a ratio of 0.97 and are not indicated to be significantly different from another at any level. For the other indices, however, it is not necessarily correlation in the first and last hour of the trading day that causes the negative bias reported in Table 4.1. If we repeat this analysis for these indices, then there is still a significant negative bias of the RV estimator for 12 (t_{HAC}), 8 (t_{HAC}), respectively 11 (t_{MOM}) indices at the one percent level. More detailed results can be found in Table 4.7 in the Appendix that repeats the analysis of Table 4.1 when excluding the first and last hour of the trading day.

4.3 Conclusion

As an ex-post measure of the quadratic variation of the price process, RV has become the standard measure for volatility estimation. While RV is often used to estimate the volatility of daily stock returns, this is only a valid approach if the log-price process is a semimartingale or a jump-diffusion.

As shown here, there are significant correlations between intra-day returns that are in contradiction to the semimartingale assumption and cause a considerable bias if RV is used as an estimate for the variance of daily or weekly index returns.

While previous research on market microstructure effects has focused on frictions in event time, these results indicate that structural effects in calendar time should be investigated further to illuminate the source of these intra-day dependencies.

Another important task for further research is the development of bias-corrected RV estimates that combine the unbiasedness of squared returns with the low variance of RV estimates. The most intuitive idea would be to add CC_t to the daily RV estimate, i.e., use $\widetilde{RV}_t$. While this solves the bias problem, it brings back the noise problem since $\widetilde{RV}_t$ is almost as noisy as r_t^2 itself. A more promising way would be to assume that the sum of the correlations between the intra-day returns is constant so that $E[CC_t/\sigma_t^2] = \rho$. In this case, $E[RV_t]$ and $E[r_t^2]$ differ by the constant $1 + \rho$ that can be estimated by $\overline{r_t^2}/\overline{RV_t}$. A similar type of estimator has been introduced in Hansen and Lunde (2005) to account for market microstructure effects.

Appendix

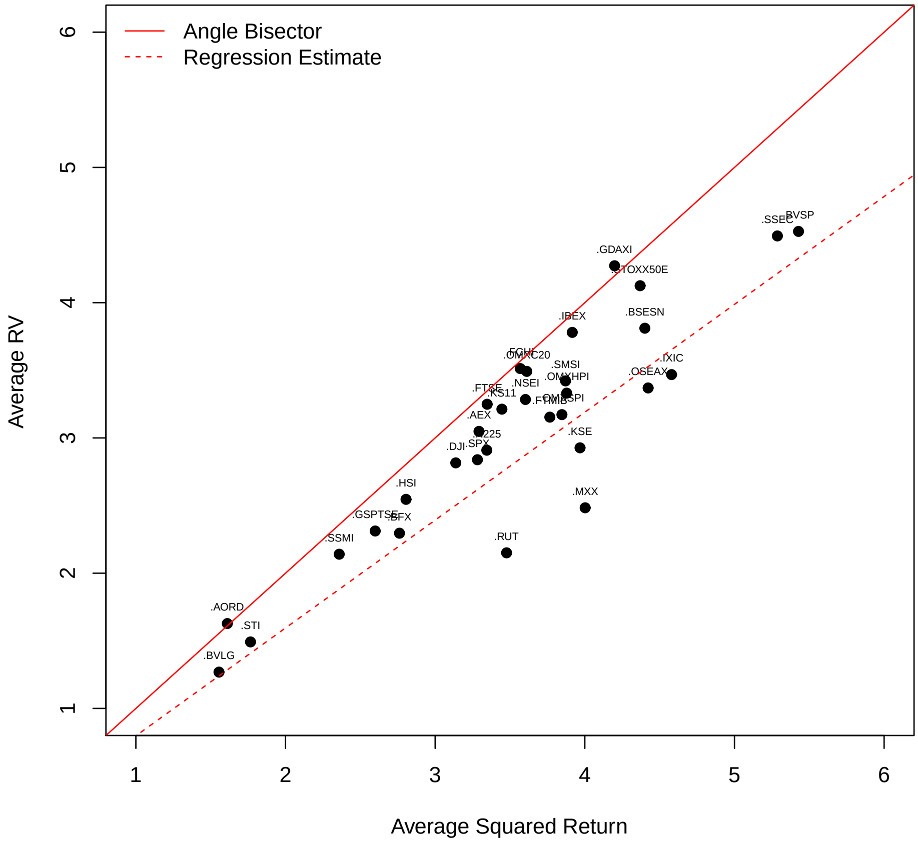

Figure 4.5: Average variance scatter plot — realized library. In analogy to Figure 4.1 with the difference that now data from *Oxford-Man Institute's realised library* (Heber et al., 2009) for 31 stock indices is considered. Here, the RVs are calculated from 10-minute data. As before, the squared returns are adjusted for overnight returns so that both estimates are based on the same time horizon.

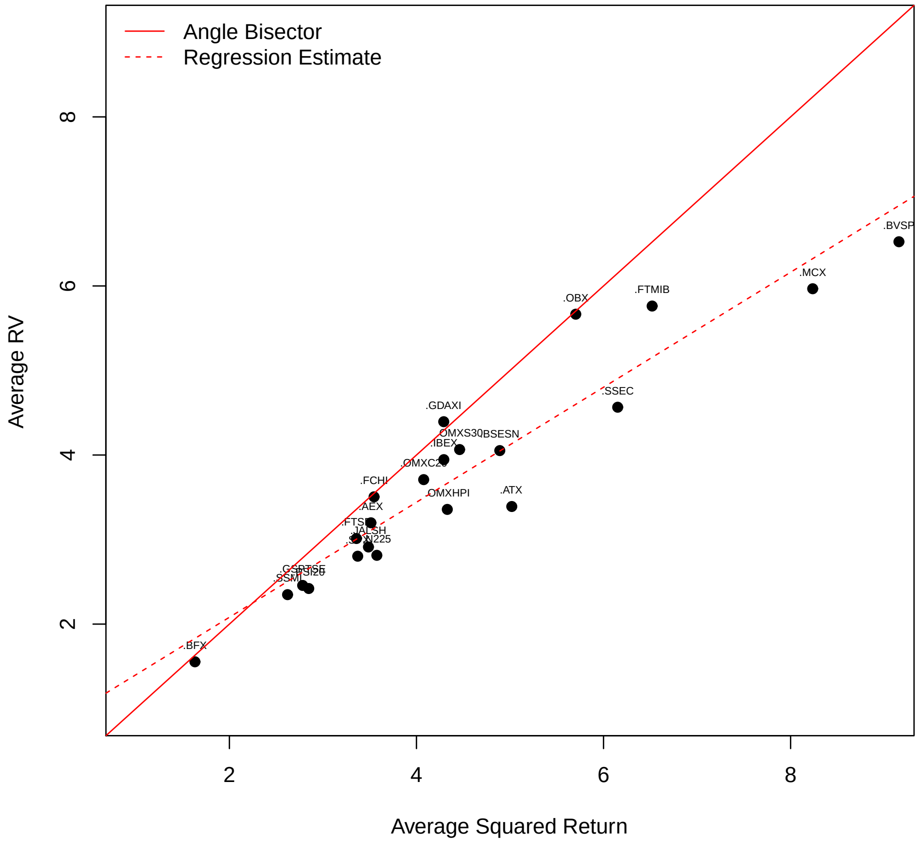

Figure 4.6: Average variance scatter plot — 5-minute data.
In analogy to Figure 4.1 with the difference that RVs are now calculated from 5-minute data. As before, squared returns are adjusted for overnight returns so that both estimates are based on the same time horizon.

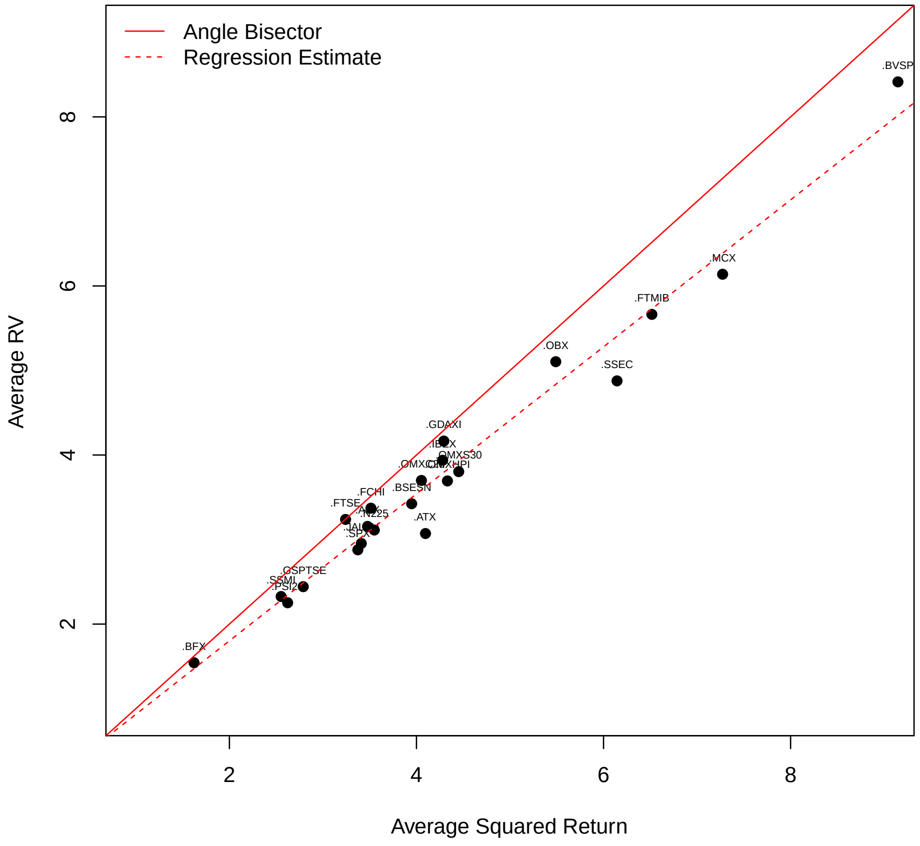

Figure 4.7: Average variance scatter plot — 30-minute data. In analogy to Figure 4.1 with the difference that RVs are now calculated from 30-minute data. As before, squared returns are adjusted for overnight returns such that both estimates are based on the same time horizon.

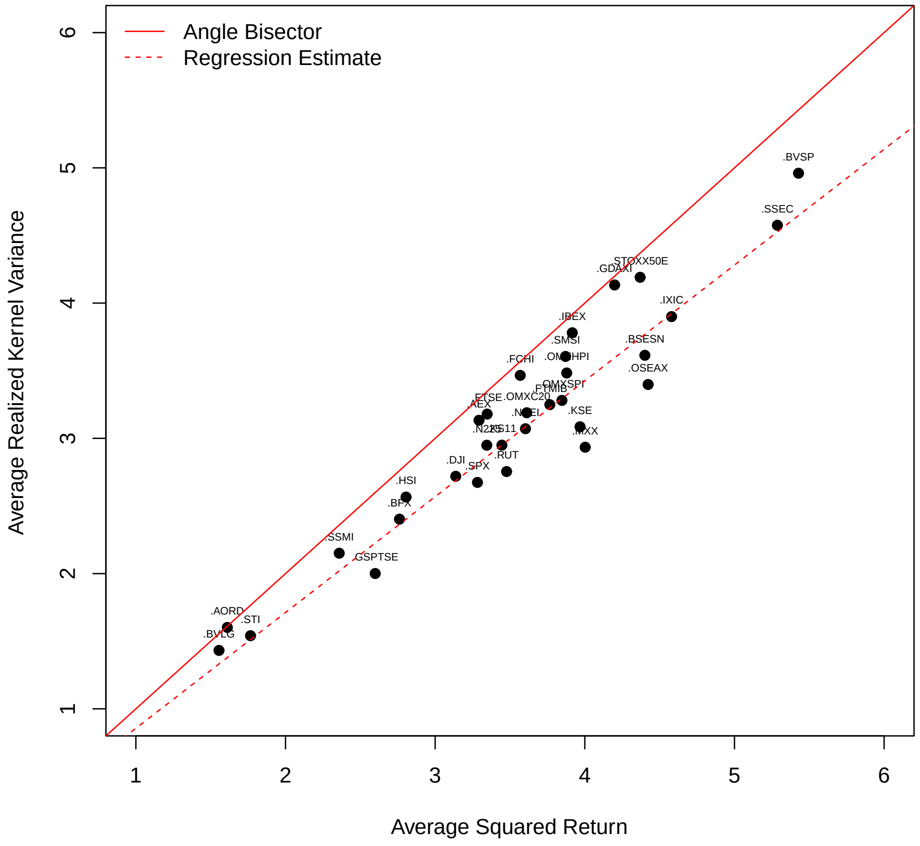

Figure 4.8: Average variance scatter plot — realized kernel variance. In analogy to Figure 4.5 again using the data from the *Oxford-Man Institute's realised library* for the 31 indices. Now, however, the realized kernel variance, which is robust to market microstructure noise, is depicted on the y-axis.

		A: 15-minute Data			B: Seasonally Adjusted			C: 5-minute Data		
RIC	Country	CC	CJ	JJ	CC	CJ	JJ	CC	CJ	JJ
.AEX	Netherlands	3.09***	0.56	0.29	2.32**	0.23	0.42	1.50	1.21	0.67
.ATX	Austria	2.81***	0.50	-0.47	2.49**	-0.18	-0.54	3.07***	1.18	0.72
.BFX	Belgium	0.74	1.63	-0.42	0.74	1.16	-0.94	0.60	0.66	0.89
.BSESN	India	2.21**	-0.40	-0.58	2.42**	-1.08	1.29	2.23**	3.51***	-0.71
.BVSP	Brazil	1.74*	1.40	-1.35	2.32**	0.46	-1.31	1.58	1.91*	-0.69
.GDAXI	Germany	-0.04	0.35	-0.66	0.14	-0.33	0.06	-0.53	1.60	1.29
.FCHI	France	1.17	-0.15	-0.69	0.64	-0.05	0.64	1.68*	1.81*	0.72
.FTMIB	Italy	4.16***	1.19	-0.40	4.15***	1.75*	-0.54	2.19**	2.90***	0.99
.FTSE	Great Britain	1.46	0.24	-1.83*	2.60***	-0.60	-1.85*	-2.70***	1.67*	3.32***
.GSPTSE	Canada	2.53**	0.05	-0.61	1.92*	-0.24	-0.87	2.68***	-0.34	-1.58
.IBEX	Spain	3.73***	0.57	-0.34	3.01***	0.76	0.41	2.80***	2.31**	0.23
.JALSH	South Africa	8.44***	2.57**	-0.96	7.96***	2.00**	-0.17	6.98***	2.55**	-1.01
.MCX	Russia	2.42**	1.88*	-0.62	1.71*	2.35**	-0.53	1.68*	1.98**	0.46
.N225	Japan	2.71***	1.16	1.13	1.33	0.65	1.71*	0.26	3.37***	2.92***
.OBX	Norway	5.51***	1.56	-0.33	7.41***	-1.08	0.36	5.36***	-2.81***	-1.69*
.OMXC20	Denmark	4.31***	-1.11	-0.56	3.77***	2.56**	-0.94	4.19***	1.89*	-1.30
.OMXHPI	Finland	5.85***	2.83***	-0.95	11.05***	1.06	0.95	4.14***	3.94***	0.46
.OMXS30	Sweden	2.86***	3.77***	-1.11	2.66***	2.49**	-1.16	2.25**	1.92*	-0.89
.PSI20	Portugal	3.24***	-0.49	1.03	4.19***	0.41	1.08	0.84	0.53	0.87
.SPX	United States	1.39	1.68*	-2.52**	1.47	0.67	-2.43**	2.26**	3.73***	-0.62
.SSEC	China	4.97***	3.10***	0.79	5.54***	3.86***	0.63	4.12***	3.21***	1.22
.SSMI	Switzerland	2.33**	0.87	1.56	1.66*	3.88***	1.54	0.96	2.38**	1.50

Table 4.4: MAC t-test statistics.
In analogy to Table 4.2 with MAC instead of HAC t-test statistics. Again, Panel A is based on 15-minute data, Panel B adjusts for intra-day seasonality and Panel C is based on 5-minute data.

		A: 15-minute Data			B: Seasonally Adjusted			C: 5-minute Data		
RIC	Country	CC	CJ	JJ	CC	CJ	JJ	CC	CJ	JJ
.AEX	Netherlands	2.84***	0.32	0.81	2.85***	1.81*	0.24	0.88	2.22**	1.09
.ATX	Austria	12.13***	1.17	-0.35	12.46***	0.48	-0.61	13.05***	0.77	2.67***
.BFX	Belgium	1.42	1.00	-0.52	1.80*	1.61	-0.79	0.70	-0.75	1.52
.BSESN	India	8.25***	-0.24	-0.96	8.75***	-0.46	0.79	6.67***	4.31***	-2.12**
.BVSP	Brazil	7.37***	0.22	-1.74*	6.22***	0.60	-1.54	11.63***	2.97***	-0.51
.GDAXI	Germany	0.85	1.02	0.04	0.87	-0.46	0.56	-2.41**	2.7***	2.04**
.FCHI	France	0.98	0.79	-0.67	0.49	0.55	0.39	-0.98	2.93***	1.46
.FTMIB	Italy	3.90***	2.64***	-0.44	4.38***	1.95*	-0.68	2.15**	3.02***	0.60
.FTSE	Great Britain	3.99***	1.44	-1.48	5.01***	-0.01	-1.86*	6.34***	1.99**	1.27
.GSPTSE	Canada	8.71***	1.34	-0.13	8.00***	-1.39	-1.33	12.14***	0.60	-1.13
.IBEX	Spain	4.94***	0.98	-0.28	4.15***	1.58	0.22	4.03***	1.71*	0.20
.JALSH	South Africa	6.97***	1.36	-0.84	6.36***	2.46**	-0.52	6.88***	3.65***	0.10
.MCX	Russia	12.04***	1.88*	-0.68	10.79***	3.57***	-0.68	13.38***	4.10***	1.33
.N225	Japan	3.87***	1.62	2.05**	2.38**	1.63	1.70*	3.91***	5.09***	3.32***
.OBX	Norway	4.19***	1.89*	-1.19	4.64***	0.35	0.08	5.21***	-0.02	-1.58
.OMXC20	Denmark	6.22***	0.43	-1.87*	4.98***	1.45	-2.11**	6.40***	1.52	-1.71*
.OMXHPI	Finland	8.12***	2.53**	-0.61	7.82***	1.29	0.91	9.56***	2.67***	3.02***
.OMXS30	Sweden	4.86***	1.74*	-1.06	5.03***	2.35**	-1.12	3.82***	1.23	-0.86
.PSI20	Portugal	6.85***	-0.34	1.00	6.64***	0.82	1.03	5.45***	1.59	0.99
.SPX	United States	5.66***	2.83***	-1.81*	5.87***	0.95	-2.35**	6.44***	3.74***	-0.10
.SSEC	China	9.43***	2.51**	0.63	9.3***	2.52**	0.83	13.29***	4.95***	1.76*
.SSMI	Switzerland	4.36***	1.16	1.46	4.36***	3.82***	1.48	2.63***	2.21**	1.47

Table 4.5: MOM t-test statistics.
In analogy to Table 4.2 with MOM instead of HAC t-test statistics. Again Panel A is based on 15-minute data, Panel B adjusts for intra-day seasonality and Panel C is based on 5-minute data.

RIC	Country	$\overline{r^2}$	$\overline{\widetilde{RV}}$	$\overline{r^2}/\overline{\widetilde{RV}}$	$\sqrt{\overline{r^2}}-\sqrt{\overline{\widetilde{RV}}}$	t_{HAC}	t_{MAC}	t_{MOM}	T
.AEX	Netherlands	3.50	3.45	1.02	0.14	0.95	0.82	0.79	4,567
.ATX	Austria	4.10	4.12	1.00	-0.04	-0.15	-0.13	0.70	4,179
.BFX	Belgium	1.63	1.61	1.01	0.09	0.74	0.86	0.61	5,275
.BSESN	India	4.21	4.26	0.99	-0.13	-0.97	-0.87	-0.81	4,948
.BVSP	Brazil	9.15	9.28	0.99	-0.22	-0.65	-0.96	0.13	4,728
.GDAXI	Germany	4.29	4.30	1.00	-0.02	-0.12	-0.11	-0.23	5,264
.FCHI	France	3.52	3.53	1.00	-0.02	-0.13	-0.15	-0.15	5,272
.FTMIB	Italy	6.53	6.30	1.04	0.44	1.00	0.69	1.84*	1,942
.FTSE	Great Britain	3.31	3.37	0.98	-0.16	-0.71	-0.91	-0.49	5,218
.GSPTSE	Canada	2.79	2.78	1.00	0.03	0.09	0.05	0.73	3,654
.IBEX	Spain	4.28	4.24	1.01	0.09	0.39	0.41	-0.14	5,188
.JALSH	South Africa	3.46	3.34	1.04	0.32	1.43	2.12**	1.41	3,216
.MCX	Russia	7.91	7.82	1.01	0.16	0.56	0.56	1.66*	3,892
.N225	Japan	3.57	3.40	1.05	0.46	2.09**	1.87*	2.65***	5,080
.OBX	Norway	5.62	5.57	1.01	0.13	0.37	1.40	0.87	2,679
.OMXC20	Denmark	4.05	4.29	0.94	-0.58	-1.88*	-1.83*	-0.81	2,821
.OMXHPI	Finland	4.33	4.20	1.03	0.30	1.08	1.58	2.10**	2,835
.OMXS30	Sweden	4.46	4.29	1.04	0.40	1.53	5.17***	2.69***	3,005
.PSI20	Portugal	2.80	2.72	1.03	0.24	1.01	1.10	1.21	4,866
.SPX	United States	3.37	3.32	1.02	0.15	0.86	1.00	0.28	5,183
.SSEC	China	6.15	6.02	1.02	0.26	1.76*	2.30**	2.67***	5,019
.SSMI	Switzerland	2.59	2.48	1.05	0.36	2.57**	3.41***	1.97**	4,770

Table 4.6: Statistics on average squared return and average corrected RV. In analogy to Table 4.1 with the difference that now the corrected RV measure $\widetilde{RV}_t = RV_t + CC_t$ instead of RV_t is considered. Again, average squared return, average corrected RV, and average deviation between the two are stated per annum in percent.

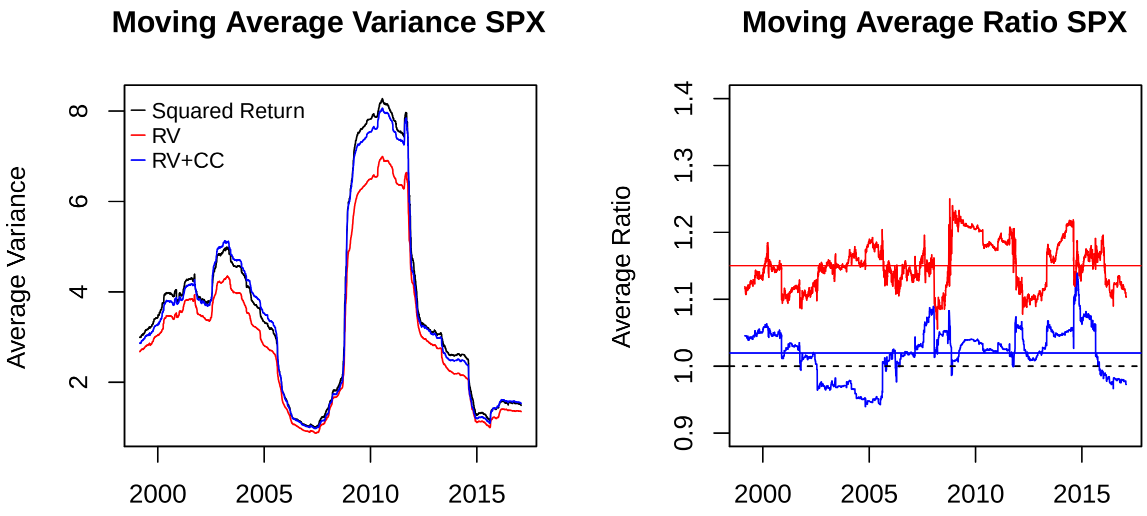

Figure 4.9: Corrected average variance time-series plots S&P 500. In analogy to Figure 4.2 but now including $\widetilde{RV} = RV_t + CC_t$. Again, both plots depict the moving average of the previous 750 observations.

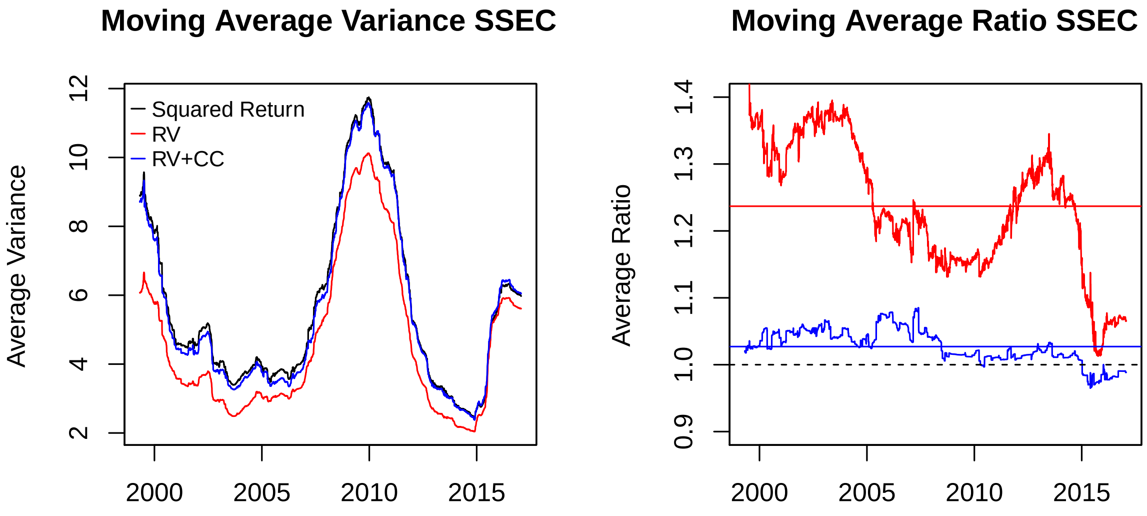

Figure 4.10: Corrected average variance time-series plots SSEC. In analogy to Figure 4.9 for the SSEC. Again, both plots depict the moving average of the previous 750 observations.

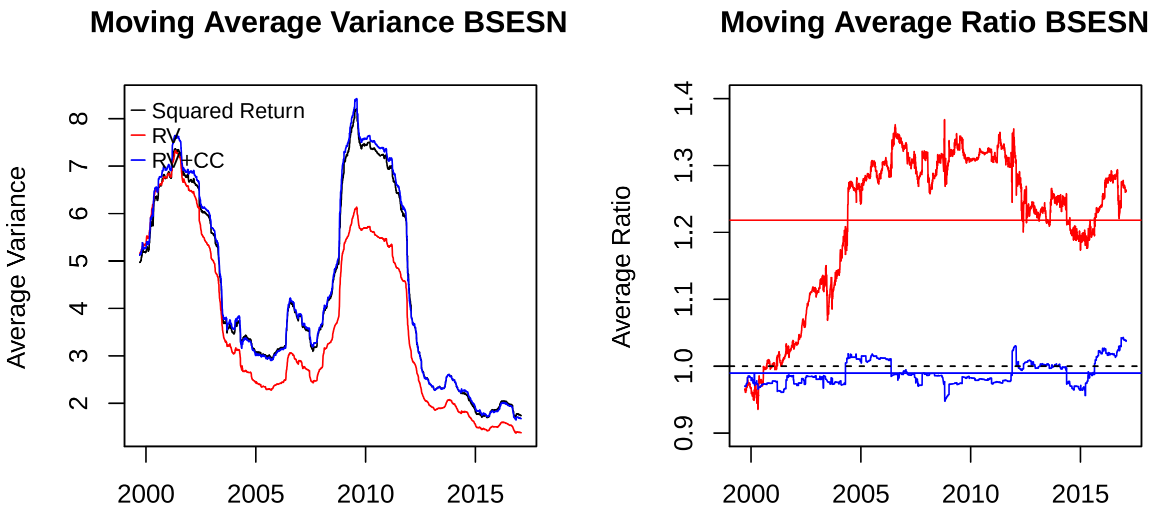

Figure 4.11: Corrected average variance time-series plots BSESN. In analogy to Figure 4.9 for the BSESN. Again, both plots depict the moving average of the previous 750 observations.

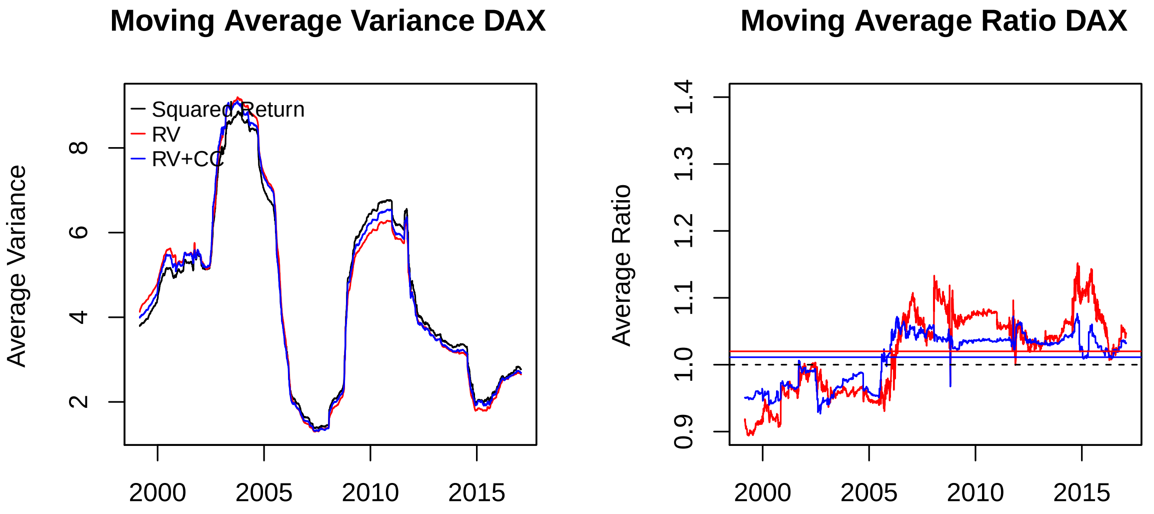

Figure 4.12: Corrected average variance time-series plots DAX. In analogy to Figure 4.9 for the DAX. Again, both plots depict the moving average of the previous 750 observations.

RIC	Country	$\overline{r^2}$	$\overline{RV}$	$\overline{r^2}/\overline{RV}$	$\sqrt{\overline{r^2}} - \sqrt{\overline{RV}}$	t_{HAC}	t_{MAC}	t_{MOM}	T
.AEX	Netherlands	2.01	1.93	1.04	0.28	1.28	0.83	1.58	4,567
.ATX	Austria	1.92	1.28	1.49	2.52	8.44***	2.79***	9.04***	4,179
.BFX	Belgium	0.88	0.85	1.05	0.21	1.46	0.65	1.76*	5,274
.BSESN	India	1.82	1.78	1.02	0.15	0.74	1.10	1.76*	4,947
.BVSP	Brazil	3.83	3.25	1.18	1.52	3.15***	3.28***	5.96***	4,728
.GDAXI	Germany	2.43	2.55	0.95	-0.40	-1.97**	-1.69*	-2.05**	5,263
.FCHI	France	2.09	2.18	0.96	-0.32	-1.68*	-0.65	-0.65	5,272
.FTMIB	Italy	3.20	2.89	1.11	0.91	2.71***	1.80*	2.96***	1,942
.FTSE	Great Britain	1.42	1.40	1.01	0.07	0.46	0.29	1.58	5,218
.GSPTSE	Canada	0.86	0.73	1.17	0.71	3.26***	5.87***	9.21***	3,654
.IBEX	Spain	2.39	2.24	1.07	0.49	2.63***	2.44**	2.90***	5,188
.JALSH	South Africa	1.30	1.14	1.14	0.74	4.12***	3.49***	5.20***	3,216
.MCX	Russia	4.05	3.30	1.23	1.95	5.74***	3.51***	7.97***	3,892
.N225	Japan	1.33	1.18	1.12	0.65	3.44***	3.39***	3.60***	5,080
.OBX	Norway	2.14	1.89	1.14	0.91	2.68***	4.51***	5.00***	2,678
.OMXC20	Denmark	1.64	1.42	1.15	0.89	3.51***	3.26***	3.93***	2,821
.OMXHPI	Finland	2.16	1.76	1.22	1.40	4.35***	2.83***	6.42***	2,835
.OMXS30	Sweden	1.97	1.86	1.06	0.39	1.52	0.71	1.19	3,005
.PSI20	Portugal	1.44	1.29	1.11	0.62	3.30***	1.66*	2.43**	4,866
.SPX	United States	1.18	1.22	0.97	-0.18	-1.16	-0.55	1.41	5,183
.SSEC	China	2.14	2.08	1.03	0.22	1.01	0.68	1.73*	5,019
.SSMI	Switzerland	1.46	1.37	1.07	0.39	1.51	1.67*	2.19**	4,770

Table 4.7: Statistics on average squared return and average RV – excluding first and last hour of the trading day.

In analogy to Table 4.1 with the difference that average RV and average squared return are now calculated without including the first and last hour of the trading day. Again, average squared return, average RV, and average deviation between the two are stated per annum in percent.

Chapter 5

The Memory of Beta

Co-authored with Fabian Hollstein, Marcel Prokopczuk, and Philipp Sibbertsen.

5.1 Introduction

In factor pricing models like the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966) or the arbitrage pricing theory (APT) (Ross, 1976), the drivers of expected returns are the stock's sensitivities to risk factors, i.e., beta factors. For many applications such as asset pricing, portfolio choice, capital budgeting, or risk management, the market beta is still the single most important factor. Indeed, Graham and Harvey (2001) document that chief financial officers of large U.S. companies primarily rely on one-factor market model cost-of-capital forecasts. In addition, Barber et al. (2016) and Berk and Van Binsbergen (2016) also show that investors mainly use the market model for capital allocation decisions. However, since beta factors are not directly observable, one needs to estimate them. For this purpose, researchers and practitioners alike typically use past information, i.e., employ time-series models.

The degree of memory is an important determinant of the characteristics of a time series. In an $I(0)$, or short-memory, process (e.g., AR(p) or ARMA(p,q)), the impact of shocks is short-lived and dies out quickly. On the other hand, for an $I(1)$, or difference-stationary, process like, for example, the random walk (RW), shocks persist infinitely. Thus, any change in a variable will have an impact on all future realizations. For an $I(d)$ process with $0 < d < 1$, shocks neither die out quickly nor persist infinitely but have a hyperbolically decaying impact. In this case, the current value of a variable depends on past shocks but the less so the further these shocks are past.

Researchers and practitioners estimate betas in several different ways. One approach is to use constant beta coefficients for the full sample (e.g., Fama and French, 1992). This relates to the most extreme $I(0)$ case possible. However, there is a strong consensus in the literature that betas vary over time. The usual approach to account for such time-variation is the use of rolling windows, where the most current estimate is taken as forecast for the next month (e.g., Fama and MacBeth, 1973; Frazzini and Pedersen, 2014). This approach inherently imposes infinite memory and resembles a random walk, i.e., presuming that the best forecast for the future beta is today's estimate.[15]

Numerous other studies employ explicit or implicit short-memory processes for modeling beta dynamics. These include, among others, AR(1) processes in Ang and Chen (2007) and Levi and Welch (2017), an AR(1) process with further latent and exogenous variables in Adrian and Franzoni (2009), and an ARMA(1,1) process in Pagan (1980). Blume (1971) imposes a joint AR(1) process for the entire beta cross-section. The implications of these differing approaches for the modeling of betas, though, vary substantially.

However, the literature on volatility modeling documents that volatility has clear long-memory properties (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996; Ding and Granger, 1996). It is thus natural to ask whether this is also true for beta. Andersen et al. (2006) tackle this issue and conclude that betas do not exhibit long memory. However, this conclusion is based on a relatively small sample of daily data and only considering tests on the autocorrelation functions. In this study, we use a large dataset of high-frequency data to comprehensively reexamine whether betas are best described by either (i) short-memory processes, (ii) difference-stationary processes, or (iii) whether beta time series instead show long-memory properties.

First, we use 30-minute high-frequency data to estimate each month the realized betas for each stock included in the S&P 500 during the 1996–2015 sample period. Next, we estimate the memory of realized beta using the two-step exact local Whittle (2ELW) estimator by Shimotsu and Phillips (2005) and Shimotsu (2010).[16] We find that betas show consistent long-memory properties. The average estimate for the long-memory parameter d is 0.56. Adjusting for potential structural breaks in the beta series decreases the average d only modestly, to 0.52. For virtually all stocks, the statistical tests clearly reject both the short-memory ($d = 0$) and difference-stationary ($d = 1$) alternatives. Thus, the vast majority of previous studies substantially misspecifies the properties of the beta time series.

Our findings differ considerably from those of Andersen et al. (2006). There are several causes for this difference. First, our study has a substantially broader focus: we consider more than 800 stocks. Second, we use high-frequency data to estimate beta factors. This enables us to obtain more precise and less noisy estimates of beta (see also Hollstein et al., 2019a). Noise in the beta series of Andersen et al. (2006) could potentially lead to a downward bias in memory estimates as found by Deo and Hurvich (2001) and Arteche (2004). In contrast, we find that changing the bandwidth in the 2ELW estimation leads to similar estimates of the memory parameter. This suggests that the noise in our beta series is small. Third, using simulations, we show that for small samples, tests based on autocorrelation functions, as opposed to direct estimates with the 2ELW estimator, have little power to detect *true* long memory.

---

[15] Black et al. (1992), for example, explicitly model beta dynamics with a random walk.

[16] In simulations, we show that, as opposed to the 2ELW estimator, the alternative, theoretically noise or structural break robust, estimators of Hurvich et al. (2005), Iacone (2010), Frederiksen et al. (2012), and Hou and Perron (2014) suffer from material biases in finite samples. Therefore, for our main analysis, we use the 2ELW estimator.

Having documented that betas exhibit distinct long-memory properties, we next examine the implications of this result for forecasting. Beta forecasts are of paramount importance for many applications in finance. For example, capital allocation decisions, portfolio risk management (Daniel et al., 2018), as well as firms' cost of capital (Levi and Welch, 2017) strongly hinge on precise forecasts of betas. We find that a FI model, which uses only the long-memory properties for beta forecasting, yields the lowest root mean squared error (RMSE). The FI model significantly outperforms both the short-memory (AR(p), ARMA(p,q)) and difference-stationary (RW) alternatives for a substantial fraction of the stocks. A full-fledged ARFIMA(p,d,q) alternative performs somewhat worse than the pure FI model but better than the AR, ARMA, and RW models. We further show that the outperformance of the FI model over alternatives gets stronger for longer-horizon beta forecasts up to one year. Thus, incorporating the long-memory property is highly important for obtaining good beta forecasts.

In a next step, we examine which firm characteristics are associated with different degrees of memory in betas. We find that higher memory in beta is to some extent linked with higher levels of a stock's beta, book-to-market ratio, and leverage. In addition, stocks with high memory typically have lower market capitalization. Furthermore, we find substantial industry effects: stocks in the Energy and Manufacturing industries have comparably high memory in beta, while stocks in the Durables, HiTec Equipment, and Wholesale industries tend to have relatively low memory in beta. The latter industries are and have been particularly prone to disruptions and creative destruction. The somewhat shorter memory of the betas of these stocks is thus consistent with what one might intuitively expect. One should note, however, that these still exhibit long memory: past shocks also have a long-lasting impact on their betas.

Finally, we document that for high-momentum stocks, liquid stocks, and those with high short interest or high idiosyncratic volatility, using a RW model instead of the FI model yields particularly high errors. On the other hand, for high-beta stocks, illiquid stocks, and those with low short interest, it is most harmful to use an ARMA(p,q) model instead of the FI model.

We run a battery of tests to document the robustness of these results. First, we show that the FI model also outperforms its competitors when using hedging errors instead of the RMSE to evaluate the forecasts. Second, we also document long-memory properties of betas for the entire Center for Research in Security Prices (CRSP) sample. For this substantially larger sample and a much longer time period, we find that the FI model also outperforms all alternatives. Third, we estimate the short-memory and difference-stationary models in a state-space framework. In addition, we consider the Vasicek (1973) and Levi and Welch (2017) estimators, a heterogeneous AR (HAR) model, as well as a FI model, for which we set the long-memory parameter d to 0.5 instead of estimating it. We find that all alternative models underperform the FI model. Instead, the FI(0.5) model performs even somewhat better than the standard FI model. Fourth, we use the alternative estimator of the d parameter of Geweke and Porter-Hudak (1983) and obtain very similar results. Finally, we consider alternative intra-day sampling frequencies, alternative rolling estimation windows, and bandwidths. Our conclusions remain unchanged.

Our paper contributes to the literature on beta estimation. Hollstein and Prokopczuk (2016) consider both $I(0)$ and $I(1)$ beta forecasts but do not take into account models that account for long memory. Further contributions that deal with beta estimation include Buss and Vilkov (2012), Levi and Welch (2017), and Hollstein et al. (2019b). We complement these studies by explicitly considering long-memory processes to make beta forecasts. To the best of our knowledge, we are the first to show that forecasting beta with long-memory models yields superior forecasts compared to both $I(0)$ and $I(1)$ models.

What is the underlying economic mechanism that creates long memory in betas? Müller et al. (1993), LeBaron (2001), LeBaron (2006), Alfarano and Lux (2007), and Corsi (2009) propose variations of models with heterogeneous agents. The primary mechanism is typically that agents incur heterogeneous planning and investment horizons. The interaction of these agents creates long memory in volatility. Kamara et al. (2016) and Brennan and Zhang (2018) put forward and examine a similar idea for systematic risk factors.

We organize the remainder of this paper as follows. Section 5.2 introduces the data and presents summary statistics. We present results about the long memory in betas in Section 5.3. In Section 5.4, we examine the impact of our findings for the forecasting of betas. We study the economic implications of our findings in Section 5.5. Section 5.6 contains several further analyses and robustness checks. In Section 5.7, we draw conclusions.

5.2 Data and Methodology

5.2.1 Data

Our dataset covers U.S. stocks for the sample period from January 1996 to December 2015. Following Bollerslev et al. (2016), for our main analysis we restrict our attention to stocks that are part of the S&P 500 index at least once during our sample period. We collect high-frequency price data from the Thomson Reuters Tick History (TRTH) database. On average, the stocks for which high-frequency data are available represent 79 percent of the entire market capitalization of ordinary common U.S. stocks.

In order to process the final high-frequency dataset, we follow the data-cleaning steps outlined in Barndorff-Nielsen et al. (2009). First, we use only data with a time stamp during the exchange trading hours, i.e., between 9:30AM and 4:00PM Eastern Standard Time. Second, we remove recording errors in prices. To be more specific, we filter out prices that differ by more than 10 mean absolute deviations from a rolling centered median of 50 observations. Afterwards, we assign prices to every 30-minute interval using the most recent entry recorded that occurred at most one day before. Finally, we follow Bollerslev et al. (2016) and supplement the TRTH data with data on stock splits and distributions from CRSP to adjust the TRTH overnight returns.

5.2.2 Beta Estimation

Following Andersen et al. (2006), we use the realized beta estimator to obtain betas. We utilize intra-day high-frequency log-returns, sampled at intervals of 30 minutes to estimate[17]

$$\beta_{i,t} = \frac{\sum_{\tau=1}^{O} r_{i,\tau} r_{M,\tau}}{\sum_{\tau=1}^{O} r_{M,\tau}^2},$$

where O is the number of high-frequency return observations during the time period under investigation. $\beta_{i,t}$ is the beta estimate for asset i using data until the end of month t. $r_{i,\tau}$ and $r_{M,\tau}$ refer to the return of asset i and the market return at time τ, respectively. For the main analysis, we consider monthly realized beta estimates.

The choice of sampling frequency underlies a delicate trade-off (Patton and Verardo, 2012). On the one hand, using low-frequency data could result in noisy estimates of beta (Andersen et al., 2005). On the other hand, pushing the analysis to a very high frequency introduces a number of microstructure issues (Scholes and Williams, 1977; Epps, 1979). To balance these effects, we focus our main analysis on a sampling frequency of 30 minutes. In Section 5.6, we show that our main results are robust to the choice of sampling frequency.

5.2.3 Long-Memory Estimation

Our estimation of the order of integration d of a beta time series relies on the 2ELW estimator as introduced in Shimotsu and Phillips (2005) and Shimotsu (2010). Given a time series y_t, we can obtain this estimator as follows. We first calculate the tapered local Whittle estimator by Velasco (1999) which is obtained by

---

[17] Note that this formula resembles the expanded formula for the variance, while neglecting both the drift term and the risk-free rate. Andersen et al. (2006) note that the effect of the drift term vanishes as the sampling frequency increases, which effectively "annihilates" the mean. Empirically, for example, the average 30-minute return of the S&P 500 index amounts to 0.0017 percent. The average daily risk-free interest rate during our sample period amounts to 0.01 percent, which is equivalent to an average risk-free rate as low as 0.0007 percent over 30-minute intervals. Thus, at this sampling frequency both the drift and the risk-free rate can indeed be neglected.

$$\hat{d}_{Vel} = \underset{d \in (-1/2,2)}{\arg\min} \left[\log \left(\frac{3}{m} \sum_{j}^{m} \lambda_j^{2d} I_y^*(\lambda_j) \right) - 2d \frac{3}{m} \sum_{j}^{m} \log \lambda_j \right].$$

Here, $I_y^*(\lambda_j)$ is the cosine-bell tapered periodogram of the series at frequency λ_j with $j = 3, 6, \ldots, m$. Furthermore, m is the bandwidth parameter which determines the number of frequencies used for estimation. Larger m imply less variance of the estimates but then the estimator will be biased in case the underlying process exhibits short-run dependencies. We follow Shimotsu (2010) and consider $m = T^{0.7}$ in the following and report qualitatively similar results for alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$ as a robustness check in Table 5.14 in the Appendix.

Under some mild assumptions, this estimator is consistent and asymptotically normal for $d \in (-1/2, 2)$. However, as the estimator considers only every third frequency of the periodogram its variance exceeds that of the the standard local Whittle estimator by Robinson (1995). To account for this, the estimate is adjusted in the second step using

$$\hat{d}_{2ELW} = \hat{d}_{Vel} - \frac{L'(\hat{d}_{Vel})}{L''(\hat{d}_{Vel})}, \quad \text{where}$$

$$L(d) = \log \left(\frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d y - \mu(d)}(\lambda_j) \right) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j .$$

Here, $I_{\Delta^d y-\mu(d)}(\lambda_j)$ is the periodogram of the demeaned series. Since the arithmetic mean $\bar{y}$ is inconsistent for $d > 1/2$, Shimotsu (2010) suggests using $\mu(d) = \bar{y}$ if $d < 1/2$, $\mu(d) = y_1$ if $d > 3/4$, and $\mu(d) = \omega(d)\bar{y} + (1 - \omega(d))y_1$ with $\omega(d) = 1/2[1 + cos(4\pi d)]$ if $d \in [1/2, 3/4]$. This two-step estimator then has the same limiting variance as the standard local Whittle estimator while being consistent and asymptotically normally distributed for $d \in (-1/2, 2)$. Consequently, the 2ELW estimator can be used to distinguish short-memory series ($d = 0$), stationary long-memory series ($0 < d < 1/2$), nonstationary long-memory series ($1/2 < d < 1$), and difference-stationary series ($d = 1$) such as the random walk. This is an advantage over the standard local Whittle estimator, which can only be used for inference for $-1/2 < d < 3/4$ as it has a non-normal limit distribution otherwise.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.561	0.112	0.998	0.999	0.523	0.136	0.993	0.999

Table 5.1: Average memory parameter estimates – realized beta. Reported are average estimates of the memory parameter of realized beta across all stocks ($\bar{\hat{d}}_i$) using the 2ELW estimator of Shimotsu and Phillips (2005) and Shimotsu (2010). Additionally, $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d = 0$ and $d = 1$, respectively, are rejected at the ten percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle and Moulines (2000).

5.3 Long Memory in Beta

5.3.1 Estimation Results

The left panel of Table 5.1 shows the average estimated d across the realized beta series of all stocks with more than 100 monthly observations (for $N = 823$ stocks we have sufficient data) using the 2ELW estimator. Additionally, we present the standard deviation of the estimates across stocks and the relative frequency with which the d estimates of different stocks are significantly different from 0 and 1, respectively, at the ten percent level. To illustrate the variation in d across stocks, Figure 5.1 additionally plots the corresponding density of the estimates.

Table 5.1 reveals that the average d is approximately 0.56 and Figure 5.1 shows that while there is some variation across stocks, most of them have a d between 0.4 and 0.8. A formal statistical test also confirms that for more than 99 percent of the stocks it holds that $0 < d < 1$ at the ten percent level. At the one percent level this is still true for more than 98 percent of the stocks.

As a firm's business may change over time, some of the considered companies could exhibit a structural break in the realized beta series. When the underlying process is stationary, i.e., $d < 1/2$, but exhibits structural breaks in mean, then the local Whittle estimator and therefore also the 2ELW estimator is positively biased (e.g., Diebold and Inoue, 2001; Granger and Hyung, 2004). One way to account for this would be to use the estimators by Iacone (2010) or Hou and Perron (2014), as these remain consistent when structural breaks are present. However, as we also show in simulations in the Appendix, these are negatively biased for sample sizes smaller than 500, making them unsuitable for our application. To examine the robustness of our results, we therefore use an alternative two-step procedure. We first estimate the points at which the series exhibit structural

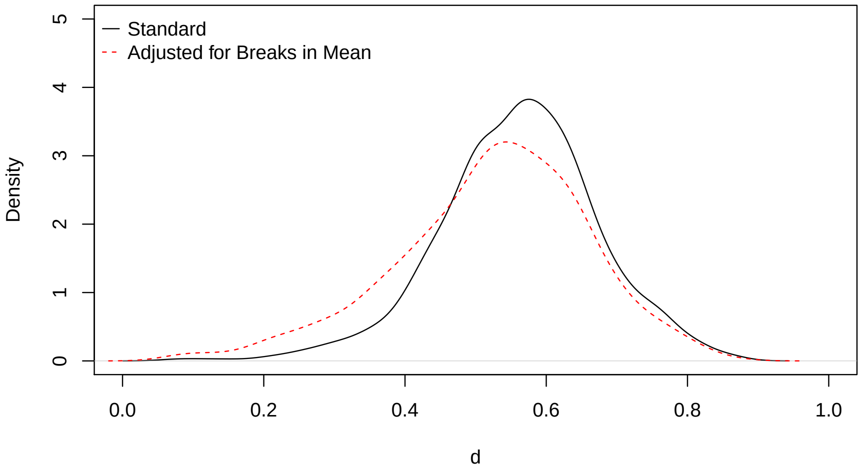

Figure 5.1: Density plots for memory parameter estimates. Density plot showing the distribution of the estimated beta memory parameters across stocks. For estimation we consider the Gaussian kernel and choose the bandwidth according to Silverman (1986).

breaks in mean using the procedure by Lavielle and Moulines (2000) and then apply the 2ELW estimator estimator for the cleaned series.[18]

The results are shown in the right panel of Table 5.1 and are visualized by the dashed line in Figure 5.1. We find that the average $\hat{d}$ decreases slightly to 0.523, implying that some stocks do indeed exhibit structural breaks in their betas time series. However, the reduction is small and for more than 99 percent of the stocks the null that $d = 0$ can still be rejected.

Our results stand in contrast to those by Andersen et al. (2006), who argue that betas are integrated of a much smaller order, often even $I(0)$. There are two main reasons for this difference in results.

First, Andersen et al. (2006) base their analysis on daily data which leads to noisy estimates of beta, as also acknowledged by the authors themselves. Deo and Hurvich (2001) and Arteche (2004) show that for perturbed series any inference on the order of integration is biased such that the series appear to be less integrated. Our beta estimates based on intra-day observations, on the other hand, are less noisy, implying that the true order of integration can be better detected. To further illustrate this, one might think of comparing the 2ELW estimates to estimates made by noise robust estimators such as those of Sun and Phillips (2003) or Frederiksen et al. (2012). However, these are positively biased when the sample size is smaller than 500, making them inappropriate for our setup. As an alternative we show in Table 5.14 of the Appendix that changing the bandwidth m in the 2ELW estimation leads to similar estimates of d. As demonstrated by Hurvich et al. (2005), this would not be the case if the series were seriously perturbed.

---

[18] Bai and Perron (1998) and Bai and Perron (2003) suggest estimating breaks in mean by minimizing the residual sum of squares (RSS) of $\beta_t = \mu_s + e_t$, where μ_s is the mean in segment s with $s = 1, \ldots, S$ and S determined by means of the BIC. Lavielle and Moulines (2000) extend this approach by adding a penalty term to the BIC criterion which is then $BIC = RSS(S) + 4S\log(T)T^{2d-1}$. This leads to a more parsimonious break point selection, as for long-memory time series the standard procedure indicates too many break points.

Second, Andersen et al. (2006) rely on graphical investigation of the first 36 autocorrelations instead of consistent estimation of the memory parameter. Particularly in small samples (Andersen et al., 2006 consider $T = 148$) this type of inference may lead to false conclusions. We illustrate this by means of a small simulation study for which we report the results in Table 5.13 in the Appendix. We simulate fractionally integrated noise, i.e., $(1 - B)^d y_t = \epsilon_t$ with B being the backshift operator, for memory parameters of $d = 0.2, 0.4, 0.6$ and sample sizes of $T = 100, 148, 240, 1000$. The table reveals that on average only 24 percent of the first 36 autocorrelations of an $I(0.4)$ process with $T = 148$ are significantly larger than zero. From this result one might falsely infer that the series exhibit short memory. In contrast, the simulation results shows that the 2ELW estimator is also unbiased in small samples, implying that the correct order of integration can be detected. For further details on the simulation setup and results we refer to the Appendix.[19]

We therefore conclude that realized betas are highly persistent and are best described by either pure long-memory processes or a combination of break and long-memory process.

5.3.2 Beta Decomposition

Since beta is actually a combination of different components, it seems interesting to investigate which of these drives the persistence. For that purpose, consider the following decomposition

$$\beta_{i,t} = \sigma_{i,M,t}\sigma_{M,t}^{-2} = \rho_{i,M,t}\sigma_{i,t}\sigma_{M,t}\sigma_{M,t}^{-2} = \rho_{i,M,t}\sigma_{i,t}\sigma_{M,t}^{-1}, \tag{5.1}$$

where $\sigma_{i,M,t}$ is the realized covariance of asset i and the market M at time t, $\rho_{i,M,t}$ their realized correlation, and $\sigma_{i,t}$ is the realized volatility. Consequently, Equation (5.1) shows that the realized beta series evolves as the product of realized correlation, realized volatility, and the inverse of realized market volatility. Leschinski (2017) shows theoretically that the products of stationary long-memory series with non-zero mean are integrated with the maximum memory of the series. This would mean that one of the components needs to exhibit the same degree of memory as realized beta, while the others could exhibit

---

[19] Table 5.13 also presents results for the estimators by Sun and Phillips (2003), Iacone (2010), Frederiksen et al. (2012), and Hou and Perron (2014) to validate our claim that these are biased in small samples. Additionally, the table presents results for the log-periodogram estimator which we consider in Section 5.6 as a robustness check. This estimator is also unbiased but exhibits a larger variance than the 2ELW estimator.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
$\rho_{i,M}$	0.559	0.096	1.000	0.999	0.557	0.099	1.000	1.000
σ_i	0.594	0.142	0.996	0.977	0.594	0.142	0.996	0.977
σ_M^{-1}	0.562	-	1.000	1.000	0.561	-	1.000	1.000

Table 5.2: Average memory parameter estimates – realized correlation, realized volatility, inverse of realized market volatility.
Reported are average estimates of the memory parameter of realized correlation (Fisher-transformed) and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the 2ELW estimator of Shimotsu and Phillips (2005) and Shimotsu (2010). $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d = 0$ and $d = 1$, respectively, are rejected at the ten percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle and Moulines (2000).

a smaller d, even $d = 0$. However, for approximately 70 percent of the stocks it holds that $d > 1/2$, meaning that the beta series exhibit nonstationary long memory. In these cases, it is theoretically unclear how products of such series behave. We therefore also estimate the order of integration of realized correlation, realized volatility, and the inverse of realized market volatility using the 2ELW estimator.[20]

The results are shown in Table 5.2.[21] Again, we consider the possibility of structural breaks and also report results when adjusting for these. The realized correlation and the inverse of realized market volatility on average exhibit a d of approximately 0.56, while the d of realized volatility is even slightly higher on average, with 0.59. Again, tests indicate that for almost all stocks the order of integration is different from 0 and 1 for all three components.

When adjusting for structural breaks, the d of the realized correlation decreases slightly, while the d of realized volatility does not. Consequently, it is rather breaks in realized correlation than breaks in volatility that drive the breaks observed in the realized betas. When comparing the actual estimate of d to the estimate of the memory of the realized beta series, it can be seen that all three components exhibit a slightly higher degree of persistence. Thus, it seems that no single component, but rather all of them, drives the persistence in realized betas.

---

[20] We obtain the realized volatility for stock i and the market ($i = M$) as $\sigma_{i,t} = \sqrt{\sum_{\tau=1}^{O} r_{i,\tau}^2}$, the realized covariance as $\sigma_{i,M,t} = \sum_{\tau=1}^{O} r_{i,\tau} r_{M,\tau}$, and the realized correlation as $\rho_{i,M,t} = \frac{\sigma_{i,M,t}}{\sigma_{i,t}\sigma_{M,t}}$.

[21] We Fisher-transform the realized correlation series to guarantee that there is no bias due to the restricted character of the variable. If we use the original series, the results are similar.

5.4 Forecasting

Having shown that betas have consistent long-memory properties, a natural next question to ask is: Can we leverage the long-memory properties in betas to make better forecasts? How big are the errors when inaccurately imposing $I(0)$ or $I(1)$ dynamics for forecasting betas? Is accounting for long memory more important for long-term beta forecasts? In this section, we set out to answer these questions. For this purpose, we compare pseudo out-of-sample forecasts for the realized beta series of models accounting for the long-memory characteristics with those for short-memory and difference-stationary processes.

5.4.1 Forecasting Methodology

For forecasting using long-memory models, we follow the approach proposed by Hassler and Pohle (2019). Given the estimated order of integration of a series, we first remove the persistence by filtering. Then, we calculate the mean of the series. In a next step, we forecast the filtered data accounting for potential short-run dependencies. Finally, we reintegrate the series to obtain a forecast.

In more detail, given the first T betas of stock i, we first compute the $\hat{d}$-th difference

$$\Delta^{\hat{d}_i}\beta_{i,t} = (1-L)^{\hat{d}_i}\beta_{i,t} = \sum_{j=0}^{t-1}\binom{\hat{d}_i}{j}(-1)^j\beta_{i,t-j}, \text{ with } t = 1,\dots,T,$$

where $\hat{d}_i$ is the estimate of the 2ELW estimator with a bandwidth of $m = T^{0.7}$. Again, we report qualitatively similar results for $m = T^{0.65}$ and $m = T^{0.75}$ in Table 5.22 of the Appendix.

We then set out to calculate the mean μ_i of the series, which is complicated by the long-memory characteristics. As discussed above, the arithmetic mean cannot be considered for nonstationary long-memory series as it does not exhibit a finite variance. We therefore consider the approach by Robinson (1994) to estimate μ_i. For this purpose, we perform the following regression

$$\Delta^{\hat{d}_i}\beta_{i,t} = \psi_{i,t}\mu_i + \eta_{i,t}, \text{ with } \psi_{i,t} = \sum_{j=0}^{t-1}\binom{\hat{d}_i}{j}(-1)^j,$$

where $\eta_{i,t}$ is the error term that contains possible short-run dynamics. This allows us to calculate the residuals

$$\varepsilon_{i,t} = \Delta^{\hat{d}_i}\beta_{i,t} - \psi_{i,t}\hat{\mu}_i,$$

which are not fractionally integrated any longer but might exhibit short-run dependencies. We can optionally account for these using an ARMA(p,q) model

$$\varepsilon_{i,t} = \phi_{i,1}\varepsilon_{i,t-1} + \ldots + \phi_{i,p}\varepsilon_{i,t-p} + \theta_{i,1}\zeta_{i,t-1} + \ldots + \theta_{i,q}\zeta_{i,t-q} + \zeta_{i,t},$$

with $t = \max\{p,q\} + 1, \ldots, T$. Here, $\zeta_{i,t}$ is the mean-zero error term and p and q are determined by means of the BIC with a maximum lag length of $12[(T/100)^{0.25}]$. This allows us to forecast the residuals h steps ahead

$$\hat{\varepsilon}_{i,T+h} = \hat{\phi}_{i,1}\hat{\varepsilon}_{i,T+h-1} + \ldots + \hat{\phi}_{i,p}\hat{\varepsilon}_{i,T+h-p} + \hat{\theta}_{i,1}\hat{\zeta}_{i,T+h-1} + \ldots + \hat{\theta}_{i,q}\hat{\zeta}_{i,T+h-q}.$$

For $\hat{\varepsilon}_{i,T+h}$, the hat indicates that it is a forecast and h denotes the forecast window in months. In a case without short-run dependencies we simply set $\hat{\varepsilon}_{i,T+h} = 0$. We then reintegrate the series to account for the long-memory characteristics by calculating $\hat{Z}_{i,t} = \Delta^{-\hat{d}_i}\hat{\varepsilon}_{i,t}$ for $t = 1, \ldots, T+h$, respectively $t = \max\{p,q\} + 1, \ldots, T+h$. Forecasts of the original sequence then evolve as

$$\hat{\beta}_{i,T+h} = \mu_i + \hat{Z}_{i,T+h}.$$

This approach allows us to forecast stationary as well as nonstationary series while also accounting for potential short-run dynamics. We denote the model with short-run components by ARFIMA in the following, while the model without short-run dependencies is referred to as FI.

As difference-stationary and short-memory competitor models, we consider the random walk model, for which $\hat{\beta}_{T+h} = \beta_T$, as well as AR(p) and ARMA(p,q) models, respectively. We estimate the latter models based on

$$\beta_{i,t} = a_i + \phi_{i,1}\beta_{i,t-1} + \ldots + \phi_{i,p}\beta_{i,t-p} + \theta_{i,1}e_{i,t-1} + \ldots + \theta_{i,q}e_{i,t-q} + e_{i,t},$$

with $t = \max\{p,q\} + 1, \ldots, T$. For the AR model we set $\theta_{i,1} = \ldots = \theta_{i,q} = 0$. Again, we choose p and q according to the BIC with a maximum lag length of $12[(T/100)^{0.25}]$.

To examine the out-of-sample forecast accuracy of the different approaches, we perform the analysis using the RMSE, a loss function commonly applied in the literature

$$\text{RMSE}_{i,h} = \sqrt{\frac{1}{\Upsilon}\sum_{T=1}^{\Upsilon}(\beta_{i,T+h} - \hat{\beta}_{i,T+h})^2},$$

where Υ is the number of out-of-sample observations of realized and predicted betas of one stock. $\beta_{i,T+h}$ is the realized beta and $\hat{\beta}_{i,T+h}$ denotes a beta forecast. The RMSE criterion is suitable since it is robust to the presence of (mean-zero) noise in the evaluation

	RW	AR	ARMA	FI	ARFIMA
RMSE	0.3149	0.2942	0.2875	0.2792	0.2800
Best	4	24	123	374	164
vs RW	0	272	344	560	518
vs AR	3	0	177	305	307
vs ARMA	1	8	0	186	179
vs FI	0	3	14	0	7
vs ARFIMA	0	2	13	24	0
N	689	689	689	689	689

Table 5.3: Forecast results.
This table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

proxy, while other commonly employed loss functions are not (Patton, 2011).[22] We test for significance in forecast differences using the modified Diebold–Mariano (DM) test proposed by Harvey et al. (1997).

5.4.2 Forecast Results

The results of the various beta forecasts can be found in Table 5.3. We use a forecast window of one month and a rolling estimation window of 100 observations. The table 5.3 presents the average RMSE across all stocks and the number of times the model yields the lowest RMSE when forecasting the realized beta of a stock. The remainder of the table indicates the number of stocks for which the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. To allow for valid inference, we only consider stocks for which we have at least 50 forecasts ($N = 689$ stocks fulfill this criterion).

Table 5.3 reveals that the FI model performs best across all considered models. It has the lowest RMSE on average and is the model with the lowest RMSE for more than 54 percent of the stocks. Second best is the ARFIMA model, which is the best model for 24 percent of the stocks. The models that do not account for the long-memory characteristics of the beta time series, on the other hand, are only the most accurate for a combined 22 percent of the stocks. The outperformance of the long-memory models is often also

---

[22] The results when using the mean absolute error criterion instead of the RMSE are qualitatively similar.

	RW	AR	ARMA	FI	ARFIMA
Bias	0.0000	0.0035	0.0023	0.0009	0.0009
Inefficiency	0.0290	0.0083	0.0062	0.0047	0.0048
Random Error	0.0910	0.0947	0.0922	0.0879	0.0887

Table 5.4: MSE decomposition.
This table shows the Mincer and Zarnowitz (1969) decomposition of the MSE as of Equation (5.2). The MSE is based on one-month forecasts of the realized beta series performed with a rolling estimation window of 100 observations. All numbers represent the average across all stocks for which at least 50 forecasts exist.

statistically significant. Compared to the RW forecasts, the FI forecasts are significantly better for 81 percent of the stocks; compared to AR and ARMA forecasts this number is 44 and 27 percent, respectively. On the other hand, the forecasts by the RW, AR, and ARMA models are almost never significantly better than those of the FI model. Consequently, we can conclude that accounting for the long-run dependence substantially improves forecasts for realized betas.

Our finding that the FI model yields significantly better forecasts than the RW model for almost all stocks has broad implications. Hollstein et al. (2019a) show that a RW model outperforms other predictors based on daily data as well as the Buss and Vilkov (2012) option-implied beta. Thus, the FI forecasts appear to be preferable not only to other time-series models but also to a broader set of potential estimators.[23]

To further investigate the causes of the differential forecast performance of the models, we follow Mincer and Zarnowitz (1969) and decompose the mean squared error (MSE) in the following fashion

$$\mathrm{MSE}_i \quad = \quad \underbrace{(\bar{\beta}_i - \bar{\hat{\beta}}_i)^2}_{\text{bias}} + \underbrace{(1-b_i)^2\sigma^2(\hat{\beta}_i)}_{\text{inefficiency}} + \underbrace{(1-\rho_i^2)\sigma^2(\beta_i)}_{\text{random error}} . \tag{5.2}$$

b_i is the slope coefficient of the regression $\beta_i = a_i + b_i\hat{\beta}_i + e_i$ and ρ_i^2 is the coefficient of determination of this regression. A bias indicates that the model is misspecified and the prediction is, on average, different from the realization. Inefficiency represents a tendency of an estimator to systematically yield positive forecast errors for low values and negative forecast errors for high values or vice versa. The remaining *random* forecast errors are unrelated to the predictions and realizations.

Table 5.4 presents the results of the MSE decomposition. Again, the numbers represent the averages across all considered stocks. We find that the RW model is on average unbiased but highly inefficient. Thus, particularly for high- and low-beta stocks, the RW approach generates sizable measurement errors. For the AR and ARMA models, the bias component is moderately larger than that of the RW model. Thus, these models appear to be somewhat misspecified. On the other hand, the inefficiency is dramatically smaller compared to the RW model. The random error component, which is the largest component for all models, is slightly higher for the AR and ARMA models than for the RW model.

---

[23] In untabulated results, we confirm this also empirically: the FI model outperforms estimators based on daily return data as well as option-implied estimators.

The models that account for long memory are approximately unbiased and yield a low inefficiency on average. In particular the FI model yields the lowest overall inefficiency component, which indicates that the model does well in particular for stocks with the most extreme betas. Finally, the FI model also yields the lowest random error. Both inefficiency and random error are slightly higher for the ARFIMA model. Thus, accounting for short-run dynamics in addition to long memory on average rather adds noise than helping to capture important parts of the variation in betas.

5.4.3 Longer Forecast Horizons

For many applications, such as capital budgeting decisions, managers typically plan over longer periods. Thus, they do not only need one-month beta forecasts but also forecasts over several months. Therefore, in this section, we also consider forecasts for three-month, six-month, and twelve-month horizons.

Table 5.5 presents the results for these forecast horizons, the table shows that the outperformance of the FI model forecasts persists and gets even stronger for horizons longer than one month. For all considered horizons, the forecasts by the FI model have the lowest average RMSE and are the best for more than half of the stocks. It can further be seen that the absolute difference in RMSE between FI forecasts and RW, AR, and ARMA forecasts increases in the forecast horizon. Consequently, it is even more beneficial to consider long-memory models when forecasting for horizons longer than one month.

Not only is the magnitude of the forecast error loss differentials larger, but also is this differential statistically significant more often for longer horizons. For the three-month forecast horizon, the FI forecasts are significantly better than the RW, AR, and ARMA forecasts for 84, 52, and 30 percent of the stocks, respectively. These numbers are only slightly smaller for the twelve-month horizons with 73, 39, and 27 percent, respectively. In addition, the forecasts of the FI model are still barely ever outperformed by forecasts of models that do not account for the long-run dependencies. This is the case for less than 3 percent of the stocks, independently of the forecasts horizon.

To summarize, using models that account for long-run dependencies, instead of short-memory or difference-stationary alternatives, does not only improve one-month forecasts but also forecasts for longer horizons up to one year.

	RW	AR	ARMA	FI	ARFIMA
	Three-Month Forecast Horizon				
RMSE	0.2918	0.2753	0.2589	0.2377	0.2417
Best	2	27	102	439	115
vs RW	0	184	303	572	617
vs AR	12	0	255	359	260
vs ARMA	4	7	0	210	130
vs FI	0	6	11	0	2
vs ARFIMA	0	9	27	149	0
N	685	685	685	685	685
	Six-Month Forecast Horizon				
RMSE	0.2955	0.2882	0.2616	0.2286	0.2367
Best	3	28	81	451	115
vs RW	0	146	274	554	614
vs AR	29	0	240	334	200
vs ARMA	3	2	0	193	77
vs FI	0	3	12	0	0
vs ARFIMA	0	16	41	208	0
N	678	678	678	678	678
	Twelve-Month Forecast Horizon				
RMSE	0.3075	0.3093	0.2766	0.2331	0.2451
Best	2	39	56	419	138
vs RW	0	139	236	478	588
vs AR	27	0	161	256	133
vs ARMA	11	4	0	177	61
vs FI	1	5	12	0	3
vs ARFIMA	0	22	45	212	0
N	654	654	654	654	654

Table 5.5: Forecast results – longer horizons.
In analogy to Table 5.3, this table illustrates the forecast performance of the models for three-, six-, and twelve-month beta forecasts from a rolling estimation window of 100 observations. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

5.5 Economic Implications

5.5.1 The Memory in Beta and Stock Characteristics

We continue the empirical analysis by examining to what extent the memory in beta factors relates to different firm characteristics. There are various candidate variables that might explain part of the difference in a stock's beta-memory. It is, for example, possible that the beta estimates of small and illiquid stocks contain more random noise, which has zero autocorrelation. Furthermore, it is possible that growth stocks, firms that invest more, or those that are most profitable change more frequently, which might make past shocks to their systematic risk die out more quickly. On the other hand, it is possible that current loser stocks or firms whose stocks experience high shorting activity are more prone to change their business models, which likely changes their systematic risk. Finally, there may be industry effects: for some industries, the business models, and with that the constituent firms' systematic risk, may be more persistent, while others experience more frequent changes.

For this analysis, we sort the stocks into five portfolios (P1 up to P5), based on their estimates for d. We do this at the end of each month using d-estimates based on a 100-month rolling window. For each portfolio we record the average of several firm characteristics at the end of that month. Subsequently, we examine whether there are systematic differences in the average firm characteristics of the different d-sorted portfolios. The variable definitions are found in the Appendix.

We present the results in Table 5.6. The quintile portfolio of the stocks with the lowest ds (P1) on average has a memory parameter of 0.35 while that of the stocks with the highest ds has a d of 0.73 on average. These averages are far away from both 0 and 1. This result is consistent with our previous finding that the betas of virtually all stocks have long-memory properties. Naturally, the difference between the memory parameters of portfolios 5 and 1 is highly statistically significant.

The second variable of interest is beta itself. We find that the average beta of high-beta-memory stocks is significantly higher than that of low-beta-memory stocks. The relation appears to be monotonic but overall economically not too strong. For the natural logarithm of a stock's market capitalization, we make an opposite observation. The stocks with the longest memory in beta appear to be somewhat smaller than those with the shortest memory in beta.

The average BtM ratio of the stocks in P1 is significantly smaller than that of P5. As the firms grow, the past shocks to their beta factors are essentially to those of different firms and their impact seems to die out more quickly. On the other hand, we detect no relation between the beta-memory and firms' investment, profitability, momentum, bid-ask spread, turnover, idiosyncratic volatility, idiosyncratic skewness, and short interest.

	P1	P2	P3	P4	P5	t-stat
d	0.3494	0.4789	0.5491	0.6173	0.7291	22.9
β	0.9326	0.9747	1.0219	1.0411	1.1182	3.02
log(Market Cap)	16.140	16.154	16.064	16.073	15.983	-2.40
BtM	0.4906	0.4972	0.4552	0.5334	0.5916	3.43
Investment	0.1073	0.0868	0.0880	0.1031	0.1033	-0.46
Profitability	0.2512	0.3398	-3.9065	-2.1864	-0.3117	-0.89
Momentum	0.1444	0.1252	0.1357	0.1368	0.1641	0.82
BAS	0.0008	0.0008	0.0008	0.0008	0.0009	1.52
Turnover	0.2339	0.2377	0.2444	0.2380	0.2503	0.90
iVol	0.0133	0.0134	0.0136	0.0134	0.0141	1.13
iSkew	0.1077	0.1169	0.1099	0.1048	0.1183	1.00
Short Interest	0.0397	0.0386	0.0402	0.0379	0.0374	-0.57
Leverage	0.5803	0.5775	0.5831	0.5759	0.6084	1.91
Age	34.330	36.799	36.839	36.302	35.992	0.75

Table 5.6: Portfolio sorts by estimated d.
At the end of each month, we sort the stocks in our sample based on the d-parameters estimated with the 2ELW estimator using a rolling window of 100 observations. Sorting the stocks into quintile portfolios, we save each portfolio's average of the firm characteristics at the end of the respective months. The main body of the table shows the average of the different firm characteristics over time. t-stat denotes the t-statistic of a test whether the firm characteristics of portfolio P5 and P1 are equal, with the standard errors calculated with the heteroscedasticity and autocorrelation robust approach by Andrews (1991), using a quadratic spectral density and data-driven bandwidth selection. Characteristics for which this difference is statistically significant at ten percent are printed in **bold**.

Low-d stocks on average exhibit lower leverage than high-d stocks. Age appears to be unrelated to the memory in betas.

Finally, we turn the focus on the stocks' industries. We present the results in Table 5.7. Stocks in the Energy and Manufacturing industries have on average the highest ds. Thus, these traditional industries tend to have higher persistence in their systematic risk than many others. For the Durables, HiTec Equipment, and Wholesale industries, the opposite holds true. These industries have in part been particularly prone to disruptions and creative destruction during the recent two decades. Thus, many of these firms and/or their market environment have experienced substantial changes and past shocks to their systematic risk die out more quickly.

In Table 5.15 of the Appendix, we also present the results of portfolios sorted on beta. We confirm that the relation of beta and beta-memory is on average positive but weakly so. There is very little difference in the d-parameters of the first three beta quintiles. Only for the two quintiles of the highest betas is the d estimate somewhat larger.

	$\hat{d}$	t-stat
Durables	0.5188	-1.70
Energy	0.6013	3.64
Healthcare	0.5529	-0.58
HiTec Equipment	0.5243	-3.85
Manufacturing	0.6079	4.77
NonDurables	0.5642	0.20
Other	0.5915	3.19
Telephone	0.5209	-1.40
Utilities	0.5489	-0.82
Wholesale	0.5036	-5.12

Table 5.7: Average memory estimate by industry.
The t-stat corresponds to t-statistics testing the null that the average d of the industry equals the average across all industries. Standard errors are calculated with the heteroscedasticity and autocorrelation robust approach by Andrews (1991), using a quadratic spectral density and data-driven bandwidth selection. Industries for which the average d is significantly higher or lower than this value at the ten percent level are printed in **bold**.

5.5.2 The Determinants of Forecast Errors

Having documented that accounting for long memory in betas substantially improves the forecasts, we next analyze for *which* stocks one makes the biggest mistakes when using short-memory processes or those that impose infinite memory. To that end, we regress the difference in absolute forecast errors on different firm characteristics. In more detail, we perform the following regressions

$$abs(\hat{\beta}_{i,t}^{RW} - \beta_{i,t}) - abs(\hat{\beta}_{i,t}^{FI} - \beta_{i,t}) = a + bx_{i,t-1} + e_{i,t},$$
$$abs(\hat{\beta}_{i,t}^{ARMA} - \beta_{i,t}) - abs(\hat{\beta}_{i,t}^{FI} - \beta_{i,t}) = a + bx_{i,t-1} + e_{i,t}.$$

Here, $\hat{\beta}_{i,t}$ are the forecasts made by the RW, ARMA, and FI models as presented in Section 5.4.2 and $x_{i,t-1}$ contains the set of explanatory variables lagged by one period.

We present the result for the forecast error differential between the RW model and the FI model in Table 5.8 and that between the ARMA(p,q) model and the FI model in Table 5.9.

Starting with the errors made when inadequately imposing a difference-stationary RW model in Table 5.8, we first obtain an economically large and statistically highly significant intercept term. This echoes our previous findings that the FI model yields substantially lower forecast errors on average than the RW model. Second, consistent with what one would intuitively expect, the slope coefficient on d is highly significantly negative. Thus, the higher the memory in betas, the less inadequate becomes the RW assumption. However, a one-standard-deviation increase in d from its average, while keeping all else

	coef	se	t-stat	p-value
Intercept	0.0253	0.0014	18.26	0.0000
d	-0.0034	0.0007	-5.28	0.0000
β	0.0013	0.0014	0.94	0.3200
log(Market Cap)	0.0006	0.0008	0.73	0.4730
BtM	0.0013	0.0037	0.34	0.7010
Investment	0.0000	0.0006	0.03	0.9790
Profitability	0.0000	0.0001	0.29	0.7200
Momentum	0.0021	0.0009	2.36	0.0170
BAS	-0.0071	0.0038	-1.84	0.0630
Turnover	-0.0039	0.0013	-2.93	0.0050
iVol	0.0127	0.0016	7.93	0.0000
iSkew	-0.0003	0.0006	-0.47	0.6420
Short Interest	0.0020	0.0012	1.71	0.0870
Leverage	0.0013	0.0008	1.71	0.0870
Age	-0.0005	0.0005	-1.01	0.3120
Durables	0.0008	0.0047	0.17	0.8620
Energy	-0.0030	0.0023	-1.33	0.1710
Healthcare	0.0010	0.0018	0.55	0.5600
HiTec Equipment	-0.0002	0.0021	-0.09	0.9160
Manufacturing	-0.0015	0.0017	-0.90	0.3540
NonDurables	0.0009	0.0023	0.40	0.6680
Telephone	-0.0069	0.0030	-2.28	0.0230
Utilities	0.0030	0.0232	0.13	0.6500
Wholesale	0.0007	0.0018	0.39	0.6590

Table 5.8: Forecast error regressions – RW.
In this table, we run regressions of the difference in absolute forecast errors from the RW and FI models on different firm characteristics variables. Firm characteristics (except for the dummy variables) are standardized to have zero mean and a volatility of one. The standard errors (se) are bootstrapped using the procedure of Cameron et al. (2008). t-stat and p-value denote the corresponding t-statistics and p-values, respectively. Characteristics which yield a statistically significant regression coefficient (coef) at ten percent are printed in **bold**.

equal, reduces the average forecast error differential (implied by the intercept term) by only one tenth.

The level of the idiosyncratic volatility has a positive effect on the forecast error differential. This effect is economically large: for an idiosyncratic volatility two-standard-deviation below the average, all else being equal, the forecast error of RW and FI processes are approximately the same. Thus, for high volatility stocks a random walk assumption appears to be less suitable.

We further observe that a one-standard-deviation increase in momentum, short interest, and leverage increases the forecast error differential by on average 0.21, 0.20, and 0.13 percentage points, respectively. It is well known that betas of stocks with extreme momen-

	coef	se	t-stat	p-value
Intercept	0.0103	0.0009	11.13	0.0000
d	-0.0005	0.0005	-0.98	0.3260
$\boldsymbol{\beta}$	0.0059	0.0011	5.41	0.0000
log(Market Cap)	-0.0004	0.0005	-0.65	0.5250
BtM	0.0003	0.0033	0.10	0.8960
Investment	0.0007	0.0007	1.00	0.3020
Profitability	-0.0002	0.0000	-4.90	0.0000
Momentum	-0.0005	0.0005	-1.03	0.3030
BAS	0.0040	0.0018	2.21	0.0310
Turnover	0.0016	0.0011	1.46	0.1490
iVol	-0.0004	0.0009	-0.41	0.6770
iSkew	-0.0002	0.0003	-0.59	0.5540
Short Interest	-0.0015	0.0007	-2.23	0.0260
Leverage	-0.0006	0.0006	-0.89	0.3590
Age	-0.0014	0.0004	-3.28	0.0010
Durables	-0.0025	0.0021	-1.21	0.2210
Energy	0.0017	0.0022	0.78	0.3870
Healthcare	0.0005	0.0017	0.27	0.7730
HiTec Equipment	0.0003	0.0014	0.23	0.8220
Manufacturing	-0.0010	0.0015	-0.65	0.5310
NonDurables	0.0014	0.0020	0.69	0.5210
Telephone	-0.0036	0.0043	-0.83	0.4780
Utilities	-0.0056	0.0059	-0.95	0.3760
Wholesale	0.0006	0.0013	0.47	0.6380

Table 5.9: Forecast error regressions – ARMA.
In this table, we run regressions of the difference in absolute forecast errors from the ARMA and FI models on different firm characteristics variables. Firm characteristics (except for the dummy variables) are standardized to have zero mean and a volatility of one. The standard errors (se) are bootstrapped using the procedure of Cameron et al. (2008). t-stat and p-value denote the corresponding t-statistics and p-values, respectively. Characteristics which yield a statistically significant regression coefficient (coef) at ten percent are printed in **bold**.

tum are highly time-varying (Grundy and Martin, 2001). Similarly, firms whose stocks exhibit very high short interest are also prone to substantial changes in systematic risk. For these stocks, in particular, it is therefore advisable to rely on the long-range dependencies when making forecasts. On the other hand, the bid-ask spread and the turnover have a negative impact on the loss differential. The beta of highly liquid stocks should therefore be predicted with long-memory models rather than the random walk.

In Table 5.9, we analyze the determinants of the ARMA and FI model error differentials. Consistent with our previous results, we also detect a strongly statistically significant intercept term of 0.0103. This intercept term is substantially smaller than that for the RW–FI forecast error differential.

The forecast error differential increases with beta. The impact of beta on these forecast error differentials is economically large: for betas two-standard-deviation below the average, all else being equal, the forecast error of ARMA and FI processes are approximately the same.

The profitability, short interest, and age all have a significant negative impact on the forecast error differential. Smaller firms and firms with higher short interest might be more prone to short-run changes in betas. Thus, the short-memory models perform a little less badly for these. The impact of each of these variables, however, is economically substantially smaller than that of the level of beta. The bid-ask spread has a positive impact on the forecast error differential. Thus, for the rather illiquid stocks the betas might contain more noise. The short-memory models might pick up too much of this noise to generate reliable forecasts.

5.6 Additional Analyses and Robustness

5.6.1 Hedging Errors

To account for the possibility that the ex-post realized betas are measured with error, we follow Liu et al. (2018) and examine the out-of-sample hedging errors of our main approaches. We compute the hedging error for each stock as

$$H_{i,T+1} \quad = \quad (r_{i,T+1} - r_{f,T+1}) - \hat{\beta}_{i,T+1}(r_{M,T+1} - r_{f,T+1}).$$

$r_{i,T+1}$ is the return of stock i in month $T+1$. $r_{f,T+1}$ and $r_{M,T+1}$ are the risk-free rate and the return on the market portfolio over the same horizon. We use one-month returns. $\hat{\beta}_{i,T+1}$ is the forecast for beta using data up to month T. Liu et al. (2018) show that under certain assumptions the hedging error variance ratio $\frac{\text{var}(H_{i,T+1})}{\text{var}(r_{M,T+1}-r_{f,T+1})}$ is approximately equal to the mean squared error relative to the *true* realized beta plus a term that is constant for all beta forecasts. We follow Liu et al. (2018) and estimate the variance ratios using rolling five-year windows to account for the possibility that the variances in the numerator and denominator change over time. We report the average ratio over time.

We present the results in Table 5.10. These are consistent with our previous results relying on the RMSE. The average hedging error of the FI-model forecasts is lowest. In particular, the average hedging error is significantly lower than both that of the difference-stationary RW and the short-memory ARMA models.

5.6.2 Entire CRSP Dataset

In our main analysis, based on the need to have high-frequency data for liquid instruments, we restrict our dataset to the S&P 500 firms and start in 1996. In this section, we examine

	RW	AR	ARMA	FI	ARFIMA
Mean	4.3890	4.3765	4.3699	4.3582	4.3640
ΔRW	0.0000	-0.0124^{**}	-0.0191^{***}	-0.0308^{***}	-0.0250^{***}
		(-2.0180)	(-2.9422)	(-5.1588)	(-3.4989)
ΔARMA	0.0191^{***}	0.0067	0.0000	-0.0117^{***}	-0.0059
	(2.9422)	(1.5468)		(-3.5983)	(-0.9233)

Table 5.10: Hedging errors.

This table presents the ratio of hedging error variances to the market variance for different approaches. For each stock, estimator, and month, we obtain the hedging error over the next month as $(r_{i,T+1} - r_{f,T+1}) - \hat{\beta}_{i,T+1}(r_{M,T+1} - r_{f,T+1})$. We estimate the hedging error and market variances using rolling five-year windows and use the average ratio over time. The table presents the average ratio of the hedging error variance to the market variance across all stocks. Additionally, ΔRW and ΔARMA report the differences between the hedging errors of RW and ARMA, respectively, and the other models. In parentheses, we present the robust Andrews (1991) t-statistics, using a quadratic spectral density and data-driven bandwidth selection, of a test for equal average hedging errors. *, **, and *** indicate significance at the ten percent, five percent, and one percent level, respectively.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.382	0.157	0.916	0.999	0.330	0.181	0.841	0.999

Table 5.11: Average memory parameter estimates – entire CRSP sample.

In analogy to Table 5.1, this table presents average estimates of the memory parameter of realized beta across all stocks ($\bar{\hat{d}}_i$) using the 2ELW estimator of Shimotsu and Phillips (2005) and Shimotsu (2010). The results are for the entire CRSP sample (3,153 stocks) and quarterly betas calculated from daily data. $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d = 0$ and $d = 1$, respectively, are rejected at the ten percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle and Moulines (2000).

whether the results found for this sample can also be generalized to a broader sample of stocks and for a longer sample period. We extend our dataset to consider the entire CRSP dataset starting from 1926. As intra-day observations are only available from 1996 onward, we calculate betas from daily returns. Since monthly beta estimates based on daily returns are too noisy, we follow Andersen et al. (2006) and consider quarterly estimates instead.[24]

---

[24] Since the zero-approximation to the risk-free rate becomes less reliable for daily returns, we deviate from the description in Equation (5.1) by using excess returns to estimate realized betas based on daily data.

	RW	AR	ARMA	FI	ARFIMA
RMSE	0.5654	0.4981	0.4881	0.4720	0.4724
Best	27	82	157	821	282
vs RW	0	721	802	1104	1052
vs AR	29	0	345	656	701
vs ARMA	15	38	0	458	480
vs FI	0	9	27	0	28
vs ARFIMA	1	7	25	27	0
N	1369	1369	1369	1369	1369

Table 5.12: Forecast results – entire CRSP sample.
In analogy to Table 5.3, this table illustrates the forecast performance of the models for quarterly beta forecasts, based on daily data, from a rolling estimation window of 100 observations. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

Table 5.11 shows the estimated order of integration of the series averaged across all stocks for which more than 100 observations are available ($N = 3,153$). Again, we present results when investigating the original series as well as when adjusting for structural breaks.

We find that the average d estimate decreases from 0.56 to 0.38 when considering the expanded sample of daily returns. This also holds when only considering the same stocks as in our main analysis, for which the average d estimate is now 0.36, and even when considering the same stocks and same time period as for our main analysis, where the average d is 0.42. Consequently, the observed reduction in d is mainly due to the change of the recording frequency and not to the expanded set of stocks and time period. As already discussed in Section 5.3, decreasing the recording frequency increases the level of noise in the realized beta time series. This then leads to a negative bias of the 2ELW estimator, which explains the reduction of the memory estimate.[25]

Even though the d estimates are negatively biased, more than 84 percent of the stocks still have a d that is significantly greater than zero. The forecast results displayed in Table 5.12 also echo this finding. It can be seen that the FI model still outperforms all models that do not account for the long-range dependencies. Its forecasts obtain the lowest average RMSE and are the most accurate for almost half of the stocks. Forecasts by RW,

---

[25] In Section 5.6.5, we explore this issue further by considering alternative intra-day sampling frequencies of 15-minutes and 75-minutes. There, we already find that increased noise in realized betas derived from 75-minute data biases the d estimates negatively.

AR, or ARMA models, on the other hand, are only the most accurate for a combined 19 percent of the stocks. The forecast of these models significantly outperform FI forecasts for less than three percent of the stocks in total. We should further note that as the FI model relies on a biased estimate of d, its performance would likely be even better if we accounted for this bias by adding a constant to each d estimate or even fixed d at a certain level for all stocks.

5.6.3 Alternative Models

Due to its great importance, there are numerous approaches and models to forecast beta. For the ease of presentation in our main analysis, we compare the performance of the long-memory models only to the performance of the most popular competitors, RW, AR, and ARMA. In this section we now consider other approaches that have been proposed in the literature.

Andersen et al. (2005) consider an AR(1) process to model beta in a state-space framework. Hollstein and Prokopczuk (2016) investigate the forecast performance of RW, AR(1), and ARMA(1,1) models in a state-space framework and find that the RW model performs somewhat better than the AR(1) and ARMA(1,1) models. Thus, in this section we also consider the forecasts from RW, AR(1), and ARMA(1,1) models when estimated as a state-space system. The measurement equation for all three models is

$$\beta_{i,t} = \tilde{\beta}_{i,t} + \xi_{i,t},$$

where $\tilde{\beta}_{i,t}$ is the unobserved *true* beta. It evolves according to one of the following transition equations for the different models

$$\begin{aligned}
\tilde{\beta}_{i,t}^{RW} &= \tilde{\beta}_{i,t-1} + v_{i,t}, \\
\tilde{\beta}_{i,t}^{AR} &= \gamma_i + \phi_i \tilde{\beta}_{i,t-1} + v_{i,t}, \text{ and} \\
\tilde{\beta}_{i,t}^{ARMA} &= \gamma_i + \phi_i \tilde{\beta}_{i,t-1} + \theta_i v_{i,t-1} + v_{i,t}.
\end{aligned}$$

We estimate those models using the Kalman filter (Pagan, 1980; Black et al., 1992) and then perform forecasts as for the standard models.

To the best of our knowledge, long-memory models in a state-space framework have only been investigated for the stationary $d < 0.5$ case (Chan and Palma, 1998; Dissanayake et al., 2016). As we investigate mostly nonstationary time series here, these models are likely inappropriate. As an alternative, we consider a ARFIMA model as before, but with the short-run dynamics now estimated with an ARMA(1,1) model in a state-space framework.

Another popular way to model and forecast long-memory time series is to use the HAR model by Corsi (2009). For the realized beta series, it evolves as

$$\beta_{i,t} = a_i + \phi_{1,i}\beta_{i,t-1} + \frac{\phi_{2,i}}{5}\sum_{j=1}^{5}\beta_{i,t-j} + \frac{\phi_{3,i}}{22}\sum_{j=1}^{22}\beta_{i,t-j} + e_{i,t},$$

where $e_{i,t}$ is a mean-zero error term. While the HAR model does not formally belong to the class of long-memory models, when applied to return volatility time series, this model has been shown to be able to reproduce long-memory patterns. We therefore also consider forecasts made by this model in the following.

Hassler and Pohle (2019) argue that although local Whittle-based approaches yield better results than other estimators, they still have a large variance. Moreover, as discussed above, the estimators are negatively biased when the degree of noise in the series becomes large. These considerations lead the authors to believe that it might be beneficial for forecasting to fix d at a certain value instead of estimating it. This eliminates estimation uncertainty, while the model is still able to capture the long-memory characteristics of the series. Based on the results of Section 5.3, we fix d to 0.5. We refer to this model as FI(0.5) in the following.

Finally, we also consider two popular shrinkage approaches. First, we apply the Vasicek (1973) estimator as modification to the RW forecast. We obtain a posterior beta by combining the RW forecast with a prior $(b_{j,t})$ in the following way

$$\beta_{i,t}^{\text{RWV}} = \frac{s_{b_{i,t}}^2}{\sigma_{\beta_{i,t}}^2 + s_{b_{i,t}}^2}\beta_{i,t} + \frac{\sigma_{\beta_{i,t}}^2}{\sigma_{\beta_{i,t}}^2 + s_{b_{i,t}}^2}b_{i,t}.$$

$\sigma_{\beta_{i,t}}^2$ and $s_{b_{i,t}}^2$ are the squared standard errors of the beta estimate and the prior, respectively. Hence, the degree of shrinkage depends on the relative precision of the historical estimate and the prior. As prior, we use the cross-sectional average beta, as suggested by Vasicek (1973).

Levi and Welch (2017) argue that a simple Vasicek (1973) shrinkage is not sufficient to create good forecasts for beta. They suggest further shrinkage using

$$\beta_{i,t}^{\text{RWLW}} = 0.75\beta_{i,t}^{\text{RWV}} + 0.25\beta_i^{\text{target}},$$

where β_i^{target} is set to 0.5 for the smallest market capitalization tercile, to 0.7 for the middle tercile, and to 0.9 for the highest market capitalization tercile. One has to bear in mind, though, that Levi and Welch (2017) optimize this double-shrinkage for betas based on daily return data. Since we rely on a highly liquid subset of stocks and use more precise estimates based on high-frequency data, it is likely that this approach does not work too well.

Table 5.16 of the Appendix shows the forecast results for these models and for comparison again the results by the FI model considered before. In line with the results by Hollstein and Prokopczuk (2016), we find that the performance of the RW model improves when estimated with a state-space framework as it on average now produces more accurate forecasts than AR and ARMA models. However, the models that account for long-range dependencies still perform substantially better and are more accurate for almost 80 percent of the stocks. The RMV model performs somewhat better than the simple RW model. The performance of the RVLW model, on the other hand, is very poor, as was expected.

Finally, it is noteworthy that the results for the FI(0.5) model are even slightly better than those by the FI model considered before. Thus, fixing d at 0.5 instead of using estimates appears to be a practical and well-performing approach for beta forecasting.

5.6.4 Alternative Long-Memory Estimator

We base our main analysis on the 2ELW estimator, as we believe it is the most suitable estimator in our setup. A popular alternative is the log-periodogram estimator by Geweke and Porter-Hudak (1983). Although the variance of log-periodogram-based approaches commonly exceeds that of local Whittle-based approaches, they are often considered due to their simplicity in application and calculation.

Table 5.17 of the Appendix shows the average estimate of d when using the log-periodogram estimator. While the average estimates of d are almost equal, the relative number of stocks for which d is significantly different from 0 and 1 decreases slightly due to the higher variance of the estimates. However, still more than 95 percent of the stocks exhibit significant long memory in beta. We can therefore conclude that with the log-periodogram estimator realized betas are also highly persistent.

Table 5.18 of the Appendix repeats the analysis of Table 5.3 and shows the forecast performance of the FI and ARFIMA model when estimating d using the log-periodogram estimator. For comparison, we also present the results for the RW, AR, and ARMA models. It can be seen that compared to the results using the 2ELW estimate, the performance of the FI and ARFIMA model slightly decreases, which is probably due to the higher variance of the estimates. However, the forecasts by the FI model still clearly outperform all forecasts by models that do not account for the long-memory characteristics.

5.6.5 Alternative Sampling Frequencies

In our main analysis, our results are based on measures calculated with 30-minute data. Since the sampling frequency influences the bias as well as the variance of the estimates, we repeat our analysis for realized betas calculated from 15-minute and 75-minute data.

Table 5.19 of the Appendix shows that decreasing the frequency to 75-minute data decreases the estimated memory in realized beta from 0.56 to 0.50. This is again due to an increase of the noise level in the ex-post realized betas, which negatively biases the 2ELW estimator. When increasing the recording frequency from 30-minute to 15-minute data the estimated d increases only slightly to 0.59, implying that the amount of noise in the betas calculated from 30-minute data is already small. Despite these smaller changes, it still holds for at least 97 percent of the stocks that the order of integration of their betas is significantly different from 0 and 1.

Concerning the order of integration of the realized correlation series, we observe a similar pattern. For 75-minute data the estimate decreases from the original value of 0.56 to 0.51 and for 15-minute data there is a small increase to 0.58. The ex-post estimates of stock and market volatility, on the other hand, seem to be less perturbed when decreasing the recording frequency. Here, the estimated memory is almost the same for 15-minute, 30-minute, and 75-minute data ranging from 0.58 to 0.60 for stock volatility and 0.55 to 0.57 for the inverse of market volatility.

Table 5.20 of the Appendix presents the forecast performance of the different models. We find that the ranking of the models stays the same for all considered frequencies. The FI model is the best independently of the sampling frequency. In addition, models that account for long-range dependencies perform substantially better than those that do not. Due to the difference in noise of the ex-post realized beta estimates, however, the average RMSE increases with decreasing sampling frequency. In line with the discussion above, this effect is more pronounced when changing from 30-minute to 75-minute data than when changing from 30-minute to 15-minute data.

Table 5.20 of the Appendix further reveals that changing the recording frequency only leads to small changes when comparing the models against each other. For 15-minute data the FI forecasts significantly outperforms the RW, AR, and ARMA forecasts for 77, 40, and 26 percent of the stocks while for 75-minute data this holds for 83, 48, 34 percent of the stocks, respectively.

Overall, the main message of Section 5.4 remains unchanged: accounting for long-range dependencies significantly improves the forecasting performance for realized betas.

5.6.6 Alternative Estimation Windows and Bandwidths

Our main analysis regarding the forecast performance of the models uses a rolling estimation window of 100 observations. To show that the results are robust to other specifications of the estimation window, Table 5.21 of the Appendix shows the results for window sizes of 75 and 125 observations.

While the smaller estimation window allows for more stocks to be included in the analysis, it can be seen that the results are qualitatively similar. The forecasts by the FI model perform the best and are outperformed by models that do not account for long-run dependencies only for a tiny number of stocks.

We also consider alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$ for forecasting as a final robustness check in Table 5.22 of the Appendix. These results are qualitatively similar as for our main bandwidth choice of $m = T^{0.7}$.

5.7 Conclusion

In this paper, we analyze the memory of beta factors. We first document that the betas of virtually all stocks exhibit long-memory properties. We further show that accounting for these long-memory properties is very important for forecasting. A pure long-memory FI model outperforms all other short-memory or difference-stationary models. For longer forecast horizons, the errors made by falsely imposing structures that do not account for long memory increase further.

Failing to account for the long-memory properties of betas can lead to very high errors, in particular for high-momentum stocks, liquid stocks, those with strong short-selling pressure or high idiosyncratic volatility, high-beta stocks, illiquid stocks, and those with low short interest. For the former four, imposing a random walk is most hurtful while for the latter three short-memory processes are particularly inadequate.

Appendix

Firm Characteristics

- **Age** (Zhang, 2006) is the number of years up to time t since a firm first appeared in the CRSP database.

- **Beta** is the median beta estimate for a certain stock across all estimation approaches considered.

- **Bid–ask spread (BAS)** is the stock's average daily relative bid–ask spread over the previous month.

- **Book-to-market (BtM)** (Fama and French, 1992) is the most current observation for "book equity" divided by the market capitalization. Following the standard literature, we assume that the book equity of the previous year's balance sheet statement becomes available at the end of June and use the market capitalization at the end of the corresponding fiscal year. Book equity is defined as stockholders' equity, plus balance sheet deferred taxes and investment tax credit, plus post-retirement benefit liabilities, minus the book value of preferred stock.

- **Idiosyncratic volatility (iVol)** (Ang et al., 2006a) is the standard deviation of the residuals $\epsilon_{i,\tau}$ in the Fama and French (1993) 3-factor model $r_{i,\tau} - r_{f,\tau} = \alpha_{i,t} + \beta^M_{i,t}(r_{M,\tau} - r_{f,\tau}) + \beta^S_{i,t} SMB_\tau + \beta^H_{i,t} HML_\tau + \epsilon_{i,\tau}$, using daily returns over the previous month. SMB_τ and HML_τ denote the returns on the Fama and French (1993) factors.

- **Idiosyncratic skewness (iSkew)** (Boyer et al., 2009) is the iSkew of the residuals $\epsilon_{i,\tau}$ in the Fama and French (1993) 3-factor model $r_{i,\tau} - r_{f,\tau} = \alpha_{i,t} + \beta^M_{i,t}(r_{M,\tau} - r_{f,\tau}) + \beta^S_{i,t} SMB_\tau + \beta^H_{i,t} HML_\tau + \epsilon_{i,\tau}$, using daily returns over the previous month.

- **Industry Classifications** employ the definition for 10 industry portfolios applied by Kenneth French. ''Durable" is Consumer Durables, "Energy" is the oil, gas, and coal extraction industry, "Healthcare" is Healthcare, Medical Equipment, and Drugs, "HiTec Equipment" is Business Equipment, "NonDurables" is Consumer Non-Durables, "Telephone" is Telephone and Television Transmission, "Wholesale" is Wholesale, Retail, Services, and "Other" contains Mines, Construction, Construction Materials, Transport, Hotels, Bus Services, Entertainment, as well as Finance.

- **Investment** (Fama and French, 2015) is the change in total assets from the fiscal year ending in year $t-2$ to that ending in $t-1$, divided by the total assets of year $t-2$. As for BtM, we assume that accounting data become available by the end of June of year t.

- **Leverage** (Bhandari, 1988) is defined as one minus book equity (see "Book-to-market") divided by total assets (Compustat: AT). Book equity and total assets are updated every 12 months at the end of June.

- **Marked Cap** (Banz, 1981) is the current market capitalization of a firm. Market capitalization is computed as the product of the stock price and the number of shares outstanding. In regressions, we take the natural logarithm to remove the extreme iSkew in this variable.
- **Momentum** (Jegadeesh and Titman, 1993) is the cumulative stock return over the period from $t - 12$ until $t - 1$.
- **Profitability** (Fama and French, 2015) is a firm's operating profitability. Operating profitability is revenues minus cost of goods sold minus selling, general, and administrative expenses minus interest expense, all divided by current book equity. As for BtM, we assume that accounting data become available by the end of June of year t.
- **Short interest (RSI)** (Boehme et al., 2006) is the ratio of short interest of a firm, obtained from Compustat, over the number of shares outstanding. If available, we use the short interest as of the end of month t, otherwise we use the last observation recorded in that month.

Simulation Study

To investigate the performance of different approaches for estimating the memory parameter d in small samples, we perform a small simulation study. For this purpose, we simulate data according to

$$(1 - B)^d y_t = \epsilon_t,$$

where $\epsilon \sim N(0, 1)$. To account for the high persistence in the series we consider a burn-in period of 250 observations.

We then infer on the order of integration of the series using various approaches. These include the two-step exact local Whittle estimator by Shimotsu (2010) (2ELW) as considered in this paper, the log-periodogram estimator by Geweke and Porter-Hudak (1983) (GPH) as considered in Section 5.6, the structural break robust estimators by Iacone (2010) (trLW) and Hou and Perron (2014) (HP), and the noise robust estimators by Hurvich et al. (2005) (LWN) and Frederiksen et al. (2012) (LPWN). Additionally, we consider the approach by Andersen et al. (2006) to infer on the order of integration. They investigate the autocorrelation function of the beta series and perform Ljung–Box tests on the residuals when estimating an AR(p) model to the realized beta series where p is determined by means of the AIC.

Table 5.13 reports results for $d = 0.2, 0.4, 0.6$ and $T = 100, 148, 240, 1000$ averaged across 1,000 repetitions.

The table reveals that the 2ELW and GPH estimators are almost unbiased, also for a small sample of size $T = 100$. We further find that the variance of the 2ELW estimator is smaller than that of the GPH estimator, which is in line with the results presented

	$d=0.2$				$d=0.4$				$d=0.6$			
	$T=100$	$T=148$	$T=240$	$T=1000$	$T=100$	$T=148$	$T=240$	$T=1000$	$T=100$	$T=148$	$T=240$	$T=1000$
$\bar{\hat{d}}_{2ELW}$	0.22	0.21	0.22	0.20	0.43	0.42	0.41	0.40	0.63	0.62	0.62	0.61
$sd(\hat{d}_{2ELW})$	0.13	0.11	0.09	0.05	0.13	0.11	0.09	0.05	0.13	0.11	0.09	0.05
$\bar{\hat{d}}_{GPH}$	0.20	0.20	0.21	0.20	0.42	0.41	0.41	0.40	0.62	0.62	0.62	0.61
$sd(\hat{d}_{GPH})$	0.16	0.13	0.11	0.06	0.16	0.14	0.12	0.06	0.17	0.14	0.11	0.06
$\bar{\hat{d}}_{HP}$	0.11	0.12	0.16	0.19	0.27	0.29	0.33	0.38	0.38	0.43	0.49	0.57
$sd(\hat{d}_{HP})$	0.19	0.16	0.12	0.06	0.23	0.21	0.15	0.06	0.36	0.30	0.22	0.08
$\bar{\hat{d}}_{trLW}$	0.30	0.19	0.14	0.17	0.47	0.38	0.32	0.35	0.66	0.55	0.53	0.55
$sd(\hat{d}_{trLW})$	0.42	0.34	0.24	0.11	0.44	0.34	0.24	0.11	0.45	0.32	0.25	0.12
$\bar{\hat{d}}_{LWN}$	0.36	0.36	0.32	0.25	0.52	0.49	0.47	0.43	0.68	0.67	0.65	0.63
$sd(\hat{d}_{LWN})$	0.29	0.26	0.20	0.08	0.21	0.17	0.13	0.06	0.16	0.14	0.11	0.06
$\bar{\hat{d}}_{LPWN}$	0.40	0.39	0.37	0.28	0.55	0.53	0.50	0.45	0.69	0.69	0.68	0.65
$sd(\hat{d}_{LPWN})$	0.36	0.33	0.29	0.16	0.29	0.26	0.20	0.09	0.24	0.19	0.15	0.08
Sign. ac (%)	4.30	6.49	10.53	30.29	13.56	22.70	37.79	88.41	28.21	45.74	67.28	99.55
Ljung–Box	0.006	0.008	0.006	0.000	0.004	0.003	0.001	0.000	0.008	0.001	0.002	0.000

Table 5.13: Simulation results.
We simulate T observations of fractional white noise that is integrated of order $I(d)$ and then compare different approaches to infer on the memory parameter d. This table reports average d estimate and standard deviation ($sd(\cdot)$) for the estimators by Shimotsu (2010) (2ELW), Geweke and Porter-Hudak (1983) (GPH), Hou and Perron (2014) (HP), Iacone (2010) (trLW), Hurvich et al. (2005) (LWN), and Frederiksen et al. (2012) (LPWN). Additionally, the table shows the average percent of the first 36 autocorrelations that are indicated to be significantly larger than zero by 95 percent Bartlett confidence intervals. This is the technique Andersen et al. (2006) use to decide on the order of integration of the series. They further consider Ljung–Box tests on the residuals of AR(p) processes, where p is selected by the AIC. In case there is significant autocorrelation in the residuals, the null is rejected, indicating that there is long memory in the series. The last row reports the power of this approach for the simulated series, i.e., the relative number of times the null hypothesis is rejected. All results are the averages over 1,000 repetitions.

in Section 5.6. Concerning the break robust estimators, it can be seen that both the HP estimator and the trLW estimator are negatively biased in sample sizes of 148 and 240. The noise robust estimators, on the other hand, are positively biased for sample sizes of 100, 148, and 240.

Andersen et al. (2006) investigate quarterly betas for which, due to the noise, the observed order of integration is decreased, such that the 2ELW estimator yields a d of 0.4 on average. They then fractionally differenced the series by 0.2, such that the resulting series should be approximately I(0.2). For such a series the simulations indicate that only 6 percent of the first 36 autocorrelations are significantly larger zero according to 95 percent Bartlett confidence intervals. It is understandable that, based on such autocorrelation functions, the authors conclude that realized betas exhibit a d of 0.2 or smaller. The simulations further reveal that Ljung–Box tests on the residuals of an $AR(p)$ with p selected by the AIC are not particularly useful to detect long-memory time series. The order p is simply chosen to be high, such that the long-memory characteristics can be captured by the AR model.

Tables Referenced in the Main Manuscript

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
Bandwidth $m = T^{0.65}$								
β_i	0.575	0.127	0.996	0.989	0.532	0.156	0.987	0.989
$\rho_{i,M}$	0.554	0.102	1.000	0.998	0.561	0.109	1.000	0.996
σ_i	0.586	0.146	0.991	0.968	0.586	0.146	0.991	0.968
σ_M^{-1}	0.544	-	1.000	1.000	0.544	-	1.000	1.000
Bandwidth $m = T^{0.75}$								
β_i	0.549	0.103	0.999	0.998	0.517	0.122	0.996	0.998
$\rho_{i,M}$	0.546	0.088	0.995	1.000	0.543	0.090	1.000	1.000
σ_i	0.592	0.137	1.000	0.989	0.592	0.137	1.000	0.989
σ_M^{-1}	0.591	-	1.000	1.000	0.590	-	1.000	1.000

Table 5.14: Average memory parameter estimates – bandwidth $m = T^{0.65}$ and $m = T^{0.75}$.

In analogy to Tables 5.1 and 5.2, this table presents average estimates of the memory parameter of realized betas, realized correlation (Fisher-transformed), and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the 2ELW estimator of Shimotsu and Phillips (2005) and Shimotsu (2010) with alternative bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$. $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d = 0$ and $d = 1$, respectively, are rejected at the ten percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle and Moulines (2000).

	P1	P2	P3	P4	P5	t-stat
β	0.4830	0.7726	0.9769	1.2075	1.6887	22.1
d	0.5333	0.5213	0.5394	0.5550	0.5822	3.57
log(Market Cap)	16.175	16.215	16.177	16.016	15.673	-4.55
BtM	0.5380	0.5060	0.5016	0.5299	0.5628	0.49
Investment	0.0938	0.1067	0.1077	0.1191	0.1177	1.59
Profitability	-0.5928	-3.4268	-0.5836	0.2827	0.1098	1.07
Momentum	0.1276	0.1329	0.1359	0.1485	0.2029	1.39
BAS	0.0009	0.0007	0.0007	0.0008	0.0010	0.98
Turnover	0.1846	0.1893	0.2114	0.2540	0.3882	6.40
iVol	0.0115	0.0115	0.0125	0.0143	0.0193	5.10
iSkew	0.0834	0.1044	0.1038	0.1269	0.1500	5.06
Short Interest	0.0302	0.0337	0.0369	0.0419	0.0558	11.2
Leverage	0.5920	0.5889	0.5774	0.5745	0.5922	0.01
Age	37.611	33.049	33.317	32.864	29.838	-11.0
Durables	0.0048	0.0107	0.0230	0.0384	0.0309	10.3
Energy	0.0309	0.0319	0.0441	0.0783	0.1137	3.42
Healthcare	0.1299	0.1096	0.0664	0.0448	0.0378	-2.79
HiTec Equipment	0.0812	0.1401	0.1790	0.1776	0.2001	2.41
Manufacturing	0.0781	0.1153	0.1655	0.1721	0.1661	9.19
NonDurables	0.1624	0.0819	0.0518	0.0364	0.0309	-18.0
Telephone	0.0397	0.0452	0.0340	0.0227	0.0132	-8.33
Utilities	0.1721	0.0806	0.0341	0.0230	0.0108	-10.0
Wholesale	0.1185	0.1374	0.1259	0.1128	0.0775	-2.97

Table 5.15: Portfolio sorts by beta.
At the end of each month, we sort the stocks in our sample based on the realized beta during the past month. Sorting the stocks into quintile portfolios, we save each portfolio's average of the firm characteristics and dummy variables at the end of the respective months. The main body of the table shows the average of the different firm characteristics over time ($T = 141$ months). t-stat denotes the t-statistic of a test whether the firm characteristics of portfolio P5 and P1 are equal with the standard errors being calculated using the heteroscedacity and autocorrelation robust approach by Andrews (1991). Characteristics, for which this difference is statistically significant at ten percent are printed in **bold**.

	RW	RWV	RWLW	AR	ARMA	ARFIMA	HAR	FI(0.5)	FI
RMSE	0.2820	0.3124	0.3862	0.2854	0.2850	0.2808	0.3176	0.2775	0.2792
Best	92	39	4	11	19	84	4	313	123
vs RW	0	2	1	2	2	53	0	85	80
vs RWV	279	0	0	248	243	315	78	394	335
vs RWLW	569	522	0	560	560	572	432	592	584
vs AR	171	6	1	0	41	81	1	107	109
vs ARMA	139	5	1	40	0	81	0	103	101
vs ARFIMA	15	1	1	9	9	0	1	68	49
vs HAR	489	81	2	425	419	495	0	440	400
vs FI(0.5)	9	1	0	3	4	8	0	0	11
vs FI	24	2	0	12	11	22	0	82	0
N	689	689	689	689	689	689	689	689	689

Table 5.16: Forecast results – alternative models.
In analogy to Table 5.3, this table illustrates the forecast performance of different additional models for one-month beta forecasts from a rolling estimation window of 100 observations. RW, AR, ARMA, and ARFIMA are estimated in a state-space framework. RWV and RWLW correspond to the forecasts from the approaches of Vasicek (1973) and Levi and Welch (2017), respectively. Finally, HAR corresponds to the model by Corsi (2009) and FI(0.5) uses a FI model with d fixed at 0.5. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
β_i	0.557	0.133	0.994	0.953	0.518	0.158	0.967	0.965
$\rho_{i,M}$	0.572	0.122	0.995	0.966	0.572	0.128	0.991	0.962
σ_i	0.593	0.164	0.978	0.930	0.593	0.164	0.978	0.930
σ_M^{-1}	0.591	-	1.000	1.000	0.590	-	1.000	1.000

Table 5.17: Average memory parameter estimates – log-periodogram estimator.
In analogy to Tables 5.1 and 5.2, this table presents average estimates of the memory parameter of realized betas, realized correlation (Fisher-transformed), and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the log-periodogram estimator by Geweke and Porter-Hudak (1983). $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d = 0$ and $d = 1$, respectively, are rejected at the ten percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle and Moulines (2000).

	RW	AR	ARMA	FI	ARFIMA
RMSE	0.3149	0.2942	0.2878	0.2812	0.2814
Best	7	30	128	332	192
vs RW	0	271	343	507	455
vs AR	3	0	155	228	259
vs ARMA	1	5	0	140	135
vs FI	1	5	22	0	21
vs ARFIMA	1	4	11	21	0
N	689	689	689	689	689

Table 5.18: Forecast results – log-periodogram estimator. In analogy to Table 5.3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. FI and ARFIMA model are now calculated using d estimates by the log-periodogram estimator instead of the 2ELW estimator. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

	Standard				Adjusted for Breaks in Mean			
	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$	$\bar{\hat{d}}_i$	$sd(\hat{d}_i)$	vs. $d_i = 0$	vs. $d_i = 1$
15-Minute Data								
β_i	0.593	0.112	0.999	0.996	0.553	0.139	0.994	0.998
$\rho_{i,M}$	0.583	0.099	1.000	0.996	0.585	0.101	1.000	0.996
σ_i	0.598	0.139	0.996	0.977	0.597	0.139	0.996	0.977
σ_M^{-1}	0.553	-	1.000	1.000	0.553	-	1.000	1.000
75-Minute Data								
β_i	0.498	0.123	0.994	0.998	0.457	0.148	0.977	0.998
$\rho_{i,M}$	0.505	0.096	0.999	0.998	0.496	0.099	1.000	1.000
σ_i	0.578	0.140	0.995	0.984	0.578	0.140	0.995	0.984
σ_M^{-1}	0.568	-	1.000	1.000	0.566	-	1.000	1.000

Table 5.19: Average memory parameter estimates – 15-minute and 75-minute data.

In analogy to Tables 5.1 and 5.2, this table presents average estimates of the memory parameter of realized betas, realized correlation (Fisher-transformed), and volatility across all stocks ($N = 823$), as well as that of the inverse of the market volatility, using the 2ELW estimator of Shimotsu and Phillips (2005) and Shimotsu (2010). The realized measures are now calculated from 15 and 75-minute data. $sd(\hat{d}_i)$ displays the standard deviation of the estimates across stocks and vs. $d_i = 0$ and vs. $d_i = 1$ indicate the relative frequency with which the null hypotheses $d = 0$ and $d = 1$, respectively, are rejected at the ten percent level. The left panel reports the results for the original series and the right panel reports results after adjusting the series for structural breaks using the procedure of Lavielle and Moulines (2000).

	RW	AR	ARMA	FI	ARFIMA
			15-Minute Data		
RMSE	0.2876	0.2713	0.2660	0.2586	0.2595
Best	5	27	124	362	171
vs RW	0	268	320	530	499
vs AR	2	0	139	277	293
vs ARMA	1	8	0	182	180
vs FI	1	4	15	0	13
vs ARFIMA	1	4	18	19	0
N	689	689	689	689	689
			75-Minute Data		
RMSE	0.3724	0.3421	0.3355	0.3241	0.3248
Best	2	25	90	432	140
vs RW	0	291	360	574	541
vs AR	4	0	156	333	353
vs ARMA	2	9	0	231	222
vs FI	0	1	4	0	12
vs ARFIMA	0	2	11	21	0
N	689	689	689	689	689

Table 5.20: Forecast results – 15-minute and 75-minute data. In analogy to Table 5.3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. For the different panels, the realized beta series are now, however, based on 15-minute and 75-minute data. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

	RW	AR	ARMA	FI	ARFIMA
			Rolling Window of 75 Observations		
RMSE	0.3102	0.2914	0.2864	0.2764	0.2775
Best	4	33	78	446	208
vs RW	0	263	318	612	570
vs AR	3	0	168	367	384
vs ARMA	2	22	0	235	241
vs FI	0	0	3	0	11
vs ARFIMA	0	2	9	28	0
N	769	769	769	769	769
			Rolling Window of 125 Observations		
RMSE	0.3017	0.2800	0.2746	0.2676	0.2683
Best	4	19	120	330	130
vs RW	0	255	302	479	433
vs AR	4	0	114	205	201
vs ARMA	3	15	0	116	113
vs FI	0	1	11	0	13
vs ARFIMA	0	1	17	25	0
N	603	603	603	603	603

Table 5.21: Forecast results – rolling window sizes of 75 and 125 observations. In analogy to Table 5.3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 75 as well as 125 observations. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

	RW	AR	ARMA	FI	ARFIMA
			Bandwidth $m = T^{0.65}$		
RMSE	0.3149	0.2942	0.2878	0.2797	0.2806
Best	3	23	108	367	188
vs RW	0	271	343	558	506
vs AR	3	0	155	270	295
vs ARMA	1	5	0	166	172
vs FI	0	2	13	0	21
vs ARFIMA	0	2	12	24	0
N	689	689	689	689	689
			Bandwidth $m = T^{0.75}$		
RMSE	0.3149	0.2942	0.2878	0.2794	0.2801
Best	6	30	100	399	154
vs RW	0	271	343	564	534
vs AR	3	0	155	310	313
vs ARMA	1	5	0	192	189
vs FI	0	4	7	0	10
vs ARFIMA	0	4	9	20	0
N	689	689	689	689	689

Table 5.22: Forecast results – bandwidth $m = T^{0.65}$ and $m = T^{0.75}$. In analogy to Table 5.3, this table illustrates the forecast performance of the models for one-month beta forecasts from a rolling estimation window of 100 observations. FI and ARFIMA model now calculated using d estimates of the 2ELW estimator calculated with bandwidths of $m = T^{0.65}$ and $m = T^{0.75}$. The first row shows average RMSEs of different models across all stocks. The row "Best" indicates the number of times a model achieves the lowest RMSE for a certain stock. Furthermore, the rows denoted by "vs. X" correspond to modified DM tests (Harvey et al., 1997), providing the number of times the forecasts of the column-model are significantly better than the forecasts of the row-model at the ten percent level. Finally, N is the number of investigated stocks. To allow for valid inference, we exclude all stocks for which we have less than 50 forecasts.

Chapter 6

Robust Multivariate Local Whittle Estimation and Spurious Fractional Cointegration

Co-authored with Christian Leschinski and Philipp Sibbertsen.

6.1 Introduction

It has been a well-established fact that level shifts among many other so-called low-frequency contaminations can be mistaken as long memory. Künsch (1986), Granger and Ding (1996), and Diebold and Inoue (2001), among others, show that various forms of low-frequency contaminations such as deterministic breaks and trends can cause spurious long memory. This leads to a bias of semiparametric estimators for the memory parameter, which mainly use these frequencies. This feature has been used by Qu (2011) to test against spurious long memory.

However, the notion of spurious long memory is not restricted to the univariate case but can as well be found in multivariate systems. An extension of the test by Qu (2011) to the multivariate case relying on the same idea can be found in Sibbertsen et al. (2018). This paper shows that working in a multivariate system can result in efficiency gains and is therefore preferable where suitable. Multivariate local Whittle estimation of the memory parameter has been considered in Shimotsu (2007). Robinson (2008) extends this to the case of possible fractional cointegration and simultaneously estimates the cointegration vector. Two series are called fractionally cointegrated if they have the same memory parameter and their linear combination has a reduced order of integration (see among many others Nielsen, 2007). Neither of these two estimators is robust against low-frequency contaminations.

The aim of this paper is to provide such a robust multivariate local Whittle estimator of the memory parameter and the fractional cointegration vector that remains consistent in case of low-frequency contaminations. Similar to the estimator of Robinson (2008), our proposed estimator requires a priori knowledge of the cointegration rank. This is because local Whittle-based methods need the inverse of the so-called G matrix of the spectral density, which becomes singular in the case of fractional cointegration. Christensen and Santucci de Magistris (2010) and Kellard et al. (2015) discuss that in case of low-frequency contaminations inference on the fractional cointegration rank is likely to be biased. For example, simultaneous breaks in the series can cause tests and estimators to falsely indicate the series to be fractionally cointegrated.

We therefore additionally suggest a robust estimator of the cointegration rank. For this purpose, we investigate what we call spurious fractional cointegration further by generalizing the definition of cobreaking in Hendry and Massmann (2007) to what we call common low-frequency contaminations. We show that low-frequency contaminations dominate the G matrix of the periodogram for frequencies close to the origin and therefore empirically effect, among others, local Whittle-based procedures. Due to this dominance of low-frequency contaminations in the observed G matrix, we find that common low-frequency contaminations spuriously indicate the presence of fractional cointegration, whereas distinct low-frequency contaminations falsely indicate the absence of fractional cointegration.

To obtain our estimators, we use the idea of Iacone (2010) of trimming away the contaminated frequencies. This idea is applied to provide a consistent estimator of the cointegration rank of the system by proposing a trimmed version of the procedure by Robinson and Yajima (2002) as well as to construct a robust multivariate local Whittle estimator for the memory parameter and the cointegration vector. As our estimators rely on properties of the periodogram of processes with low-frequency contaminations, we find it useful to additionally provide some deeper understanding of the behavior of the periodogram in this situation.

The paper is structured as follows. First, we provide some results for the periodogram of a contaminated process in a rather general framework of low-frequency contaminations generalizing previously obtained results in Section 6.2. Section 6.3 formally defines common low-frequency contaminations and contains our robust procedure to estimate the cointegration rank while Section 6.4 has the robust multivariate local Whittle estimator. Section 6.5 contains some Monte Carlo and Section 6.6 an empirical example. Section 6.7 concludes. All the proofs are gathered in the Appendix and additional simulation results are provided in a Supplementary Appendix.

6.2 The Periodogram of Spurious Long-Memory Processes

In this section, we obtain some properties of the periodogram for a very general class of low-frequency contaminations, which are partly needed later but are also of an interest on its own. We therefore discuss it in more detail then necessary for our robust estimators and see this as an additional contribution of the paper. Although the focus of this paper is on multivariate estimation, we derive the results in this section in a univariate setup to avoid notational complexity. We will later assume the trend process to be independent from the noise process, which means that an extension of the results to a multivariate framework is straightforward. It further implies that effects of additional noise components are irrelevant for the mean process so that we are only concerned with the behavior of the pseudo-periodogram of a time-varying mean process in this section.

For the mean process, we use a very general specification allowing for deterministic mean shifts, smooth deterministic trends, and random level shifts with rare shift asymptotics as well as random level shifts with medium rare shifts, where the number of shifts tends to infinity with sample size but with a slower rate. This model embeds many of the processes discussed in the literature to generate spurious long memory such as the fractional trend of Bhattacharya et al. (1983) or the STOPBREAK model of Engle and Smith (1999).

The mean process that can be either deterministic or stochastic is represented by

$$\mu_t = \mu_0 + \sum_{k=1}^{K} \Delta\mu_k 1(t \geq T_k), \tag{6.1}$$

or

$$\mu_t = \mu_0 + \sum_{k=0}^{K} \mu_k 1(T_{k-1} \leq t < T_k). \tag{6.2}$$

Here, K is either a fixed number or a random variable giving the number of breaks, T is the length of the series, 1 corresponds to the indicator function, T_k denotes the breakpoint, and $\mu_0 = \mu_1 - 1/T \sum_{t=1}^{T} \mu_t$.

The expression in Equation (6.1) is a suitable representation of the mean for processes which have a nonstationary nature. Examples include deterministic trends or mean shifts as well as nonstationary random level-shift models or the STOPBREAK model. The model is accumulative in the sense that the break at time t depends on all shifts that occurred before t. It also nests a random walk. The model in (6.2), on the other hand, has a stationary character and seems appropriate for models such as the Markov-Switching model or stationary random level-shift models. It has a non-cumulative structure and nests the White Noise.

In the following, we denote by $I_z(\lambda_j) = w_z(\lambda_j)w_z^*(\lambda_j)$ the periodogram of the series z_t at frequency λ_j. Here, $w_z(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} z_t e^{i\lambda_j t}$ is the Fourier transform of the series z_t and the asterisk denotes complex conjugation. $I_\mu(\lambda_j)$ is then the pseudo-periodogram of the mean process. We focus on the behavior of this pseudo-periodogram at the Fourier frequencies $\lambda_j = \frac{2\pi j}{T}$ for $\lambda_j \to 0_+$.

We now derive the properties of the induced periodogram of the process (6.1) and (6.2). Let us first consider the case of a smooth trend $h(s, T)$ and assume:

Assumption A1. $|h(s, T)|$, $\left|\frac{\partial h(s,T)}{\partial s}\right| < \infty$, for $s \in [0, 1]$.

We have the following Lemma:

Lemma 1. If $\mu_t = h(t/T, T)$, under Assumption A1 we have

$$I_\mu(\lambda_j) \sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi j s) ds \right]^2 + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} (1 - \cos(2\pi j s)) ds \right]^2 \right\}.$$

Since the integrals in Lemma 1 are functions of j (and possibly T), it can be seen that the exact rate of the periodogram $I_\mu(\lambda_j)$ depends on the derivative of the trend function. Therefore, if the trend function is known and the integrals have a closed form solution, it is possible to determine the exact order. If this is not the case, we can still recover the upper bound on the rate of decay for increasing j that was established by Künsch (1986), Iacone (2010), and Qu (2011). To see this, note that $\sin(2\pi js) \leq 1$ and $1 - cos(2\pi js) \leq 2$ for all j and s. It therefore follows immediately for $\mu_t = h(s,T) = h(s)$ that the periodogram is $I_\mu(\lambda_j) = O(Tj^{-2})$.

We now turn to the behavior of the periodogram of abrupt level-shift processes. To simplify the exposition, let ζ_k denote either $\Delta\mu_k$ or μ_k, depending on whether the accumulative structural-change model (6.1) or the non-accumulative model (6.2) is considered.

To characterize the behavior of different groups of processes, we require different groups of assumptions. First, in the case of deterministic structural breaks, we assume:

Assumption A2. $|\zeta_k| < \infty$ and the $\delta_k = T_k/T$ are deterministic with $0 < \delta_k < 1$ for $k = 1, \ldots, K < \infty$.

For stochastic level shifts we require the following assumptions.

Assumption A3. $E[\zeta_k] = 0$ and $Var[\zeta_k] = \sigma_\Delta^2 T^{-\beta}$ for some $0 \leq \beta \leq 1$ and $0 < \sigma_\Delta^2 < \infty$.

Assumption A4. $P(t \in \{T_1, \ldots, T_K\}) = p_t$, where $0 \leq p_t \leq 1$, and $E[p_t] = \tilde{p}T^{-\alpha}$ for some $0 \leq \alpha \leq 1$. Furthermore, the dependence in p_t is limited such that $E[K] = \tilde{p}T^{1-\alpha}$, $E\left[((T_k - T_{k-1})/T)^2\right] = \frac{2\tilde{D}}{\tilde{p}^2} T^{2(\alpha-1)}$, and $E\left[((T_k - T_{k-1})/T)^4\right] = O(T^{4(\alpha-1)})$, for some $0 < \tilde{p}, \tilde{D} < \infty$.

Assumption A5. p_t is independent of ζ_k for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$. Additionally, $Var[\zeta_k]\tilde{C} = \sum_{\tau=1}^{\infty} |E[\zeta_k \zeta_{k-\tau}]|$ for $k = 1, 2, \ldots$ and $0 \leq \tilde{C} < \infty$.

The rate $T^{-\beta}$ in Assumption A3 is required to nest a number of mean-change processes from the literature, such as the STOPBREAK process of Engle and Smith (1999). For other processes, setting $\beta = 0$ gives the familiar setup with non-degenerate breaks.

Assumption A4 imposes a structure on the nature of the mean change process. The nature of the dependence in p_t is restricted by the additional requirement that the expected squared length of the k-th regime expressed as a fraction of the sample is $\frac{2\tilde{D}}{\tilde{p}^2} T^{2(\alpha-1)}$, which means that the second moment of the regime lengths is still of the same order as that of a geometric distribution. In this context, the constant $\tilde{D}$ depends on the dependence in p_t, and it is equal to one if $p_t = p$ for all $t = 1, \ldots, T$.

Since there are T observations in the sample, the expected number of mean shifts in the series is $E[K] = \tilde{p}T^{1-\alpha}$. The parameter α controls the asymptotic frequency of level changes. The expected number of shifts remains constant for $\alpha = 1$, whereas it goes to infinity for $\alpha < 1$. The first case ($\alpha = 1$) is referred to as *rare shifts* asymptotics or *low-frequency contaminations*. We refer to the second case ($\alpha < 1$) as *intermediate-frequency contaminations*. Here, we have $K \to \infty$ but $K/T \to 0$ as $T \to \infty$. That means we asymptotically have an infinite number of shifts but also an infinite number of observations between shifts. Finally, for $\alpha > 1$ shifts are so rare that we will no longer observe any in a sample, asymptotically.

Even though it may seem unusual to tie the properties of the process to the sample size, this is a common approach in the related literature. Guégan (2005) refers to this practice as a thought experiment. The validity of this approach depends on the purpose of the analysis. Obviously, it is unreasonable to assume that structural changes will become less common in the future if the objective is to forecast a time series. On the other hand, if the objective is statistical inference based on a given sample, we argue that assuming that the frequency of structural change is tied to the sample size T can be thought of as an asymptotic framework that is better suited to approximate the statistical properties of the quantities of interest than keeping p fixed. The latter would imply, for example, that level changes are so frequent that the mean between two shifts cannot be estimated consistently.

Finally, we require some bound on the degree of dependence between the means or mean changes ζ_k in consecutive segments. This is imposed by Assumption A5 according to which the autocovariance function of the ζ_k has to be absolutely summable. We then obtain the following result.

Lemma 2. Denote by $\kappa > 0$ a finite constant and by $|\kappa_T| \leq 1$ a sequence of constants. Then, for $j/T \to 0$ and level-shift processes characterized by (6.1),

i.) $I_\mu(\lambda_j) \sim \frac{T}{4\pi^3 j^2}\kappa$, under Assumption A2.

ii.) $E\left[I_\mu(\lambda_j)\right] \sim \frac{\sigma_\Delta^2 \tilde{p} T^{2-\alpha-\beta}}{4\pi^3 j^2}\left(1 + \kappa_T \tilde{C}\right)$, for $\alpha \leq 1$, and under Assumptions A3, A4, and A5.

Lemma 2 establishes the properties of the periodogram of the accumulative mean-change process in (6.1). The first case i.) derives the growth rate of the peak near the origin and the rate of decay for frequencies further away from the zero frequency for a deterministic structural break process. This order was previously established by McCloskey and Perron (2013). Interesting is the contrast to case ii.), where rare random level shifts are considered. In contrast to i.), the periodogram becomes stochastic instead of deterministic. Furthermore, the scaling factor $T^{-\beta}$ influences the scaling of the peak local to the origin, which is of order $T^{1-\beta}$ instead of T. In i.) the periodogram is a deterministic function. In ii.) there is a well defined expectation and the process is not ergodic for $\alpha = 1$, the expected number of shifts in the sample is always given by $E[K] = \tilde{p}$. The case of $\alpha < 1$, on the other hand, covers intermediate frequency contaminations so that the expected number of shifts is $E[K] = \tilde{p}T^{1-\alpha}$ and the process is ergodic. In this situation, the scaling of the peak near the origin is determined by both, α and β. Since $\alpha < 1$, the growth rate is always faster than that in case i.) and for $\alpha = 1$. The rate of decay for increasing j, however, is the same for all three types of processes.

Similar results to these can be obtained for the non-accumulative mean-change process in (6.2).

Lemma 3. Denote by $\kappa' > 0$, $|\kappa'_T| \leq 1$, and $\kappa'_{P,T}$, a finite constant, a sequence of constants, and a sequence of positive valued random variables with constant expectation and finite variance, respectively. Then, for $j/T \to 0$ and level-shift processes characterized by (6.2),

i.) $I_\mu(\lambda_j) \sim \frac{T}{2\pi^3 j^2}\kappa'$, under Assumption A2.

ii.) $I_\mu(\lambda_j) \sim \frac{\sigma_\Delta^2 \tilde{p} T^{1-\beta}}{2\pi^3 j^2}\kappa'_{P,T}$, for $\alpha = 1$, and under Assumptions A3 and A4.

iii.) $E\left[I_\mu(\lambda_j)\right] \sim \frac{\sigma_\Delta^2 \tilde{D}}{\pi \tilde{p}} T^{\alpha-\beta}\left(1 + \kappa'_T \tilde{C}\right)$, for $\alpha < 1$, and under Assumptions A3, A4, and A5.

As one can see, the orders for cases i.) and ii.) in Lemma 3 are identical to those in Lemma 2. This means that accumulative and non-accumulative structural change have the same impact on the periodogram local to zero as long as the mean changes are deterministic or rare. In contrast to that, the case $\alpha < 1$ is remarkably different and needs to be treated separately. In presence of intermediate frequency contaminations, when the process becomes ergodic, the order of the peak is reduced to $T^{\alpha-\beta}$ instead of $T^{2-\alpha-\beta}$. Furthermore, the peak local to zero no longer decays for increasing j. This is a behavior similar to a white noise reflecting the stationary structure of the non-accumulative approach.

Important special cases of both the accumulative and the non-accumulative process are obtained for $\alpha = 0$. In this case, the accumulative process boils down to a unit root process and the non-accumulative process becomes a simple stationary short-memory process. In this situation, case ii.) in Lemma 2 and iii.) in Lemma 3 reduces to the well known result that the periodogram of the unit root process local to the origin is of order $O_P(T^2/j^2)$ and that of the short-memory process is $O_P(1)$.

6.3 Robust Fractional Cointegration Rank Estimator

In this section, we first provide evidence that existing semiparametric estimators and tests for the fractional cointegration rank are biased in case of low-frequency contaminations.

We then derive our robust fractional cointegration rank estimator as an extension of the rank estimator by Robinson and Yajima (2002).

Point of departure is a vector valued long-memory process y_t with low-frequency contaminations. For expositional simplicity, we focus on a bivariate system, i.e., $y_t = (y_{at}, y_{bt})'$, extensions to higher dimensions are straightforward. The process under investigation is

$$y_t = x_t + \mu_t, \tag{6.3}$$

where $\mu_t = (\mu_{at}, \mu_{bt})'$ is a bivariate low-frequency contamination process that is independent of x_t and where for μ_{at} and μ_{bt} either of the Assumptions A1, A2, or A3 with $\beta = 0$ holds. Moreover, $x_t = (x_{at}, x_{bt})'$ is a bivariate long-memory process whose spectral density matrix $f(\lambda_j)$ at frequency λ_j fulfills

$$f(\lambda_j) \sim \Lambda_j(d) G_x \Lambda_j^*(d), \tag{6.4}$$

where $\Lambda_j(d) = \mathrm{diag}(\lambda_j^{-d_a} e^{i(\pi-\lambda_j)d_a/2}, \lambda_j^{-d_b} e^{i(\pi-\lambda_j)d_b/2})$ with $i = \sqrt{-1}$, and $d = (d_a, d_b)$ are the memory parameters. Further, $A \sim B$ denotes that $A/B \to 1$ as $\lambda \to 0_+$, and A^* denotes the complex conjugate of A.

As shown in Marinucci and Robinson (2001), the matrix G_x is positive definite if and only if x_t is not (fractionally) cointegrated and it becomes singular otherwise. Consequently, by estimating the rank of G_x we can investigate whether the time series are (fractionally) cointegrated as suggested by Robinson and Yajima (2002). However, if instead of the pure memory process we observe a contaminated process such as (6.3), then our estimate of the rank will be based on G_y which comprises the influence of G_x and G_μ.

To illustrate this, note that the periodogram as an estimate of G_y is given by

$$I_y(\lambda_j) = I_\mu(\lambda_j) + I_x(\lambda_j) + I_{\mu x}(\lambda_j) + I_{x\mu}(\lambda_j),$$

where $I_{\mu x}(\lambda_j) = w_\mu(\lambda_j) w_x^*(\lambda_j)$ and $I_{x\mu}(\lambda_j) = w_x(\lambda_j) w_\mu^*(\lambda_j)$ are the cross periodograms of μ_t and x_t. Consequently, $E[I_y(\lambda_j)] = E[I_\mu(\lambda_j)] + E[I_x(\lambda_j)]$ if x_t and μ_t are assumed to be independent.

For the long-memory component it holds that $E[I_x(\lambda_j)] = \left(\frac{T}{j}\right)^{2\max\{d_a,d_b\}} G_x$ as $j/T \to 0$. The properties of $E[I_\mu(\lambda_j)]$ can be derived based on our results presented in Section 6.2. Here, we are interested in the empirically relevant situations of a smooth trend, a deterministic break, or a random level-shift process with rare shifts. These low-frequency contaminations can be distinct, i.e., each series faces different contaminations, or they can be common as discussed in Hendry and Massmann (2007). Their definition is limited to contemporaneous mean cobreaking, i.e., common deterministic structural changes. Furthermore, their definition refers to changes relative to some initial parametrization so that processes with a stable monotonous trend, for example, are not included. We therefore propose the following slightly modified definition:

Definition 1 (CLFC)**.** The bivariate process y_t in (6.3) has common low-frequency components if there exists a 2×1 matrix $\Phi \neq 0_{21}$, such that $\Phi' (\mu_t - \mu_1) = 0$ for all $t = 1, \ldots, T$.

Now, we are able to derive the order of the expected pseudo-periodogram of μ_t.

Theorem 1. Suppose y_t is generated by (6.3) and $j/T \to 0$, we have

$$E[I_\mu(\lambda_j)] = \frac{T}{j^2} G_\mu,$$

and G_μ has rank 1 if and only if μ_t is a common low-frequency component according to Definition 1.

It is obvious from the rate in the theorem that the G_μ matrix dominates the G_y matrix for low frequencies, while the G_x matrix is the dominating one for higher frequencies. If we now observe time series which are not fractionally cointegrated but exhibit joint breaks, then the estimator by Robinson and Yajima (2002) might spuriously identify a fractional cointegration relation since the G_μ matrix is singular. On the other hand, if we observe fractionally cointegrated time series which exhibit distinct breaks, then the same estimator might wrongly identify no fractional cointegration relation since the G_μ matrix has full rank. These problems do not only arise for the estimator by Robinson and Yajima (2002) but for all existing semiparametric estimators and tests concerning the fractional cointegration relation, since all of them use the G_y matrix in some form. We will demonstrate this by simulations in Section 6.5.

It should further be noted that testing the homogeneity of fractional difference parameters, as suggested by Robinson and Yajima (2002), is also not possible in case of low-frequency contaminations. This is due to two reasons. First, the estimates of the memory parameter will be biased when using the standard local Whittle estimator. This issue can be overcome by considering a robust estimator such as those by Iacone (2010), McCloskey and Perron (2013), or Hou and Perron (2014). However, the test statistics also includes an estimate of the G_x matrix which is based on the first m_1 frequencies. In case of joint breaks, this matrix might be estimated to be singular letting the statistics converge to zero no matter if the order of integrations are truly equal.

Let us now introduce an approach to estimate the G_x matrix consistently also in case of low-frequency contaminations. We know from Theorem 1 that the G_μ matrix only dominates for the low frequencies. If we trim these away, then we can estimate the rank of the G_x matrix without distortions no matter if low-frequency contaminations are present or not. We first show that the estimated G_y matrix trimmed by the first frequencies converges to the estimated G_x matrix. For this purpose, we need to introduce the following assumptions.

Assumption B1. It holds that

$$x_t - E[x_t] = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j},$$

where $\sum_{j=0}^{\infty} \|A_j\|^2 < \infty$, and $\|\cdot\|$ denotes the supremum norm. It is further assumed that $E[\varepsilon_t \varepsilon_t^{'} | \mathfrak{F}_{t-1}] = I_q$, $E[\varepsilon_t | \mathfrak{F}_{t-1}] = 0$ a.s. for $t = 0, \pm 1, \pm 2, \dots$ where $\mathfrak{F}_t$ denotes the σ-field generated by ε_s and I_q is an identity matrix, $s \leq t$. Furthermore, there exists a scalar random variable ε such that $E[\varepsilon^2] < \infty$ and for all $\tau > 0$ and some $C > 0$ it is $P(\|\varepsilon_t\|^2 > \tau) \leq CP(\varepsilon^2 > \tau)$.

Assumption B2. As $T \to \infty$,

$$\frac{l}{m_1} + \frac{m_1}{T} \to 0,$$

where l is a trimming parameter with $l = \max(1, [c_l T^{\delta_l}])$ and $m_1 = \max(l+1, [c_{m_1} T^{\delta_{m_1}}])$ is the bandwidth parameter with $0 \leq \delta_l < \delta_{m_1} < 1$ and $c_l, c_{m_1} \in (0, \infty)$.

Let us further assume for the moment that the order of integration is known and denote

$$\begin{aligned}
\hat{G}_y(d, l, m_1) =& (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_y(\lambda_j) \Lambda_j^*(d) \\
=& (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_x(\lambda_j) \Lambda_j^*(d) \\
&+ (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_\mu(\lambda_j) \Lambda_j^*(d) \\
&+ (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_{x\mu}(\lambda_j) \Lambda_j^*(d) \\
&+ (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_{\mu x}(\lambda_j) \Lambda_j^*(d) \\
\text{and} \quad \hat{G}_x(d, l, m_1) =& (m_1 - l + 1)^{-1} \sum_{j=l}^{m_1} \Lambda_j(d) I_x(\lambda_j) \Lambda_j^*(d).
\end{aligned}$$

Theorem 2. Suppose y_t is generated by (6.3), Assumptions B1 and B2 hold with $m_1 = T^{\delta_{m_1}}$ and $l = T^{\delta_l}$ for some $0 \leq \delta_l < \delta_{m_1} < 1$, and $T \to \infty$, it is

$$\hat{G}_y(d, l, m_1) \xrightarrow{p} \hat{G}_x(d, 1, m_1),$$

if either

i.) $l = 1$ and $d_a^0 + d_b^0 > 1$, or $l = 1$, $d_a^0 + d_b^0 < 1$, and $\delta_{m_1} > 1 - d_a^0 - d_b^0$.

ii.) $d_a^0 + d_b^0 < 1$, $l = O\left(T^{(d_a^0+d_b^0-1)/(d_a^0+d_b^0-2)+\upsilon}\right)$ for some $\upsilon > 0$, and $(d_a^0 + d_b^0 - 1)/(d_a^0 + d_b^0 - 2) + \upsilon < ((d_a - d_a^0) + (d_b - d_b^0) + \delta_{m_1})/((d_a - d_a^0) + (d_b - d_b^0) + 1)$.

Here and in the rest of the paper, the superscript 0 denotes the true value of a parameter, for example, d_a^0 is the true memory parameter of series a.

The first part of the theorem shows that for nonstationary long-memory processes the long-memory component always dominates the mean component and no trimming is needed. Here, we can use the procedure by Nielsen and Shimotsu (2007) to determine the fractional cointegration rank even when low-frequency contaminations are present. In the case of stationary long memory, however, this is not the case. Here, the second part of the condition gives the frequency from which onwards the long-memory component becomes dominant. This depends on the true order of integration, which we assumed to be known so far. If this is not the case, a feasible choice would be to trim away $l = \sqrt{T}$ frequencies. We could also estimate d using univariate approaches that are robust to low-frequency contaminations and then choose l based on these estimates. However, unreported simulations indicate that setting $l = \sqrt{T}$ yields superior results.

Nevertheless, we still have to estimate d for determining $\hat{G}_y(d, l, m_1)$. For this purpose, we need an estimator that is robust to low-frequency contaminations and converges with a rate of $\log m$, which is the standard rate for semiparametric estimators. Moreover, as discussed in Robinson and Yajima (2002) and Nielsen and Shimotsu (2007), we require an estimate of d that converges faster than the estimate of G_y such that the effect of estimating d vanishes asymptotically and we cannot rely on multivariate estimators since these require knowledge of the cointegration rank. Possible estimators are those of Iacone (2010), McCloskey and Perron (2013), or Hou and Perron (2014). Denote the bandwidth for estimating d by m for which the following assumption holds.

Assumption B3. For any $\psi > 0$,

$$\frac{m_1^{1/2-\psi} T^{\psi}}{m^{1/2}} + \frac{m^{1+2\psi} \log(m)^2}{T^{2\psi}} \to 0 \text{ as } T \to \infty.$$

Theorem 3. Suppose y_t is generated by (6.3), Assumptions B1 to B3 hold, and let $\hat{d}(m) - d^0 = o(\log m)$, then

$$\hat{G}_y(\hat{d}(m), l, m_1) \xrightarrow{p} \hat{G}_y(d^0, l, m_1).$$

Theorem 3 in conjunction with 2 ii.) and Proposition 3 of Robinson and Yajima (2002) implies that G can be estimated consistently and $\hat{G}_y(\hat{d}(m), l, m_1)$ is asymptotically Gaussian given the additional assumptions made in Robinson and Yajima (2002). Note that these assumptions restrict the series to be stationary. This might seem restrictive but as discussed before, for nonstationary series low-frequency contaminations are not troublesome such that the standard extension by Nielsen and Shimotsu (2007) can be considered.

We can then follow the route of Robinson and Yajima (2002) to estimate the fractional cointegration rank of the series. For the sake of completeness, we will briefly outline the steps. We can test whether two series have equal memory using

$$\hat{T}_{TRE} = \frac{m^{1/2}(\hat{d}_a - \hat{d}_b)}{(1/2(1 - \hat{G}_{ab}^2/(\hat{G}_{aa}\hat{G}_{ab})))^{1/2} + n(T)},$$

where G_{ab} are the respective elements of the estimated matrix $\hat{G}_y(\hat{d}(m), l, m_1)$, d_a and d_b are estimated using a robust estimator as discussed above, and $n(T) > 0$. Consistency of the test follows under the same additional assumptions as in Robinson and Yajima (2002).

If the test indicates the two series to have equal memory, we can then determine whether they are fractionally cointegrated by estimating the fractional cointegration rank. To do so, denote by $q_{1,G}$ the first eigenvalue of $G_y(\hat{d}(m), l, m_1)$ and let $\hat{q}_{1,G}$ denote its empirical counterpart. If $\text{rank}(A) = 2$, we have $q_{1,A} > q_{2,A} > 0$, whereas for $\text{rank}(A) = 1$, it is $q_{1,A} > q_{2,A} = 0$. Define furthermore $\sigma_{v,G} = \sum_{i=1}^{v} q_{i,G}$ and $N(T) > 0$ such that $N(T) + m_1^{-1/2} N(T)^{-1} \to 0$ as $T \to \infty$. We can estimate the fractional cointegration rank by minimizing the loss function

$$L(u) = N(T)(2 - u) - \hat{\sigma}_{2-u,G}.$$

The estimator for the fractional cointegration rank is

$$\widehat{rk}_{TRE} = \arg\min_{u=0,1} L(u).$$

Again, consistency of this estimator follows directly from Robinson and Yajima (2002) given the same additional assumptions.

6.4 Robust Multivariate Local Whittle Estimator

After determining the fractional cointegration rank, we aim to estimate the cointegration vector and the memory parameter robust to potential low-frequency contaminations. In this section, we obtain a robust local Whittle estimator for the parameter θ containing of the memory parameters $d = (d_a, d_b)$ and the possible cointegration vector β.

It should be mentioned that our robust local Whittle estimator depends on the a priori specified fractional cointegration rank as much as the original local Whittle estimator in Robinson (2008). The cointegration rank is needed to know the dimension of the parameter to be estimated. It does not enter the estimation procedure as a nuisance parameter. Therefore, we see the estimation of the cointegration rank in a first step as part of the model specification procedure and find the assumption of a known cointegration rank when it comes to parameter estimation justified. The same assumption is implicitly used by

Robinson (2008) and Shimotsu (2012) and is therefore standard in the literature. A short Monte Carlo analysis underpinning this can be found in the Supplementary Appendix.

The expectation of the periodogram of the contaminated multivariate long-memory process has an additive structure. Consequently, a robust multivariate local Whittle estimator can be constructed in a similar fashion as for the fractional cointegration rank estimator, i.e., by trimming away the frequencies dominated by the low-frequency contamination. Our estimator is then based on the univariate trimmed local Whittle estimator by Iacone (2010). In Section 6.2, we showed that $O_P(Tj^{-2})$ is an upper bound for the pole of the periodogram at the zero frequency for a fairly general class of processes. Therefore, trimming the periodogram by the first $l = \sqrt{T}$ frequencies eliminates the influence of the low-frequency contaminations and leads to a robust estimate of the memory parameter.

We aim to estimate the parameter $\theta = (d, \beta)'$, where we restrict the series to be stationary, i.e., $-1/2 < d_a, d_b < 1/2$. As discussed in Section 6.3, for nonstationary time series trimming is not needed. Our trimmed multivariate local Whittle estimator

$$\hat{\theta} = \arg\min R(\theta)$$

is defined by the objective function using here and in what follows the superscript *tri* to indicate the trimmed version

$$R(\theta) = \log\det \hat{\Omega}^{tri}(\theta) - 2(d_a + d_b)\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j,$$

where

$$\Omega^{tri}(\theta) = \frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}[\Lambda_j(d)BI_y^{tri}(\lambda_j)B'\Lambda_j^*(d)] \quad \text{with } B = \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix}.$$

To show consistency of this estimator, we need to make the following assumptions.

Assumption C1. As $\lambda \to 0_+$,

$$f_{x,ab}(\lambda) = \exp\left(i\pi\left(d_a^0 - d_b^0\right)/2\right)\lambda^{-d_a^0-d_b^0}G_{x,ab}^0 = O\left(\lambda^{-d_a^0-d_b^0}\right),$$

where $f_{x,ab}$ and $G_{x,ab}$ are the respective element of the matrices f_x and G_x of x_t.

Assumption C2. Assumption B1 holds.

Assumption C3. In a neighborhood $(0, \alpha)$ of the origin, $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ is differentiable and

$$\frac{\partial}{\partial\lambda}\,{}_aA(\lambda) = O\left(\lambda^{-1}\|{}_aA(\lambda)\|\right) \text{ as } \lambda \to 0_+,$$

where ${}_aA(\lambda)$ is the a-th row of $A(\lambda)$.

Assumption C4. As $T \to \infty$,

$$\frac{l}{m} + \frac{m}{T} \to 0,$$

where l is a trimming parameter with $l = \max(1, [c_l T^{\delta_l}])$ and $m = \max(l+1, [c_m T^{\delta_m}])$ is the bandwidth parameter with $0 \leq \delta_l < \delta_m < 1$ and $c_l, c_m \in (0, \infty)$.

Assumptions C1, C2, and C3 are analogous to Assumptions A1 to A3 of Lobato (1999) respectively Assumptions 1 to 3 of Shimotsu (2007) and Assumption C4 corresponds to A4 of Shimotsu (2007). We furthermore denote in what follows $\nu^0 = d_b^0 - d_a^0$.

Theorem 4. Suppose y_t is generated by (6.3) and Assumptions C1 to C4 hold with the trimming parameter $l = \sqrt{T}$, it is

$$\hat{d} - d^0 \xrightarrow{P} 0, \qquad \hat{\beta} = \beta^0 + o_P\left(\left(\frac{m}{T}\right)^{\nu^0}\right).$$

As usual, the assumptions required to prove the normality of the estimator are somewhat stronger than those needed for consistency. Here, we assume

Assumption D1. For $\zeta \in (0, 2]$ and as $\lambda \to 0_+$,

$$f_{x,ab}(\lambda) = \exp\left(i\left(\pi - \lambda\right)\left(d_a^0 - d_b^0\right)/2\right)\lambda^{-d_a^0 - d_b^0} G_{x,ab}^0 = O\left(\lambda^{-d_a^0 - d_b^0 + \zeta}\right).$$

Assumption D2. Assumption B1 holds and in addition it holds for $a, b, c, d = 1, 2$, $t = 0, \pm 1, \pm 2, \ldots$ that

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}|\mathfrak{F}_{t-1}) = \mu_{abc} \qquad a.s.$$

and

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}|\mathfrak{F}_{t-1}) = \mu_{abcd} \qquad a.s.,$$

where $|\mu_{abc}| < \infty$ and $|\mu_{abcd}| < \infty$.

Assumption D3. Assumption C3 holds.

Assumption D4. As $T \to \infty$, it holds for any $\tau > 0$

$$\frac{l}{m} + \frac{m^{1+2\tau}(\log m)^2}{T^{2\tau}} + \frac{\log T}{m^\tau} \to 0.$$

Assumption D5. There exists a finite real matrix Q such that

$$\Lambda_j(d^0)^{-1} A(\lambda_j) = Q + o(1) \text{ as } \lambda_j \to 0.$$

These assumptions allow for non-Gaussianity. Assumption D1 and D5 are satisfied by multivariate ARFIMA processes. Assumption D4 is necessary for the Hessian of the objective function of the local Whittle function to converge. It should be mentioned that

Assumption D4 gives a sharp upper bound for the number of frequencies m which can be used for the local Whittle estimator. It is $m = o(T^{0.8})$.

For the robust estimator, we obtain the following results.

Theorem 5. Suppose y_t is generated by (6.3), Assumptions D1 to D5 hold with $T \to \infty$, and $\Delta_T = \text{diag}(\lambda_m^{-\nu^0}, 1, 1)$, then

$$\sqrt{m}\Delta_T(\hat{\theta} - \theta^0) \xrightarrow{d} N(0, \Xi^{-1}),$$

where $\Xi_{aa} = 2\tilde{\mu}[(1-2\nu^0)^{-1} - (1-\nu^0)^{-2}cos^2(\tilde{\gamma}^0)]G_{bb}/G_{aa}$, $\Xi_{ab} = \Xi_{21} = -2\tilde{\mu}\nu^0(1-\nu^0)^{-2}cos(\tilde{\gamma}^0)G_{ab}/G_{aa} + (\pi/2)2\tilde{\mu}(1-\nu^0)^{-1}sin(\tilde{\gamma}^0)(G_{ab}/G_{aa})$, $\Xi_{13} = \Xi_{31} = -\Xi_{ab}$, $\Xi_{bb} = \Xi_{33} = 4 + (\pi^2/4 - 1)2\tilde{\mu}\rho^2$, $\rho = G_{ab}/(G_{aa}G_{bb})^{1/2}$, $\tilde{\mu} = (1-\rho^2)^{-1}$, and $\tilde{\gamma}^0 = (\pi/2)\nu^0$.

Consequently, the estimator is consistent and asymptotically normal with the same limiting variance as the GSE estimator in Robinson (2008). We can therefore robustify the estimator without an asymptotic loss in efficiency. If $G_\mu = 0$, the estimator is reduced to the standard multivariate estimator of Robinson (2008).

6.5 Monte Carlo Simulation

In this section, we show the behavior of our proposed methods in finite samples by means of a simulation study. We first investigate the behavior of our robust fractional cointegration rank estimator and then consider the robust multivariate local Whittle estimator. Our framework as stated in Section 6.2 allows for various forms of low-frequency contaminations. For the ease of the presentation, we present results for nonstationary random level-shift processes with rare shifts in the following as this seems to be the empirical most relevant case and move the qualitatively similar results for stationary random level-shift processes and deterministic trends to Tables 6.8-6.13 in the Supplementary Appendix.

As data generating process (DGP), we consider the following bivariate stationary long-memory process with random level shifts

$$y_{at} = \zeta_a\mu_t + \xi\tilde{\mu}_t + x_t + (1-L)^{-(d-\tilde{d})}u_t \tag{6.5}$$

$$y_{bt} = \zeta_b\mu_t + x_t, \tag{6.6}$$

where

$$\mu_t = \mu_{t-1} + \pi_t\eta_t, \qquad \pi_t \sim B(5/T), \qquad \eta_t \sim N(0,1), \tag{6.7}$$

$$\tilde{\mu}_t = \tilde{\mu}_{t-1} + \tilde{\pi}_t\tilde{\eta}_t, \qquad \tilde{\pi}_t \sim B(5/T), \qquad \tilde{\eta}_t \sim N(0,1), \tag{6.8}$$

$$x_t = (1-L)^{-d}e_t, \qquad \text{and} \qquad \begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right). \tag{6.9}$$

Here, L is the usual lag operator such that $(1-L)^d = \sum_{k=0}^{\infty}\binom{d}{k}(-1)^k L^k$ with $\binom{d}{k} = \frac{d(d-1)(d-2)\dots(d-(k-1))}{k!}$. This model allows for fractional cointegration, distinct structural breaks, and joint structural breaks.

Concerning the fractional cointegration component, we investigate the case of no fractional cointegration with $\tilde{d} = 0$ and the case of fractional cointegration with $\tilde{d} = d$ and the cointegration vector being $\beta = (1, -1)'$. Concerning the low-frequency contamination component, we investigate the situations of no low-frequency contaminations, i.e., $\zeta_a = \zeta_b = \xi = 0$, of distinct structural breaks, i.e., $\zeta_a = 0$ but $\zeta_b = \xi = 1$, and of joint breaks with $\zeta_a = \zeta_b = 1$ and $\xi = 0$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$.

We present results for cross-sectionally uncorrelated ($r = 0$) series with orders of integration of $d = 0.2, 0.4$. Qualitatively similar results for $r = 0.5$ can be found in Tables 6.5 to 6.7 of the Supplementary Appendix. We consider sample sizes of $T = 250, 1000$ with a burn-in period of 250 observations and all presented results are the averages over 5,000 replications.

6.5.1 Fractional Cointegration

We first consider the results for the estimation of the fractional cointegration rank. To put the performance of our estimator into perspective, we also report the results of all other procedures applicable when the series exhibit stationary long memory. This includes the rank estimator by Robinson and Yajima (2002) (RY02) and the fractional cointegration tests by Chen and Hurvich (2006) (CH06) and Souza et al. (2018) (SRF). Parameter values are all chosen according to the authors recommendation.

For our robust estimator, abbreviated TRE in the table, we need to choose l, m, m_1, N, and the type of robust univariate estimator. Unreported simulations indicate that $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}, N = m_1^{-0.2}$, and the univariate estimator by Iacone (2010) yielded the best trade-off between correctly and spuriously identifying fractional cointegration in our simulations. We therefore show the results for this parameter combination in the following and recommend it to be considered in empirical applications.

The results can be found in Table 6.1. In the table, “NO” indicates no low-frequency contaminations, “DIS” means that the series exhibit distinct breaks, and “COB” refers to joint breaks. Furthermore, “TRUE” means that the series are fractionally cointegrated, whereas “FALSE” indicates that they are not. We then state the mean estimated cointegration rank for the rank estimators, which should be 1 for “TRUE” and 0 otherwise, and the mean rejection rate for the tests, which should be 1 for “TRUE“ and 0.05 (the significance level) otherwise as all procedures test the null of no fractional cointegration.

Table 6.1 shows that our procedure works well for all of the considered scenarios. It correctly identifies fractional cointegration respectively no fractional cointegration in most

		Cointegration	TRUE				FALSE			
d	T	Breaks	TRE	RY02	CH06	SRF	TRE	RY02	CH06	SRF
0.2	250	NO	1.00	0.99	0.29	0.09	0.00	0.00	0.16	0.00
		DIS	0.96	0.35	0.24	0.10	0.00	0.00	0.51	0.16
		COB	1.00	1.00	0.81	0.66	0.00	0.02	0.75	0.37
	1000	NO	1.00	1.00	0.63	0.29	0.00	0.00	0.18	0.00
		DIS	0.99	0.23	0.44	0.54	0.00	0.00	0.63	0.61
		COB	1.00	1.00	0.97	0.92	0.00	0.00	0.90	0.72
0.4	250	NO	1.00	1.00	0.80	0.39	0.00	0.00	0.23	0.00
		DIS	0.98	0.68	0.18	0.10	0.00	0.00	0.38	0.04
		COB	1.00	1.00	0.93	0.71	0.00	0.00	0.52	0.10
	1000	NO	1.00	1.00	1.00	0.90	0.00	0.00	0.19	0.01
		DIS	1.00	0.78	0.22	0.42	0.00	0.00	0.41	0.13
		COB	1.00	1.00	1.00	0.98	0.00	0.00	0.57	0.22

Table 6.1: Simulation results fractional cointegration rank

Reported is the mean estimated fractional cointegration rank for a bivariate fractionally integrated system. The DGP is based on Equations (6.5) to (6.9) with $r = 0$. In case of fractional cointegration, $\beta = (-1, 1)'$ and $b = d$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. Our estimator (TRE) is considered with $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}$, $N = m_1^{-0.2}$, and the univariate estimator by Iacone (2010) to estimate d. For the procedures by Robinson and Yajima (2002) (RY02), Chen and Hurvich (2006) (CH06), and Souza et al. (2018) (SRF), parameter values are chosen according to the authors recommendation.

of the cases. In Section 6.3, we mentioned that two scenarios are of particular importance, the case when no fractional cointegration is present but joint breaks as standard procedure might spuriously indicate fractional cointegration, and the case of fractional cointegration but distinct breaks as standard procedure might have problems detecting the fractional cointegration relation in this case. The table shows that our estimator works well in both cases with minor distortions for the first case in a small sample size of $T = 250$. When increasing the sample size, these vanish completely.

In contrast to this, all other procedures show problems in at least one of the two cases. In the first case, the rejection rates of CH06 and SRF increase with increasing T implying that asymptotically the tests always indicate fractional cointegration in the case of joint breaks, no matter if the series truly are fractionally cointegrated. RY02 has serious problems in the second case where, at least for $d = 0.2$, the estimated cointegration rank decreases with increasing T implying that asymptotically the estimator always indicate no fractional cointegration in the case of distinct breaks.

Consequently, there is a risk in empirical applications that these procedures miss a fractional cointegration relation due to structural breaks or falsely indicate a fractional

		Estimator	TMLW				GSE			
			Bias		RMSE		Bias		RMSE	
d	T	Breaks	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
0.2	250	NO	-0.03	-0.02	0.19	0.20	-0.02	-0.02	0.08	0.08
		DIS	0.02	0.01	0.20	0.20	0.18	0.18	0.22	0.22
		COB	0.01	0.01	0.19	0.20	0.17	0.17	0.20	0.20
	1000	NO	-0.02	-0.01	0.09	0.09	-0.01	-0.01	0.04	0.04
		DIS	0.01	0.01	0.09	0.09	0.17	0.17	0.19	0.19
		COB	0.01	0.01	0.09	0.09	0.16	0.16	0.18	0.18
0.4	250	NO	-0.04	-0.04	0.20	0.20	-0.02	-0.02	0.08	0.08
		DIS	-0.01	-0.01	0.19	0.20	0.08	0.08	0.12	0.12
		COB	-0.01	-0.01	0.19	0.19	0.07	0.07	0.11	0.11
	1000	NO	-0.02	-0.02	0.09	0.09	-0.01	-0.01	0.04	0.04
		DIS	-0.01	-0.01	0.09	0.09	0.06	0.06	0.09	0.09
		COB	-0.01	-0.01	0.09	0.09	0.06	0.06	0.08	0.08

Table 6.2: Simulation results order of integration — no fractional cointegration

Reported are bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally integrated system. The DGP is based on Equations (6.5) to (6.9) with $r = 0$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. We use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

cointegration relation due to joint breaks in the series. In contrast, our procedure delivers the required robustness to correctly detect whether fractional cointegration is present.

6.5.2 Order of Integration

After providing evidence that we can robustly determine the fractional cointegration rank also in small samples, we now focus on robust estimation of the memory parameter. We consider the same DGP as before and again use $m = T^{0.75}$ and $l = T^{0.5}$ for robust estimation. Furthermore, we state the results for the standard multivariate estimator by Shimotsu (2007) and Robinson (2008). For comparison, we assume the fractional cointegration rank to be known. Tables 6.14 and 6.15 in the Supplementary Appendix show the qualitatively similar results when first estimating the fractional cointegration rank and then estimating the order of integration.

Table 6.2 has bias and RMSE of the estimators when no fractional cointegration is present and Table 6.3 states bias and RMSE when fractional cointegration is present.

In the case of no fractional cointegration, we estimate the two memory parameters d_1 and d_2 of the two time series. Table 6.2 shows that in this case bias and RMSE of our estimator (TMLW) are small in all scenarios and that they decrease with increasing

	Estimator		TMLW						GSE					
			Bias			RMSE			Bias			RMSE		
d	T	Breaks	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$
0.2	250	NO	-0.02	-0.03	-0.28	0.20	0.19	2.75	-0.02	-0.02	0.01	0.08	0.07	1.15
		DIS	0.08	0.01	0.11	0.22	0.19	2.82	0.37	0.17	1.32	0.39	0.21	3.88
		COB	-0.02	0.01	-0.21	0.20	0.19	2.65	-0.02	0.18	0.00	0.08	0.22	0.32
	1000	NO	-0.01	-0.02	-0.02	0.09	0.09	1.27	-0.01	-0.01	0.00	0.04	0.04	0.17
		DIS	0.06	0.01	0.10	0.12	0.09	1.84	0.37	0.16	1.50	0.39	0.18	3.92
		COB	-0.01	0.01	-0.03	0.09	0.09	1.18	-0.01	0.17	0.00	0.04	0.19	0.06
0.4	250	NO	-0.02	-0.04	-0.12	0.19	0.20	1.99	-0.02	-0.01	0.00	0.08	0.08	0.18
		DIS	0.08	-0.01	0.05	0.22	0.19	2.16	0.38	0.07	0.58	0.40	0.12	3.15
		COB	-0.02	-0.02	-0.13	0.20	0.20	1.84	-0.02	0.08	0.00	0.08	0.13	0.09
	1000	NO	-0.01	-0.02	-0.01	0.09	0.09	0.25	-0.01	-0.01	0.00	0.04	0.04	0.05
		DIS	0.06	-0.01	-0.02	0.12	0.09	0.59	0.39	0.05	0.46	0.40	0.08	2.88
		COB	-0.01	-0.01	0.00	0.09	0.09	0.22	-0.01	0.06	0.00	0.04	0.09	0.03

Table 6.3: Simulation results order of integration — fractional cointegration Reported are bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally cointegrated system with cointegration vector $\beta = (1, -1)'$. The DGP is based on Equations (6.5) to (6.9) with $r = 0$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. We use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

sample size. If we consider the standard estimator (GSE), we can see that it is upward biased in case of low-frequency contaminations. The bias is large for $d = 0.2$ and does not seem to vanish asymptotically when increasing T. However, it does decrease when increasing d as implied by Theorem 2. The table further reveals that the variance of the standard estimator is lower than those of our procedure since more frequencies are used for estimation. When increasing T, the difference becomes smaller as implied by Theorem 5.

In the case of fractional cointegration, we estimate the memory parameter d, the reduction of memory through the fractional cointegration relation $d - \tilde{d}$, and the cointegration vector β. As Table 6.3 shows, for the robust estimation of the memory parameter the same conclusion as for the case without a cointegration relation can be drawn. Bias and RMSE are small even for $T = 250$ and they decrease with increasing sample size. Concerning the estimation of $\widehat{d - \tilde{d}}$, it can be observed that our estimator works well when considering the processes without breaks or with joint breaks. For the process with distinct breaks, there is a positive bias which vanishes with increasing sample size. We could decrease this bias by increasing l which, however, comes at the cost of an increased variance of the estimator. Concerning the estimate of the cointegration vector β, we can see that the estimate exhibits some variation in small samples resulting in rather large RMSEs.

As in the case of no fractional cointegration, the standard estimator shows a substantial positive bias for d, which decreases in d but not in T, when low-frequency contaminations are present. The estimate of $d - \tilde{d}$ is accurate in the case of joint breaks but enormously upward biased in the case of distinct breaks. The same conclusion can also be drawn for the estimate of β. As before, it can again be observed that the variance of the standard estimator is smaller than of our procedure resulting in smaller RMSEs when no low-frequency contaminations are present.

To summarize, the standard estimators are upward biased in the case of low-frequency contaminations. In contrast, our estimator is robust to low-frequency contaminations in the case of no fractional cointegration as well as in the case of fractional cointegration. The price for this robustness is an increased variance which might be problematic when estimating β in small samples, as the variance of the standard estimator is already large in this case. When increasing T, this problem disappears making our estimator well suited for estimating the order of integration and the cointegration vector.

6.6 Empirical Example

To demonstrate the empirical relevance of our procedures, we consider an example investigating the daily realized beta of two American stocks, namely Chevron (CVX) and ExxonMobil (XOM), relative to the S&P 500 between January 1996 and February 2017 ($T = 5,238$). Realized betas measure the systematic risk of a stock and are defined as the realized covariance of the stock with the market divided by the realized variance of the market. To construct these series, we use 5-minute returns obtained from the Thomson Reuters Tick History database. These returns are cleaned following the recommendations of Barndorff-Nielsen et al. (2009) to account for the typical high frequency data quality issues.

To give a first graphical impression, Figure 6.1 plots the realized betas and the corresponding autocorrelation function and periodogram for the two stocks. It can be seen that the autocorrelation function and periodogram indicate the series to be highly persistent with significant positive correlation even after 200 lags and a pole at the origin.

Despite such evidence for persistence, realized betas have been sparsely investigated concerning their order of integration so far. A noteworthy study in this context is the one by Andersen et al. (2006), who find that quarterly betas in the time period 1969 until 1999 were best described by a process with $d \approx 0.2$. Due to the small number of observations, their analysis is based on graphical investigation rather than consistent and robust estimation of the memory parameter. The two constituents of realized beta, realized variance and realized covariance, on the other hand, have been investigated more extensively. Depending on the investigated asset and time period, it is found that realized variances can be best described by pure long-memory processes (e.g., Andersen et al., 2003) or a combination of long-memory process and shift process (e.g., Liu and Maheu, 2007). For realized covariances, although less considered in the literature, a similar conclusion might be drawn (e.g., Asai and McAleer, 2015).

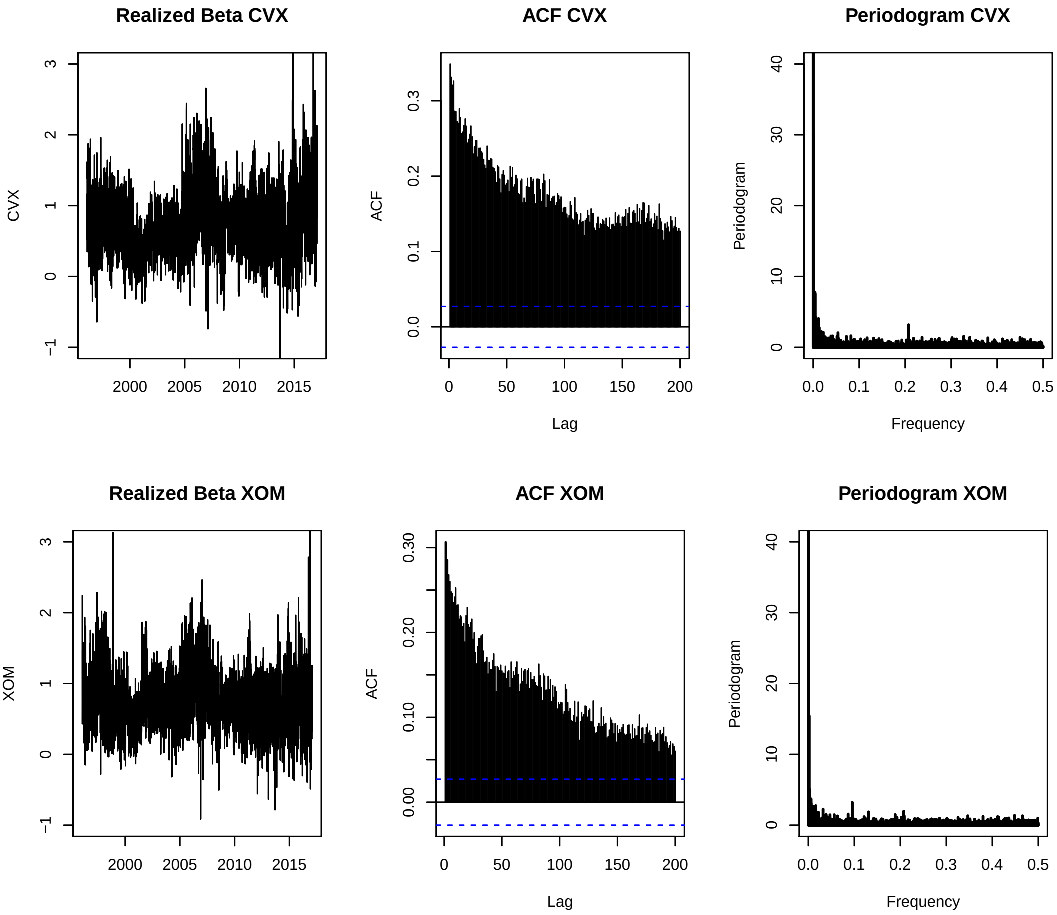

Figure 6.1: Realized beta plots.
Left: daily realized betas of Chevron (CVX) and ExxonMobile (XOM) relative to the S&P 500 from January 1996 to February 2017. Middle: corresponding autocorrelation functions excluding lag zero. Right: corresponding periodograms.

To summarize, there is evidence that realized betas are fractionally integrated. Furthermore, since we investigate two companies who both mainly operate in the same industry, it seems reasonable to assume that they face the same relation to the systematic risk factor. This would imply a fractional cointegration relation between the two series. However, the series might also exhibit low-frequency contaminations caused by structural breaks in the realized variance or the realized covariance. Therefore, estimating the memory of realized betas and thereby estimating the fractional cointegration relation should be done using our robust methods.

To demonstrate this, we test whether the two series exhibit equal order of integration and then estimate the fractional cointegration rank of the two series using the standard procedure by Robinson and Yajima (2002) and our robust procedure. As mentioned in Section 6.3, for the test we need to decide for a robust univariate estimator. Here, we consider the estimator by Iacone (2010), the results are, however, qualitatively similar when considering the estimators by McCloskey and Perron (2013) or Hou and Perron (2014).

	$\hat{T}_{RY}$	$\hat{T}_{TRE}$	$\widehat{rk}_{RY}$	$\widehat{rk}_{TRE}$	MLWS	GSE	TMLW	$\hat{\hat{d}}$	$\hat{\beta}$
Chevron	0.21	0.40	0	1	3.84	0.346	0.180	0.059	0.491
ExxonMobile						0.375			

Table 6.4: Estimation results.
$\hat{T}_{RY}$ and $\widehat{rk}_{RY}$ correspond to the test statistics for equality of d and the fractional cointegration rank estimator by Robinson and Yajima (2002) and $\hat{T}_{TRE}$ and $\widehat{rk}_{TRE}$ to our robust procedures. In analogy, GSE corresponds to the d estimate by the standard multivariate local Whittle estimator from Shimotsu (2007) and Robinson (2008) and TMLW to the estimate by our trimmed estimator for which we also state the estimated reduction in memory $\hat{b}$ and the estimated cointegration vector $\hat{\beta}$. Finally, MLWS corresponds to the test statistic of the multivariate test for spurious long memory by Sibbertsen et al. (2018). For our procedures, we employ the parameter combination used in Section 6.5. For the procedures by Robinson and Yajima (2002), Shimotsu (2007) and Robinson (2008), and Sibbertsen et al. (2018) we consider the parameter combinations recommended by the authors.

We are then able to compute the multivariate local Whittle estimator by Shimotsu (2007) and Robinson (2008) and our robust multivariate local Whittle estimator. Additionally, we apply the multivariate MLWS test by Sibbertsen et al. (2018) to test for low-frequency contaminations. For our methods, we use the parameter combinations recommended in Section 6.5 and for the other methods we use the parameter combinations recommended by the authors of the procedures. The results are displayed in Table 6.4.

For estimating the fractional cointegration rank, we first need to test whether the series exhibit an equal order of integration. Otherwise, fractional cointegration can be excluded right away. The table shows that both, robust and standard test statistic are clearly below the five percent critical value of 1.96. Therefore, the null hypothesis that the series are equally integrated cannot be rejected meaning that it is sensible to investigate whether the two series are fractionally integrated. It can be seen that the procedure by Robinson and Yajima (2002) indicates the cointegration rank to be zero, i.e., no fractional cointegration, while our robust procedure estimates it to be 1. Furthermore, the multivariate MLWS by Sibbertsen et al. (2018) shows a test statistic of 3.84 which is above the one percent critical value of 1.517 implying that the series exhibit low-frequency contaminations. These then dominate the G matrix in the lower frequencies letting the matrix appear to have full rank which makes the standard estimator by Robinson and Yajima (2002) unable to detect the cointegration relation. In contrast, our procedure trims those frequencies and is therefore robust to the contaminations.

If we then estimate the order of integration using the standard multivariate local Whittle estimator, our error is twofold. First, we have a positive bias of the estimates caused by the low-frequency contaminations. Second, we ignore the fact that the two series are fractionally cointegrated causing a bias as well. This reflects in the estimates made by the standard multivariate local Whittle estimator and our trimmed estimate stated in column four and five of Table 6.4. While the standard procedure estimates the memory parameters to be 0.35 respectively 0.38, our robust estimator yields a significantly lower value of 0.18 for both series which is in line with the considerations by Andersen et al. (2006) that realized betas have a d of approximately 0.2. The estimator further states the reduction in memory to be 0.06 and the fractional cointegration vector to be $(1, -0.49)'$.

6.7 Conclusion

It is well known that low-frequency contaminations bias inference on the order of integration of a series. Therefore, several authors have proposed robust approaches that yield valid inference also in the case of low-frequency contaminations. These are, however, all univariate. As working in multivariate systems often yields efficiency gains, we suggest a multivariate local Whittle estimator robust to low-frequency contaminations in this paper. This does not only yield an estimate of the order of integration of the series but also of the cointegration vector.

Our estimator requires a priori knowledge of the fractional cointegration rank. We provide theoretical as well as simulation-based evidence that low-frequency contaminations bias inference on the fractional cointegration rank. On the one hand, common low-frequency contaminations cause spurious fractional cointegration. On the other hand, distinct low-frequency contaminations cause tests and estimators to falsely indicate the absence of fractional cointegration. We therefore also propose an estimator of the fractional cointegration rank robust to low-frequency contaminations. Using this estimator, we find that the realized beta series of ExxonMobil and Chevron are fractionally cointegrated. As the series exhibit low-frequency contaminations, non-robust procedures were unable to detect this relation.

Appendix

Before proving Lemmas 1, 2, and 3, we need two auxiliary lemmas. For the structural-change processes in (6.1) and (6.2), we have the following result.

Lemma 4. The discrete Fourier transform (DFT) of the process in (6.1) can be represented as

$$w_\mu(\lambda_j) = -\frac{1}{\sqrt{2\pi T}} \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda_j),$$

and that of the process in (6.2) can be represented as

$$w_\mu(\lambda_j) = -\frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{K} \mu_k \left(D_{T_k-1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \right),$$

where $D_{T_k}(\lambda_j) = \sum_{t=1}^{T_k} e^{i\lambda_j t}$ is a version of the Dirichlet kernel.

Note that Lemma 4 is completely algebraic and we do not impose any conditions on the $\Delta\mu_k$, μ_k, or T_k.

Proof of Lemma 4:

From (6.1), we have

$$\begin{aligned} w_\mu(\lambda_j) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left\{ \mu_0 + \sum_{k=1}^{K} 1(t \geq T_k)\Delta\mu_k \right\} e^{i\lambda_j t} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda_j t} + \sum_{k=1}^{K} \Delta\mu_k \sum_{t=1}^{T} 1(t \geq T_k) e^{i\lambda_j t} \right\}. \end{aligned}$$

Here,

$$\sum_{t=1}^{T} 1(t \geq T_k) e^{i\lambda_j t} = \sum_{t=T_k}^{T} e^{i\lambda_j t} = \sum_{t=1}^{T} e^{i\lambda_j t} - \sum_{t=1}^{T_k} e^{i\lambda_j t} = D_T(\lambda_j) - D_{T_k}(\lambda_j).$$

Therefore,

$$\begin{aligned} w_\mu(\lambda_j) &= \frac{1}{\sqrt{2\pi T}} \left\{ D_T(\lambda_j)\mu_0 + \sum_{k=1}^{K} \Delta\mu_k \left[D_T(\lambda_j) - D_{T_k}(\lambda_j) \right] \right\} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \left[\mu_0 + \sum_{k=1}^{K} \Delta\mu_k \right] D_T(\lambda_j) - \sum_{k=1}^{K} \Delta\mu_k D_{T_k}(\lambda_j) \right\}. \end{aligned}$$

Furthermore, we have

$$D_T(\lambda_j) = \frac{e^{i(T+1)\lambda_j} - e^{i\lambda_j}}{e^{i\lambda_j} - 1} = e^{i(T-1)\lambda_j/2} \frac{\sin(T\lambda_j/2)}{\sin(\lambda_j/2)}, \tag{6.10}$$

cf. Beran et al. (2013), p. 327. Note that $\lambda_j T = 2\pi j$, $e^{i(T+1)\lambda_j} = e^{i\lambda_j T} e^{i\lambda_j}$, and $e^{i2\pi j} = \cos(2\pi j) + i\sin(2\pi j) = \cos(2\pi) + i\sin(2\pi) = 1$. Therefore, $D_T(\lambda_j) = \frac{e^{i\lambda_j} - e^{i\lambda_j}}{e^{i\lambda_j} - 1} = 0$, which proves the first part of the lemma.

Similarly, for the second part of the lemma, we have from (6.2) that

$$\begin{aligned} w_\mu(\lambda_j) &= \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \left\{ \mu_0 + \sum_{k=0}^{K} \mu_k 1(T_{k-1} \leq t < T_k) \right\} e^{i\lambda_j t} \\ &= \frac{1}{\sqrt{2\pi T}} \left\{ \mu_0 \sum_{t=1}^{T} e^{i\lambda_j t} + \sum_{k=0}^{K} \mu_k \sum_{t=1}^{T} 1(T_{k-1} \leq t < T_k) e^{i\lambda_j t} \right\}. \end{aligned}$$

Here,

$$\begin{aligned} &\sum_{t=1}^{T} 1(T_{k-1} \leq t < T_k) e^{i\lambda_j t} = \sum_{t=1}^{T} \{1(T_k > t) - 1(T_{k-1} \geq t)\} e^{i\lambda_j t} \\ &= \sum_{t=1}^{T_k - 1} e^{i\lambda_j t} - \sum_{t=1}^{T_{k-1}} e^{i\lambda_j t} = D_{T_k - 1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \end{aligned}$$

Therefore, since $D_T(\lambda_j) = 0$, we have

$$w_\mu(\lambda_j) = \frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{K} \mu_k \left\{ D_{T_k - 1}(\lambda_j) - D_{T_{k-1}}(\lambda_j) \right\}.$$

□

Since Lemma 4 implies that the properties of the DFT and thus the properties of the periodogram of a structural-change process are directly related to those of the Dirichlet kernel, the following lemma provides an approximation for the Dirichlet kernel at frequencies local to zero.

Lemma 5. We have for $T_k/T = \delta_k \in (0,1)$ and $j/T \to 0$,

$$\begin{aligned} D_{T_k}(\lambda_j) =& T j^{-1} \frac{\sin(2\delta_k \pi j)}{2\pi} + \sin^2(\pi j \delta_k) \\ &+ i \left[T j^{-1} \frac{\sin^2(\pi \delta_k j)}{\pi} - \frac{1}{2}\sin(2\pi\delta_k j) \right] + O_P(jT^{-1}). \end{aligned}$$

Clearly, from Lemma 5, both the real and the imaginary part of the Dirichlet kernel are $O(Tj^{-1})$ for deterministic δ_k and $O_p(Tj^{-1})$ if any of the δ_k are stochastic. Furthermore, the order is exact. Again, this is an approximation based on a Laurent expansion that holds irrespective of the stochastic properties of the T_k.

Proof of Lemma 5:

From the second expression in (6.10) in the proof of Lemma 4, we can decompose the real and the imaginary parts of the DFT at the Fourier frequencies $\lambda_j = 2\pi j/T$ as follows,

$$\begin{aligned} D_{T_k}(\lambda_j) &= \frac{e^{i(T_k-1)\lambda_j/2}\sin(T_k\lambda_j/2)}{\sin(\lambda_j/2)} \\ &= \frac{[\cos((T_k-1)\lambda_j/2) + i\sin((T_k-1)\lambda_j/2)]\sin(T_k\lambda_j/2)}{\sin(\lambda_j/2)}. \end{aligned}$$

It follows by the sum-to-product identities that

$$\begin{aligned} D_{T_k}(\lambda_j) &= \frac{\sin(\frac{T_k-1}{T}\pi j + \frac{T_k}{T}\pi j) - \sin(\frac{T_k-1}{T}\pi j - \frac{T_k}{T}\pi j)}{2\sin(\frac{\pi j}{T})} \\ &\quad + i\frac{\cos(\frac{T_k-1}{T}\pi j - \frac{T_k}{T}\pi j) - \cos(\frac{T_k-1}{T}\pi j + \frac{T_k}{T}\pi j)}{2\sin(\frac{\pi j}{T})} \\ &= \frac{\sin(2\delta_k\pi j - \frac{\pi j}{T}) + \sin(\frac{\pi j}{T})}{2\sin(\frac{\pi j}{T})} + i\frac{\cos(\frac{\pi j}{T}) - \cos(2\delta_k\pi j - \frac{\pi j}{T})}{2\sin(\frac{\pi j}{T})}. \end{aligned}$$

By a Laurent series approximation around $\lambda_j = 0$, we obtain

$$\begin{aligned} D_{T_k}(\lambda_j) =& Tj^{-1}\frac{\sin(2\delta_k\pi j)}{2\pi} + \sin^2(\pi j\delta_k) + O_P(jT^{-1}) \\ &+ i\left[Tj^{-1}\frac{\sin^2(\pi\delta_k j)}{\pi} - \frac{1}{2}\sin(2\pi\delta_k j) + O_P(jT^{-1})\right], \end{aligned}$$

where the Laurent series is obtained from separate Taylor approximations for each of the trigonometric functions. This proves the lemma. □

We can now prove the Lemmas 1, 2, and 3.

Proof of Lemma 1:

For $\mu_t = h(t/T, T)$, by combining Lemma 4 with Lemma 5, we have

$$\begin{aligned} I_\mu(\lambda_j) =& \left| -\frac{1}{\sqrt{2\pi T}}\sum_{t=1}^{T}\Delta\mu_t D_t(\lambda_j)\right|^2 \\ =& (2\pi T)^{-1}\left\{\left[T\sum_{t=1}^{T}\frac{\Delta\mu_t\sin(2\pi jt/T)}{2\pi j} + \sum_{t=1}^{T}\Delta\mu_t\sin^2(\pi jt/T) + T^{-1}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right. \\ &\left. + \left[T\sum_{t=1}^{T}\Delta\mu_t\frac{\sin^2(\pi jt/T)}{\pi j} - 1/2\sum_{t=1}^{T}\Delta\mu_t\sin(2\pi jt/T) + T^{-1}\sum_{t=1}^{T}\Delta\mu_t O_P(j)\right]^2\right\}. \end{aligned}$$

Factoring out T from the square brackets gives

$$2\pi I_\mu(\lambda_j)T^{-1} = \left\{ \left[\sum_{t=1}^{T} \frac{\Delta\mu_t \sin(2\pi jt/T)}{2\pi j} + T^{-1} \sum_{t=1}^{T} \Delta\mu_t \sin^2(\pi jt/T) + T^{-2} \sum_{t=1}^{T} \Delta\mu_t O_P(j) \right]^2 \right.$$
$$\left. + \left[\sum_{t=1}^{T} \Delta\mu_t \frac{\sin^2(\pi jt/T)}{\pi j} - \frac{1}{2T} \sum_{t=1}^{T} \Delta\mu_t \sin(2\pi jt/T) + T^{-2} \sum_{t=1}^{T} \Delta\mu_t O_P(j) \right]^2 \right\}.$$

Now, using $\Delta\mu_t = h\left(t/T, T\right) - h\left((t-1)/T, T\right)$, we have

$$\lim_{T\to\infty} \Delta\mu_t T = \lim_{T\to\infty} \frac{\partial h(t/T,T)}{\partial(t/T)},$$

so that

$$2\pi I_\mu(\lambda_j)T^{-1} \sim$$
$$\left\{ \left[\frac{1}{2\pi jT} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \sin(2\pi jt/T) + \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \frac{1}{T^2} \sin^2(\pi jt/T) + \frac{1}{T^3} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} O_P(j) \right]^2 \right.$$
$$\left. + \left[\sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \frac{\sin^2(\pi jt/T)}{\pi jT} - \frac{1}{2} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} \frac{1}{T^2} \sin(2\pi jt/T) + \frac{1}{T^3} \sum_{t=1}^{T} \frac{\partial h(t/T,T)}{\partial(t/T)} O_P(j) \right]^2 \right\},$$

where $A \sim B$ means that the ratio of A and B converge to 1 as $T \to \infty$.

By the definition of a Riemann integral,

$$2\pi I_\mu(\lambda_j)T^{-1} \sim$$
$$\left\{ \left[\frac{1}{2\pi j} \int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js)ds + \frac{1}{T} \int_0^1 \frac{\partial h(s,T)}{\partial s} \sin^2(\pi js)ds + T^{-2} \int_0^1 \frac{\partial h(s,T)}{\partial s} O_P(j)ds \right]^2 \right.$$
$$\left. + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \frac{\sin^2(\pi js)}{\pi j} ds - \frac{1}{2T} \int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js)ds + T^{-2} \int_0^1 \frac{\partial h(s,T)}{\partial s} O_P(j)ds \right]^2 \right\}.$$

Clearly, both parts of this expression are dominated by the first term in the respective square bracket, such that

$$I_\mu(\lambda_j) \sim \frac{T}{8\pi^3 j^2} \left\{ \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} \sin(2\pi js)ds \right]^2 + \left[\int_0^1 \frac{\partial h(s,T)}{\partial s} (1 - \cos(2\pi js))ds \right]^2 \right\},$$

which finishes our proof. □

Proof of Lemma 2:

First, by (6.10) in the proof of Lemma 4, we have

$$D_{T_k}(\lambda_j) D^*_{T_u}(\lambda_j) = e^{i(T_k - T_u)\lambda_j/2} \frac{\sin(T_k\lambda_j/2)\sin(T_u\lambda_j/2)}{\sin^2(\lambda_j/2)} = 2e^{i\pi j(\delta_k - \delta_u)} \frac{\sin(\delta_k \pi j)\sin(\delta_u \pi j)}{1 - \cos(\lambda_j)}.$$

By a Laurent expansion around $\lambda_j = 0$, we have

$$D_{T_k}(\lambda_j)D^*_{T_u}(\lambda_j) = 2e^{i\pi j(\delta_k-\delta_u)}\frac{\sin(\delta_k\pi j)\sin(\delta_u\pi j)}{1-[1-2\pi^2(j/T)^2+O((j/T)^4)]}$$
$$= \frac{T^2}{\pi^2 j^2}e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j) + O_P(1). \tag{6.11}$$

In particular, for the case when $T_k = T_u = t$, we obtain

$$D_t(\lambda_j)D^*_t(\lambda_j) = \frac{T^2}{2\pi^2 j^2}(1-\cos(2\pi t j/T)) + O(1). \tag{6.12}$$

Furthermore, we have

$$\sum_{k=1}^{K}\Delta\mu_k D_{T_k}(\lambda_j) = \sum_{t=1}^{T}\Delta\mu_t D_t(\lambda_j), \quad \text{where} \quad \Delta\mu_t = \begin{cases}\Delta\mu_k, & \text{if } t = T_k\\ 0, & \text{otherwise.}\end{cases}$$

In addition to that,

$$I_\mu(\lambda_j) = A + B = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D^*_s(\lambda_j)$$
$$= \frac{1}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 D_t(\lambda_j)D^*_t(\lambda_j) + \frac{1}{2\pi T}\sum_{t\neq s}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D^*_s(\lambda_j).$$

Consequently, we have for term A and from (6.12) above

$$A = \frac{1}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2 D_t(\lambda_j)D^*_t(\lambda_j) = \frac{T}{4\pi^3 j^2}\sum_{t=1}^{T}(\Delta\mu_t)^2(1-\cos(2\pi j t/T)) + \frac{O(1)}{2\pi T}\sum_{t=1}^{T}(\Delta\mu_t)^2$$
$$= \frac{T}{4\pi^3 j^2}\left\{\sum_{t=1}^{T}(\Delta\mu_t)^2 - \sum_{t=1}^{T}(\Delta\mu_t)^2\cos(2\pi j t/T)\right\} + O_P(1). \tag{6.13}$$
$$= \frac{T}{4\pi^3 j^2}\left\{\sum_{k=1}^{K}(\Delta\mu_k)^2 - \sum_{k=1}^{K}(\Delta\mu_k)^2\cos(2\pi j\delta_k)\right\} + O_P(1). \tag{6.14}$$

To deal with term B, we revert back to the original representation in which the sum is random and write

$$B = \frac{1}{2\pi T}\sum_{t\neq s}\Delta\mu_t\Delta\mu_s D_t(\lambda_j)D^*_s(\lambda_j) = \frac{1}{2\pi T}\sum_{k\neq u}\Delta\mu_k\Delta\mu_u D_{T_k}(\lambda_j)D^*_{T_u}(\lambda_j),$$

where $k, l = 1, \ldots, K$. Similar to the approach above, we have from (6.11)

$$B = \frac{T}{2\pi^3 j^2}\sum_{k\neq u}\Delta\mu_k\Delta\mu_u e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j) + \frac{1}{2\pi T}\sum_{k\neq u}\Delta\mu_k\Delta\mu_u O_P(1). \tag{6.15}$$

The first part of the lemma i.) follows immediately from (6.14) and (6.15).

For the second part of the lemma, from Assumption A4 we have $E[K] = E[p_t T] = \tilde{p}T^{1-\alpha}$ and, from (6.13) and (6.14),

$$
\begin{aligned}
E[A] &= \frac{T}{4\pi^3 j^2}\left\{E\left[\sum_{k=1}^{K}(\Delta\mu_k)^2\right] - E[\Delta\mu_t^2]\sum_{t=1}^{T}\cos(2\pi jt/T)\right\} + O(1) \\
&= \frac{T}{4\pi^3 j^2}E\left[\sum_{k=1}^{K}(\Delta\mu_k)^2\right] + O(1),
\end{aligned}
$$

since $\sum_{t=1}^{T}\cos(2\pi jt/T) = 0$.

Now, from Assumption A3 we have $E[(\Delta\mu_k)^2] = \sigma_\Delta^2 T^{-\beta}$ so that by the generalized Wald identity of Brown (1974) and Assumption A4

$$
E[A] = \frac{T}{4\pi^3 j^2}E[K]E[(\Delta\mu_k)^2] = \frac{\tilde{p}\sigma_\Delta^2 T^{2-\alpha-\beta}}{4\pi^3 j^2} + o(1).
$$

Similarly, from (6.15)

$$
E[B] = \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}\Delta\mu_k\Delta\mu_u e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)\right] + \frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}\Delta\mu_k\Delta\mu_u\right]
$$

and by the generalized Wald identity of Brown (1974) in conjunction with Assumption A5

$$
\begin{aligned}
E[B] =& \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}E\left[\Delta\mu_k\Delta\mu_u\right]E\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)\right]\right] \\
&+ \frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}E\left[\Delta\mu_k\Delta\mu_u\right]\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
|E[B]| \leq& \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}\left|E\left[\Delta\mu_k\Delta\mu_u\right]\right|\left|E\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j)\right]\right|\right] \\
&+ \left|\frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}E\left[\Delta\mu_k\Delta\mu_u\right]\right]\right| \\
\leq& \frac{T}{2\pi^3 j^2}E\left[\sum_{k\neq u}\left|E\left[\Delta\mu_k\Delta\mu_u\right]\right|\right] + \frac{|O(1)|}{2\pi T}E\left[\sum_{k\neq u}\left|E\left[\Delta\mu_k\Delta\mu_u\right]\right|\right].
\end{aligned}
$$

Assumption A5 combined with Assumptions A3 and A4 implies that

$$
E\left[\sum_{k\neq u}|E[\Delta\mu_k\Delta\mu_u]|\right] = 2E\left[\sum_{k=2}^{K}\sum_{\tau=1}^{k-1}|E[\Delta\mu_k\Delta\mu_{k-\tau}]|\right] \leq 2E\left[K\right]Var(\Delta\mu_k)\tilde{C} = 2\tilde{p}\tilde{C}\sigma_\Delta^2 T^{1-\alpha-\beta},
$$

so that $|E[B]| \leq \frac{\tilde{p}\sigma_\Delta^2\tilde{C}}{\pi^3 j^2}T^{2-\alpha-\beta} + \left|O(T^{-\alpha-\beta})\right|$. □

Proof of Lemma 3:

From $w_\mu(\lambda_j) = -\frac{1}{\sqrt{2\pi T}}\sum_{k=0}^{K}\mu_k[D_{T_k-1}(\lambda_j) - D_{T_{k-1}}(\lambda_j)]$, as shown in Lemma 4, we have

$$\begin{aligned} I_\mu(\lambda_j) = \tilde{A} + \tilde{B}& \\ =&\frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2[D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j) - D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j) - D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j) \\ &+ D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j)] + \frac{1}{2\pi T}\sum_{k\neq u}\mu_k\mu_u[D_{T_k-1}(\lambda_j)D^*_{T_{u-1}}(\lambda_j) - D_{T_{k-1}}(\lambda_j)D^*_{T_{u-1}}(\lambda_j) \\ &- D_{T_{k-1}}(\lambda_j)D^*_{T_{u-1}}(\lambda_j) + D_{T_{k-1}}(\lambda_j)D^*_{T_{u-1}}(\lambda_j)]. \end{aligned}$$

Denoting $(T_k - 1)/T = \tilde{\delta}_k$, we have from (6.11) for the term in square brackets in $\tilde{A}$

$$\begin{aligned} \tilde{a}_k =&[D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j) - D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j) - D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda_j) + D_{T_{k-1}}(\lambda_j)D^*_{T_{k-1}}(\lambda)] \\ =&\frac{T^2}{\pi^2 j^2}\Big[\sin^2(\tilde{\delta}_k\pi j) + \sin^2(\delta_{k-1}\pi j) - e^{i\pi j(\tilde{\delta}_k - \delta_{k-1})}\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j) \\ &-e^{i\pi j(\delta_{k-1} - \tilde{\delta}_k)}\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\Big] + O_P(1) \\ =&\frac{T^2}{\pi^2 j^2}[1 - \frac{1}{2}\Big[\cos(2\tilde{\delta}_k\pi j) + \cos(2\delta_{k-1}\pi j)\Big] - 2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))] \\ &+ O_P(1), \end{aligned}$$

from Euler's formula. By the sum-to-product identity for the cosine, it follows

$$\begin{aligned} \tilde{a}_k =&\frac{T^2}{\pi^2 j^2}\Big[1 - \frac{1}{2}\Big[2\cos(\pi j(\tilde{\delta}_k + \delta_{k-1}))\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\Big] \\ &-2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\Big] + O_P(1) \\ =&\frac{T^2}{\pi^2 j^2}\Big[1 - \cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\Big[\cos(\pi j(\tilde{\delta}_k + \delta_{k-1})) + 2\sin(\tilde{\delta}_k\pi j)\sin(\delta_{k-1}\pi j)\Big]\Big] + O_P(1). \end{aligned}$$

Now, by the product-to-sum identity of the sine

$$\begin{aligned} \tilde{a}_k =&\frac{T^2}{\pi^2 j^2}\Big[1 - \cos(\pi j(\tilde{\delta}_k - \delta_{k-1}))\Big[\cos(\pi j(\tilde{\delta}_k + \delta_{k-1})) \\ &+ \cos(\pi j(\tilde{\delta}_k - \delta_{k-1})) - \cos(\pi j(\tilde{\delta}_k + \delta_{k-1}))\Big]\Big] + O_P(1) \\ =&\frac{T^2}{\pi^2 j^2}\Big[1 - \cos^2(\pi j(\tilde{\delta}_k - \delta_{k-1}))\Big] + O_P(1). \end{aligned}$$

Therefore, we have

$$\tilde{A} = \frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2\tilde{a}_k = \frac{T}{2\pi^3 j^2}\sum_{k=0}^{K}\mu_k^2[1 - \cos^2(\pi j(\tilde{\delta}_k - \delta_{k-1}))] + \frac{(K+1)O_P(1)}{2\pi T}. \tag{6.16}$$

For $\alpha < 1$, by a Taylor expansion of the squared cosine at zero

$$\begin{aligned}\tilde{a}_k &= \frac{T^2}{\pi^2 j^2}\left[1 - \left[1 - \pi^2 j^2 (\tilde{\delta}_k - \delta_{k-1})^2 + \pi^4 j^4 O((\tilde{\delta}_k - \delta_{k-1})^4)\right]\right] + O_P(1)\\ &= T^2\left[(\tilde{\delta}_k - \delta_{k-1})^2 - \pi^2 j^2 O((\tilde{\delta}_k - \delta_{k-1})^4)\right] + O_P(1).\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\tilde{A} &= \frac{1}{2\pi T}\sum_{k=0}^{K}\mu_k^2\left\{T^2\left[(\tilde{\delta}_k - \delta_{k-1})^2 - \pi^2 j^2 O((\tilde{\delta}_k - \delta_{k-1})^4)\right] + O_P(1)\right\}\\ &= \frac{T}{2\pi}\sum_{k=0}^{K}\left\{\mu_k^2(\tilde{\delta}_k - \delta_{k-1})^2\right\} - \frac{T\pi j^2}{4}\sum_{k=0}^{K}\mu_k^2 O\left((\tilde{\delta}_k - \delta_{k-1})^4\right) + \frac{O_P(1)}{2\pi T}\sum_{k=0}^{K}\mu_k^2.\end{aligned}$$

By applying the Wald identity for dependent random sums of Brown (1974) and then using Assumption A5, we obtain

$$\begin{aligned}E[\tilde{A}] =& E[K+1]E[\mu_k^2]\left\{\frac{T}{2\pi}E[(\tilde{\delta}_k - \delta_{k-1})^2] - \frac{T\pi j^2}{4}E[O((\tilde{\delta}_k - \delta_{k-1})^4)] + \frac{O(1)}{2\pi T}\right\}\\ =& (\tilde{p}T^{1-\alpha} + 1)\sigma_\Delta^2 T^{-\beta}\left\{\frac{T}{\pi}\frac{\tilde{D}}{\tilde{p}^2}T^{2(\alpha-1)} - \frac{T\pi j^2}{2}O(T^{4(\alpha-1)}) + O(T^{-1})\right\}\\ =& \frac{\sigma_\Delta^2\tilde{D}}{\pi\tilde{p}}T^{\alpha-\beta} - \frac{\sigma_\Delta^2\tilde{p}\pi j^2}{2}O(T^{-2+3\alpha-\beta}) + \sigma_\Delta^2\tilde{p}O(T^{-\alpha-\beta})\\ &+ \frac{\sigma_\Delta^2\tilde{D}}{\pi\tilde{p}}T^{-1+2\alpha-\beta} - \pi j^2 O(T^{-3+4\alpha-\beta}) + O(T^{-1-\beta}).\end{aligned} \tag{6.17}$$

For $\tilde{B}$, we have from (6.11),

$$\begin{aligned}\tilde{B} =& \frac{T}{2\pi^2 j^2}\sum_{k\neq u}\mu_k\mu_u\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_k\pi j)\sin(\delta_u\pi j) - e^{i\pi j(\delta_k-\delta_{u-1})}\sin(\delta_k\pi j)\sin(\delta_{u-1}\pi j)\right.\\ &\left.-e^{i\pi j(\delta_{k-1}-\delta_u)}\sin(\delta_{k-1}\pi j)\sin(\delta_u\pi j) + e^{i\pi j(\delta_{k-1}-\delta_{u-1})}\sin(\delta_{k-1}\pi j)\sin(\delta_{u-1}\pi j)\right]\\ &+ \frac{O_P(1)}{2\pi T}\sum_{k\neq u}\mu_k\mu_u.\end{aligned} \tag{6.18}$$

Denote the term in the square bracket by $\tilde{b}$, and let $\tilde{b}_1$ denote the first two summands and $\tilde{b}_2$ the last two summands so that $\tilde{b} = \tilde{b}_1 + \tilde{b}_2$. We have

$$\begin{aligned}\tilde{b}_1 =& \sin(\delta_k\pi j)\left[e^{i\pi j(\delta_k-\delta_u)}\sin(\delta_u\pi j) - e^{i\pi j(\delta_k-\delta_{u-1})}\sin(\delta_{u-1}\pi j)\right]\\ =& \sin(\delta_k\pi j)\left[\cos(\pi j(\delta_k-\delta_u))\sin(\delta_u\pi j) - \cos(\pi j(\delta_k-\delta_{u-1}))\sin(\delta_{u-1}\pi j)\right.\\ &\left.+i\left\{\sin(\pi j(\delta_k-\delta_u))\sin(\pi j\delta_u) - \sin(\pi j(\delta_k-\delta_{u-1}))\sin(\pi j\delta_{u-1})\right\}\right].\end{aligned}$$

Now, let $\gamma_u = \delta_u - \delta_{u-1}$. Then, by a Taylor approximation at $\gamma_u = 0$

$$\tilde{b}_1 = \pi j\gamma_u\sin(\delta_k\pi j)e^{i\pi j(\delta_k-2\delta_u)} + O_P(\gamma_u^2).$$

Similarly, we have for the third and fourth term in the square bracket

$$\tilde{b}_2 = -\sin(\delta_{k-1}\pi j)\left[e^{i\pi j(\delta_{k-1}-\delta_u)}\sin(\delta_u\pi j) - e^{i\pi j(\delta_{k-1}-\delta_{u-1})}\sin(\delta_{u-1}\pi j)\right],$$

and by a Taylor approximation at $\gamma_u = 0$,

$$\tilde{b}_2 = -\pi j\gamma_u\sin(\delta_{k-1}\pi j)e^{i\pi j(\delta_{k-1}-2\delta_u)} + O_P(\gamma_u^2).$$

Therefore, we have

$$\tilde{b} = -\pi j\gamma_u\left[\sin(\delta_k\pi j)e^{i\pi j(\delta_k-2\delta_u)} - \sin(\delta_{k-1}\pi j)e^{i\pi j(\delta_{k-1}-2\delta_u)}\right].$$

Defining $\gamma_k = \delta_k - \delta_{k-1}$ and approximating at $\gamma_k = 0$, we obtain

$$\tilde{b} = \pi^2 j^2\gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)} + O_P(\gamma_u^2) + O_P(\gamma_k^2)$$

so that

$$\begin{aligned}\tilde{B} =& \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_u\left[\pi^2 j^2\gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)} + O_P(\gamma_u^2) + O_P(\gamma_k^2)\right] + \frac{O_P(1)}{2\pi T}\sum_{k\neq u}\mu_k\mu_u\\ =& \frac{T}{2}\sum_{k\neq u}\mu_k\mu_u\gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)} + \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_u O_P(\gamma_u^2) + \frac{T}{2\pi^2 j^2}\sum_{k\neq l}\mu_k\mu_u O_P(\gamma_k^2)\\ &+ \frac{O_P(1)}{2\pi T}\sum_{k\neq u}\mu_k\mu_u.\end{aligned}$$

Similar to the proof of Lemma 2, we have from the Wald identity of Brown (1974) and Assumption A5

$$\begin{aligned}E[\tilde{B}] =& \tilde{B}_1 + \tilde{B}_2 + \tilde{B}_3\\ =& \frac{T}{2}E\left[\sum_{k\neq l}E\left[\mu_k\mu_u\right]E\left[\gamma_u\gamma_k e^{2i\pi j(\delta_k-\delta_u)}\right]\right] + \frac{T}{2\pi^2 j^2}E\left[\sum_{k\neq l}E\left[\mu_k\mu_u\right]E[O_P(\gamma_k^2)]\right]\\ &+ \frac{O(1)}{2\pi T}E\left[\sum_{k\neq u}E[\mu_k\mu_u]\right].\end{aligned}$$

For the first term,

$$\begin{aligned}\left|E[\tilde{B}_1]\right| \leq& \frac{T}{2}E\left[\sum_{k\neq l}\left|E\left[\mu_k\mu_u\right]\right|\left|E\left[\gamma_u\gamma_k\right]\right|\right] = TE\left[\sum_{k=1}^{K}\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\left|E\left[\gamma_k\gamma_{k-\tau}\right]\right|\right]\\ \leq& TE\left[\sum_{k=1}^{K}\left|E\left[\gamma_k^2\right]\right|\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\right].\end{aligned}$$

Therefore, by Assumptions A4 and A5

$$\begin{aligned}\left|E[\tilde{B}_1]\right| &\leq TE\left[\sum_{k=1}^{K}\left|E\left[\gamma_k^2\right]\right|Var(\mu_k)\tilde{C}\right] = TE[K]E\left[\gamma_k^2\right]Var(\mu_k)\tilde{C}\\ &= \frac{2\tilde{C}\tilde{D}\sigma_\Delta^2}{\tilde{p}}T^{\alpha-\beta} + \tilde{C}\left[O(T^{2\alpha-1-\beta}) + O(T^{-\beta}) + O(T^{\alpha-\beta-1})\right]. \end{aligned} \tag{6.19}$$

Similarly, for the second term

$$\begin{aligned}\left|E[\tilde{B}_2]\right| &\leq \frac{2T}{\pi^2 j^2}E\left[\sum_{k=1}^{K}\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]E\left[O_P(\gamma_k^2)\right]\right|\right]\\ &\leq \frac{2T}{\pi^2 j^2}E\left[\sum_{k=1}^{K}E\left[O_P(\gamma_k^2)\right]\sum_{\tau=1}^{k-1}\left|E\left[\mu_k\mu_{k-\tau}\right]\right|\right] \leq \frac{2\tilde{C}\sigma_\Delta^2}{\pi^2 j^2}T^{1-\beta}E\left[\sum_{k=1}^{K}O_P(\gamma_k^2)\right]\\ &= \frac{2\tilde{C}\sigma_\Delta^2\tilde{p}}{\pi^2 j^2}T^{2-\alpha-\beta}\left[O(T^{2(\alpha-1)}) + O(T^{\alpha-2})\right] = \frac{2\tilde{C}\sigma_\Delta^2\tilde{p}}{\pi^2 j^2}\left[O(T^{\alpha-\beta}) + O(T^{-\beta})\right]. \end{aligned} \tag{6.20}$$

The term $\tilde{B}_3$ is of order $O(T^{-\alpha-\beta})$ by the same arguments as in the proof of Lemma 2. Consequently, parts i.) and ii.) of the Lemma follow directly from Equations (6.16) and (6.18). Similarly, part iii.) is the direct consequence of Equation (6.17) and Equations (6.19) and (6.20). □

Proof of Theorem 1:

From Lemma 1 to 3, we have $I_\mu(\lambda_j) \sim Tj^{-2}\kappa$ under Assumption A2, $I_\mu(\lambda_j) \sim Tj^{-2}\tilde{\kappa}$ under Assumption A1, and $I_\mu(\lambda_j) \sim Tj^{-2}\kappa_{P,T}$ under Assumption A3 with $\beta = 0$, where κ, $\tilde{\kappa}$ and $\kappa_{P,T}$ are two finite constants and a random variable with finite variance, respectively. $E[I_\mu(\lambda_j)] = G_\mu \frac{T}{j^2}$ therefore follows immediately.

To prove that the rank of G_μ is reduced if and only if μ_t has common low-frequency contaminations, we first show that co-shifting according to Definition 1 implies a reduced rank of G_μ and then we show that G_μ has full rank if μ_t is not co-shifting.

For the first part, note that $\Phi'\mu_t = 0 \Leftrightarrow \mu_{at} = \phi_b/\phi_a\mu_{bt}$. Let $\mu_{bt} = \omega_t$, then

$$\mu_t = \begin{pmatrix}\mu_{at}\\ \mu_{bt}\end{pmatrix} = \begin{pmatrix}\phi_b/\phi_a\\ 1\end{pmatrix} \otimes \begin{pmatrix}0 & \phi_b/\phi_a\\ 0 & 1\end{pmatrix}\begin{pmatrix}0\\ \omega_t\end{pmatrix}.$$

Therefore,

$$\begin{aligned} f_\mu(\lambda_j) &= \begin{pmatrix}0 & \phi_b/\phi_a\\ 0 & 1\end{pmatrix}\begin{pmatrix}0 & 0\\ 0 & f_\omega(\lambda_j)\end{pmatrix}\begin{pmatrix}0 & \phi_b/\phi_a\\ 0 & 1\end{pmatrix}'\\ &= f_\omega(\lambda_j)\begin{pmatrix}(\phi_b/\phi_a)^2 & \phi_b/\phi_a\\ \phi_b/\phi_a & 1\end{pmatrix}\end{aligned}$$

so that $\det\left(f_\mu(\lambda_j)\right) = (\phi_b/\phi_a)^2 - (\phi_b/\phi_a)^2 = 0$.

For the second part, let $\Phi'\mu_t = c_t$, where $\exists\ c_t \neq 0$, then $\Phi'\mu_t = c_t \Leftrightarrow \mu_{at} = \phi_a^{-1} c_t - \phi_a/\phi_b \mu_{bt}$ so that for $\mu_{bt} = \omega_t$

$$\mu_t = \begin{pmatrix} \mu_{at} \\ \mu_{bt} \end{pmatrix} = \begin{pmatrix} \phi_a^{-1} & -\phi_a/\phi_b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_t \\ \omega_t \end{pmatrix}.$$

Then, denoting $\tilde{\omega}_t = (c_t,\ \omega_t)'$, we have

$$f_{\mu_t}(\lambda_j) = \begin{pmatrix} \phi_a^{-1} & -\phi_b/\phi_a \\ 0 & 1 \end{pmatrix} f_{\tilde{\omega}}(\lambda_j) \begin{pmatrix} \phi_a^{-1} & 0 \\ -\phi_b/\phi_a & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \phi_a^{-2} f_{c\omega}(\lambda_j) - \phi_b\phi_a^{-2} f_{\omega c}(\lambda_j) - \phi_b\phi_a^{-2} f_{c\omega}(\lambda_j) + \phi_b^2\phi_a^{-2} f_{\omega\omega}(\lambda_j)^2 & \phi_a^{-1} f_{c\omega}(\lambda_j) - \phi_b\phi_a^{-1} f_{\omega\omega}(\lambda_j) \\ \phi_a^{-1} f_{\omega c}(\lambda_j) - \phi_a\phi_b^{-1} f_{\omega\omega}(\lambda_j) & f_{\omega\omega}(\lambda_j) \end{pmatrix}$$

so that $\det\left[f_{\mu_t}(\lambda_j)\right] = \phi_a^{-1}\left(f_{cc}(\lambda_j) f_{\omega\omega}(\lambda_j) - f_{c\omega}(\lambda_j) f_{\omega c}(\lambda_j)\right) \neq 0$. □

To prove Theorem 2, we need the following Lemma.

Lemma 6. For $e > -1$

$$\lim_{m_1\to\infty} \sum_{j=1}^{m_1} j^e = \zeta(-e) + \frac{m_1^{e+1}}{e+1} + O(m_1^e),$$

where ζ is the Riemann zeta function.

Proof of Lemma 6:

From the extension of the Faulhaber formula derived by McGown and Parks (2007),

$$\lim_{m_1\to\infty} \left[(e+1)\sum_{j=1}^{m_1} j^e - m^\gamma F_e(m_1)\right] = (e+1)\zeta(-e)$$
$$\lim_{m_1\to\infty} (e+1)\sum_{j=1}^{m_1} j^e = (e+1)\zeta(-e) + m_1^\gamma F_e(m_1)$$
$$\lim_{m\to\infty} \sum_{j=1}^{m_1} j^e = \zeta(-e) + \frac{m_1^\gamma F_e(m_1)}{(e+1)},$$

where

$$F_e(m_1) = m_1^{\lfloor e\rfloor+2} + \sum_{k=1}^{\lfloor e\rfloor+1} (-1)^k \binom{e+1}{k} B_k m_1^{\lfloor e\rfloor+2-k}$$
$$= m_1^{\lfloor e\rfloor+2} + O\left(m_1^{\lfloor e\rfloor+1}\right),$$

the constants B_k are the Bernoulli numbers, and $\gamma = -(\lfloor e \rfloor) + 1 - e = e - \lfloor e \rfloor - 1$, so that

$$\lim_{m\to\infty} \sum_{j=1}^{m_1} j^e = \zeta(-e) + \frac{m_1^{e-\lfloor e \rfloor -1}\left(m_1^{\lfloor e \rfloor +2} + O\left(m_1^{\lfloor e \rfloor +1}\right)\right)}{e+1}$$
$$= \zeta(-e) + \frac{m_1^{e+1} + O\left(m_1^{e}\right)}{e+1}.$$

□

Proof of Theorem 2:

To prove the theorem, we show that the difference of $\hat{G}_y(d,l,m_1)$ and $\hat{G}_x(d,1,m_1)$ vanishes in probability. Let therefore $\Delta(d) = \|\hat{G}_y(d,l,m_1) - \hat{G}_x(d,1,m_1)\|$, then

$$\Delta(d) = \|(m_1-l+1)^{-1}\sum_{j=l}^{m_1} \Lambda_j(d) I_\mu(\lambda_j)\Lambda_j^*(d) - m_1^{-1}\sum_{j=1}^{l-1}\Lambda_j(d)I_x(\lambda_j)\Lambda_j^*(d) + Z + R\|,$$

where $Z = (m_1-l+1)^{-1}\sum_{j=l}^{m_1}\Lambda_j(d)I_{x\mu}(\lambda_j)\Lambda_j^*(d) + (m_1-l+1)^{-1}\sum_{j=l}^{m_1}\Lambda_j(d)I_{\mu x}(\lambda_j)\Lambda_j^*(d)$ and $R = (l-1)(m_1(m_1-l+1))^{-1}\sum_{j=l}^{m_1}\Lambda_j(d)I_x(\lambda_j)\Lambda_j^*(d)$. Furthermore, define $\nu^- = (d_a - d_a^0) - (d_b - d_b^0)$, $\nu^+ = (d_a - d_a^0) + (d_b - d_b^0)$, and $I_x(\lambda_j) = \lambda_j^{-d_a^0 - d_b^0}\chi_j e^{i\pi(d_a^0 - d_b^0)/2}$, where χ_j is a random matrix with $E[\chi_j] = I$ and $Var[\chi_j] < \infty$. Similarly, $I_\mu(\lambda_j) = \kappa_j \lambda_j^{-2}T^{-1}$, where $E[\kappa_j] < \infty$ and $Var[\kappa_j] < \infty$. Then,

$$\Delta(d) = \|\frac{e^{i\pi(d_a-d_b)/2}}{(m_1-l+1)T}\sum_{j=l}^{m_1}\lambda_j^{d_a+d_b-2}\kappa_j - \frac{e^{i\pi\nu^-/2}}{m_1}\sum_{j=1}^{l-1}\lambda_j^{\nu^+}\chi_j + Z + R\|$$
$$= \|\frac{(2\pi)^{d_a+d_b-2}e^{i\pi(d_a-d_b)/2}}{(m_1-l+1)T^{d_a+d_b-1}}\sum_{j=l}^{m_1} j^{d_a+d_b-2}\kappa_j - \frac{(2\pi)^{\nu^+}e^{i\pi\nu^-/2}}{m_1 T^{\nu^+}}\sum_{j=1}^{l-1} j^{\nu^+}\chi_j + Z + R\|.$$

Due to the independence of x_t and μ_t, $Z \xrightarrow{p} 0$ as $T \to \infty$. Obviously, $R \xrightarrow{p} 0$ for $T \to \infty$ holds as well.

For $l = 1$, the second sum is empty and for the first sum it holds with $A = \|(m_1 - l + 1)^{-1}T^{-d_a-d_b+1}\sum_{j=l}^{m_1} j^{d_a+d_b-2}\kappa_j\|$ and $d_a + d_b > 1$ from Lemma 6,

$$A \leq \max\|\kappa_j\|\frac{O(m_1^{d_a+d_b-2})}{T^{d_a+d_b-1}} = o_p(1).$$

For $d_a + d_b < 1$, we have by definition of the Riemann ζ-function

$$A \leq \frac{\max\|\kappa_j\|\zeta(-d_a-d_b+2)}{m_1 T^{d_a+d_b-1}},$$

which is $o_P(1)$, for $\delta_{m_1} > 1 - d_a - d_b$.

For the second part of the theorem, we have $\sum_{j=l}^{m_1} j^{d_a+d_b-2} \le m_1 l^{d_a+d_b-2}$. Therefore, with $d_a + d_b < 1$,

$$A \le \frac{m_1 l^{d_a+d_b-2}}{(m_1 - l + 1)T^{d_a+d_b-1}} = O(l^{d_a+d_b-2} T^{-(d_a+d_b-1)}) = o_P(1),$$

for $\delta_{m_1} > \delta_l$ and $l = T^{(d_a+d_b-1)/(d_a+d_b-2)+\upsilon}$. Furthermore, let $B = \|m_1^{-1} T^{-\nu^+} \sum_{j=1}^{l-1} j^{\nu^+} \chi_j\|$, then

$$B \le \frac{\max_{j<l} \|\chi_j\|}{m_1 T^{\nu^+}} \sum_{j=1}^{l-1} j^{\nu^+} = \frac{\max_{j<l} \|\chi_j\|}{m_1 T^{\nu^+}} O(l^{\nu^+ + 1}) = o_P(1),$$

for $l = o(T^{(\nu^+ + \delta_{m_1})/(\nu^+ + 1)})$. □

Proof of Theorem 3:

The proof directly follows from a Taylor expansion of the matrix $\hat{G}_y(\hat{d}(m), l, m_1)$ at d^0 and is omitted here.

Proof of Theorem 4:

The proof follows ideas in Robinson (2008). For any $c > 0$, define neighborhoods $N_\beta(c) = \{\beta : |\beta - \beta^0| < c]$ and $N_d(c) = \{d : \|d - d^0\| < c]$. Furthermore, fix $\varepsilon > 0$ and define $N(\varepsilon) = N_\beta(\varepsilon^{-1}(\frac{m}{T})^{\nu^0}) N_d(\varepsilon)$, $\overline{N(\varepsilon)} = \Theta \backslash N(\varepsilon)$, and $\zeta_i = d_i - d_i^0$.

We split the parameter space into two. For a constant $0 < C \le \frac{1}{8}$, define $\Theta_{da} = \{d \in \Theta_d : \zeta_a \ge -\frac{1}{2} + C; \zeta_b \ge -\frac{1}{2} + C]$ and $\Theta_{db} = \Theta_d \backslash \Theta_{da}$. Since $P(\hat{\theta} \in N(\varepsilon)) \le P(\inf_{\bar{N(\varepsilon)}} \{R(\theta) - R(\theta^0)] \le 0)$, the consistency of $\hat{\Theta}$ follows if we show

$$P\left(\inf_{\overline{N(\varepsilon)} \cap \{\Theta_\beta \times \Theta_{da}\}} \{R(\theta) - R(\theta^0)\} \le 0 \right) \to 0 \quad \text{as} \quad T \to \infty \quad \text{and} \tag{6.21}$$

$$P\left(\inf_{\overline{N(\varepsilon)} \cap \{\Theta_\beta \times \Theta_{db}\}} \{R(\theta) - R(\theta^0)\} \le 0 \right) \to 0 \quad \text{as} \quad T \to \infty. \tag{6.22}$$

First we show (6.21). Rewrite $R(\theta) - R(\theta^0)$ as

$$R(\theta) - R(\theta^0) = \log\det\left[\hat{\Omega}^{tri}(\theta) \hat{\Omega}^{tri}(\theta^0)^{-1}\right] - 2(\zeta_a + \zeta_b) \frac{1}{m - l + 1} \sum_{j=l}^{m} \log \lambda_j,$$

where $\hat{\Omega}^{tri}(\theta) = \frac{1}{m-l+1} \sum_{j=l}^{m} \mathrm{Re}[\Lambda_j(d) B I_{yy}^{tri}(\lambda_j) B' \Lambda_j(d)^*]$ instead of $\hat{\Omega}(\theta)$ in Robinson (2008). Define a vector type II $I(d_a^0, d_b^0)$ process as

$$\xi_t = \binom{\xi_{at}}{\xi_{bt}} = B^0, z_t = \begin{pmatrix} (1-L)^{-d_a^0} & y_{at} & 1(t \ge 1) \\ (1-L)^{-d_b^0} & y_{bt} & 1(t \ge 1) \end{pmatrix}, B^0 = \begin{pmatrix} 1 & -\beta^0 \\ 0 & 1 \end{pmatrix}.$$

Define further analogously to Robinson (2008)

$$
\begin{aligned}
H_j &= (h_{k_1k_2j}) = \Lambda_j(d^0) I^{tri}_{\xi j}(\lambda_j)\Lambda_j(d^0)^* \quad \text{and}\\
\hat{G}^{(1)}(d) &= (\hat{g}^{(1)}_{k_1k_2}), \quad \text{where } \hat{g}^{(1)}_{k_1k_1} = \frac{1}{m-l+1}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2\zeta_{k_1}} h_{k_1k_1j} \quad \text{and}\\
\hat{g}^{(1)}_{ab} &= \hat{g}^{(1)}_{ba} = \frac{1}{m-l+1}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{\zeta_a+\zeta_b}\left(e^{i(\pi-\lambda_j)(\zeta_b-\zeta_a)/2}h_{abj} + e^{-i(\pi-\lambda_j)(\zeta_b-\zeta_a)/2}h_{baj}\right).
\end{aligned}
$$

Similar to Robinson (2008), we obtain $R(\theta) - R(\theta^0) = U_d(d) + U_\beta(\theta)$ with

$$
\begin{aligned}
U(d) &= \log\det\left[\Upsilon(d)\hat{G}^{(1)}(d)\Upsilon(d)\hat{G}^{(1)}(d^0)^{-1}\right] + \Phi_a(d,l) + u(d) + \Phi_b(d,l) \quad \text{and}\\
U_\beta(\theta) &= \log\det\left[\hat{\Omega}^{tri*}(\theta)\hat{G}^{(1)}(d)^{-1}\right] - \Phi_a(d,l) + \Phi_b(d,l).
\end{aligned}
$$

Here,

$$
\begin{aligned}
&\Upsilon(d) = \operatorname{diag}\left((2\zeta_a+1)^{1/2}, (2\zeta_b+1)^{1/2}\right),\\
&\hat{\Omega}^{tri*}(\theta) = \Xi(\theta)\hat{\Omega}^{tri}(\theta)\Xi(\theta),\\
&\Xi(\theta) = \operatorname{diag}\left(\lambda_m^{-\zeta_a}, \lambda_m^{-\zeta_b}\right),\\
&\Phi_1(d,l) = \log\left[(l-1)^2(l^{2\zeta_a+1})^{-1}(l^{2\zeta_b+1})^{-1}\right],\\
&\Phi_2(d,l) = 2(\zeta_a+\zeta_b)(l-1)^{-1} l\log l, \quad \text{and}\\
&u(d) = \sum_{i=a,b}\left[2\zeta_i - \log(2\zeta_i+1) + 2\zeta_i\left(\log m - \frac{1}{m-l+1}\sum_{j=l}^{m}\log j - 1\right)\right].
\end{aligned}
$$

The functions $\Phi_a(d,l)$ and $\Phi_b(d,l)$ control effects of taking summations from l by application of the Euler-McLaurin formula as in Lemma 2(a) of Shimotsu (2010). In contrast to Robinson (2008), all matrices here are defined by the trimmed periodogram and we do not have the parameter γ.

Now, (6.21) follows if we show that, as $T \to \infty$,

$$
P\left(\inf_{\overline{N(\epsilon)}\cap\Theta_{d_a}} U_d(d) \leq 0\right) \to 0 \quad \text{and} \tag{6.23}
$$

$$
P\left(\inf_{\overline{N_\beta(\frac{1}{\epsilon}(\frac{T}{m})^{\nu^0})}\times\Theta_d} U_\beta(d) \leq 0\right) \to 0. \tag{6.24}
$$

The proof of (6.23) is similar to Robinson (2008). Define the population analogue of $\hat{g}^{(1)}_{k_1k_2}$ as

$$
g^{(1)}_{k_1k_2} = \omega_{k_1k_1}\frac{1}{l}\int_l^1 x^{2\zeta_{k_1}}dx
$$

and

$$
g^{(1)}_{ab} = g^{(1)}_{ba} = \omega_{ab}\frac{1}{l}\int_l^1 x^{\zeta_a+\zeta_b}dx\cos\tau,
$$

where

$$\tau = (\zeta_b - \zeta_a)\frac{\pi}{2}.$$

Then, (6.23) holds if

$$\sup_{\Theta_{da}} \left|\left|\Upsilon(d)[\hat{G}^{(1)}(d) - G^{(1)}(d)]\Upsilon(d)\right|\right| \xrightarrow{P} 0, \tag{6.25}$$

$$\sup_{\Theta_{da}} \left|\left| \left[\Upsilon(d)G^{(1)}(d)\Upsilon(d)\right]^{-1} \right|\right| < \infty, \tag{6.26}$$

$$\underline{\inf}_{N_d(\varepsilon)\cap\Theta_{d_a}} \left[\log\det\left[\Upsilon(d)G^{(1)}(d)\Upsilon(d)G^{(1)}(d^0)^{-1}\right] + \Phi_a(\theta,l)\right] \geq 0, \quad \text{and} \tag{6.27}$$

$$\lim_{T\to\infty} \underline{\inf}_{N_d(\varepsilon)\cap\Theta_{d_a}} [u(d) - \Phi_b(d,l)] > 0. \tag{6.28}$$

These conditions correspond to (7.5) - (7.8) in Robinson (2008).

The proof of (6.25) follows from observing that

$$\begin{aligned}&\left|\left|\frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}\left[\Lambda_j(d)BI_y^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right|\right| \\ =&\left|\left|\frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}\left[\Lambda_j(d)B(I_x^{tri}(\lambda_j) + I_{x\mu}^{tri}(\lambda_j) + I_{\mu x}^{tri}(\lambda_j) + I_{\mu}^{tri}(\lambda_j))B'\Lambda_j(d)^*\right]\right|\right|.\end{aligned}$$

Now,

$$\frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}\left[\Lambda_j(d)BI_\mu^{tri}(\lambda_j)B'\Lambda_j(d)^*\right] = O_P\left(\frac{1}{m}\left(\frac{\delta_M}{T}\right)^{2\left(\frac{d_a+d_b}{2}-\frac{1}{2}\right)}\right) = o_P(1).$$

In addition to this, we have because $||I_{\mu x}(\lambda_j)||^2 = I_\mu(\lambda_j)I_x(\lambda_j)$

$$\begin{aligned}
&\Big|\Big|\frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}\left[\Lambda_j(d)BI_{\mu x}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\Big|\Big| \\
\leq&\frac{1}{m-l+1}\sum_{j=l}^{m}\Big|\Big|\mathrm{Re}\left[\Lambda_j(d)BI_{\mu x}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\Big|\Big| \\
\leq&\frac{1}{m-l+1}\sum_{j=l}^{m}\left(\mathrm{Re}\left[\Lambda_j(d)BI_{x}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}} \\
&\cdot\left(\mathrm{Re}\left[\Lambda_j(d)BI_{\mu}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}} \\
\leq&\underbrace{\left(\frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}\left[\Lambda_j(d)BI_{x}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}}}_{O_P(1)} \\
&\cdot\underbrace{\left(\frac{1}{m-l+1}\sum_{j=l}^{m}\mathrm{Re}\left[\Lambda_j(d)BI_{\mu}^{tri}(\lambda_j)B'\Lambda_j(d)^*\right]\right)^{\frac{1}{2}}}_{O_P(1)}.
\end{aligned}$$

Applying now the same arguments as in the proof of (17) in Shimotsu (2012) gives (6.25). Also the proof of (6.26), (6.27), and (6.28) is equal to Shimotsu (2012) proving his Equations (18), (19), and (20).

We proceed to show (6.24). Define $\hat{g}^{(i)}_{k_1k_2}$ similarly to Robinson (2008) but using $\frac{1}{m-l+1}\sum_{j=l}^{m}$ and setting $\tau = (\zeta_b-\zeta_a)\frac{\pi}{2}$ and $\gamma^0 = (d_b^0-d_a^0)\frac{\pi}{2}$. Let further

$$\begin{aligned}
\hat{\alpha}_a &= \left(\hat{g}^{(2)}_{aa}\hat{g}^{(1)}_{bb} - 2\hat{g}^{(1)}_{ab}\hat{g}^{(2)}_{ab}/\det(\hat{G}^{(1)}(d)\right) \quad \text{and} \\
\hat{\alpha}_b &= \left(\hat{g}^{(3)}_{aa}\hat{g}^{(1)}_{bb} - (\hat{g}^{(2)}_{ab})^2)/\det(\hat{G}^{(1)}(d)\right).
\end{aligned}$$

Define $g^{(i)}_{k_1k_2}$, the population counterpart of $\hat{g}^{(i)}_{k_1k_2}$, analogously to $g^{(1)}_{k_1k_2}$: for example,

$$\begin{aligned}
g^{(2)}_{ab} &= g^{(2)}_{ba} = \frac{1}{l}\omega_{bb}\cos\gamma\int_l^1 x^{d_a-d_b^0+\zeta_b}dx \quad \text{and} \\
g^{(3)}_{aa} &= \frac{1}{l}\omega_{bb}\int_l^1 x^{2(d_a-d_b^0)}dx,
\end{aligned}$$

where $\gamma = (d_b-d_a)\frac{\pi}{2}$. Using summation by parts and Lemma 1(b) of Shimotsu (2012), we obtain

$$\sup_{\Theta_d}|\hat{g}^{(i)}_{k_1k_2} - g^{(i)}_{k_1k_2}| \xrightarrow{P} 0,$$

for $i = 1,2,3$, $k_1, k_2 = a, b$, and as $T\to\infty$. Rewrite $U_\beta(d) = \log Q(b_n(\beta)) - \Phi_a(d,l) + \Phi_b(d,l)$, where $Q(s) = 1 + \hat{a}_1 s + \hat{a}_2 s^2$ and $b_n(\beta) = \lambda_j^{-\nu^0}(\beta^0-\beta)$. Define

$$\begin{aligned}
\alpha_a &= \left(g^{(2)}_{aa}g^{(1)}_{bb} - 2g^{(1)}_{ab}g^{(2)}_{ab}/\det(G^{(1)}(d)\right) \quad \text{and} \\
\alpha_b &= \left(g^{(3)}_{aa}g^{(1)}_{bb} - (g^{(2)}_{ab})^2)/\det(G^{(1)}(d)\right).
\end{aligned}$$

Following Robinson (2008), with $\rho = \sup_{\Theta_d} |\Phi_a(d,l) - \Phi_b(d,l)| < \infty$, the probability in (6.24) is bounded by

$$P\left(\log\left[1 - \left(\sup_{\Theta_d} \frac{|\hat{\alpha}_a|}{\varepsilon} + \inf_{\Theta_d} \frac{|\hat{\alpha}_b|}{\varepsilon^2}\right)\right] \leq \rho\right) + P\left(\sup_{\Theta_d} \frac{|\hat{\alpha}_a|}{2|\hat{\alpha}_b|} > \frac{1}{\varepsilon}\right)$$
$$\leq 2P\left(\sup_{\Theta_d} |\hat{\alpha}_a - a_a| + \frac{2}{\varepsilon} \sup_{\Theta_d} |\hat{\alpha}_b - \alpha_b| + \varepsilon\rho \geq \frac{1}{\varepsilon} \inf_{\Theta_d} \alpha_b - \sup_{\Theta_d} |\alpha_a|\right),$$

which has an additional term $\varepsilon\rho$ compared to (7.13) in Robinson (2008). (6.24) follows now exactly as in Shimotsu (2012).

It remains to show (6.22). Write

$$R(\theta) - R(\theta^0) = U_d^*(d) + U_\beta^*(\theta),$$

where

$$\begin{aligned} U_d^*(d) =& \log\ \det\left[\Xi(d)\hat{G}^{(1)}(d)\Xi(d)\hat{G}^{(1)}(d^0)^{-1}\right] \\ & - 2(\zeta_a + \zeta_b)\frac{1}{m-l+1}\sum_{j=l}^{m} \log\ \lambda_j, \end{aligned}$$

and

$$U_\beta^*(\theta) = \log\ \det[\hat{\Omega}^{tri*}(\theta)\hat{G}^{(1)}(d)^{-1}] = U_\beta(\theta) + \Phi_a(d,l) - \Phi_b(d,l).$$

Then,

$$P\left(\inf_{N_\beta(\varepsilon^{-1})(\frac{T}{m})^{\nu^0}) \times \Theta_d} U_\beta^*(\theta) \leq 0\right) \to 0$$

follows from the proof of (6.24), so it suffices to show

$$P\left(\inf_{\Theta_{db}} U_d^*(d) \leq 0\right) \to 0.$$

Rewrite $U_d^*(d)$ as

$$U_d^*(d) = \log\det\hat{D}(d) - \log\det\hat{D}(d^0),$$

where

$$\hat{D}(d) = \frac{1}{m-l+1}\sum_{j=l}^{m}\begin{pmatrix} \left(\frac{j}{q}\right)^{2\zeta_a} h_{aaj} & \left(\frac{j}{q}\right)^{\zeta_a+\zeta_b} \mathrm{Re}\left(e^{i(\pi-\lambda_j)\frac{(\zeta_b-\zeta_a)}{2}} h_{abj}\right) \\ \left(\frac{j}{q}\right)^{\zeta_a+\zeta_b} \mathrm{Re}\left(e^{i(\pi-\lambda_j)\frac{(\zeta_b-\zeta_a)}{2}} h_{abj}\right) & \left(\frac{j}{q}\right)^{2\zeta_b} h_{bbj} \end{pmatrix},$$

and

$$q = \exp\left(\frac{1}{m-l+1}\sum_{j=l}^{m}\log j\right).$$

Define K(d) as $\hat{D}(d)$ but $h_{k_1k_2j}$ is replaced with $\omega_{k_1k_2}$.
Now,

$$\sup_{\Theta_{db}} |D(d) - K(d)| \xrightarrow{P} 0$$

follows from Lemma 1 of Shimotsu (2010) and the proof of Theorem 1 of Shimotsu (2007). Furthermore, it follows from Shimotsu (2007) and Shimotsu (2012) that there exists an $\varepsilon \in (0, 0.1)$ and $l < m$ such that

$$\inf_{\Theta_{db}} \det K(d) \geq (1+\varepsilon) \det G^0 + o(1).$$

Therefore,

$$\det \hat{D}(d) \geq (1+\varepsilon) \det G^0 + o(1).$$

Since

$$\det \hat{D}(d^0) = \det \hat{G}^{(1)}(d^0) = \det \Omega^0 + o_P(1)$$

from (6.25), we establish (6.22). □

Proof of Theorem 5:

$\hat{\theta}$ has now the stated limiting distribution if for any $\tilde{\theta}$ such that $\tilde{\theta} - \theta^0 = O_P(m^{-1/2})$,

$$\sqrt{m}(\Delta_T)^{-1}\frac{dR(\theta^0)}{d\theta}|_{\theta^0} \overset{d}{\Rightarrow} N(0, \Xi)$$

and

$$(\Delta_T)^{-1}\frac{d^2R(\tilde{\theta})}{d\theta d\theta'}(\Delta_T)^{-1} \xrightarrow{P} \Xi.$$

For the score vector approximation, denote by $s_k(\theta)$ the k-th element of $\frac{dR(\theta)}{d\theta}$ and by E_{ij} the matrix of zeros where the ij-th element has been replaced by a one. Following exactly the lines of Theorem 4 of Robinson (2008), we have that

$$\begin{aligned}
s_1(\theta^0) &= -tr\frac{1}{m-l+1}\sum_{j=l}^{m}\lambda_j^{d_b^0-d_a^0}(E_{12}\,\mathrm{Re}[\Lambda_j(d^0)B^0 I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]e^{i\pi(d_b-d_a)/2} \\
&\quad +\mathrm{Re}[\Lambda_j(d^0)B^0 I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{21}e^{-i\pi(d_b-d_a)/2})\hat{G}(d^0)^{-1} \\
&= -tr\frac{1}{m-l+1}\sum_{j=l}^{m}\lambda_j^{d_b^0-d_a^0}(E_{12}\,\mathrm{Re}[\Lambda_j(d^0)B^0 I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]e^{i\pi(d_b-d_a)/2} \\
&\quad +\mathrm{Re}[\Lambda_j(d^0)B^0 I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{21}e^{-i\pi(d_b-d_a)/2})\hat{G}(d^0)^{-1}+o_P(1),
\end{aligned}$$

$$\begin{aligned}
s_2(\theta^0) &= itr\Bigg[\frac{1}{m-l+1}\sum_{j=l}^{m}(\mathrm{Re}[\Lambda_j(d^0)B^0 I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{22} \\
&\quad -E_{22}\,\mathrm{Re}[\Lambda_j(d^0)B^0 I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*])\hat{G}(d^0)^{-1}\Bigg] \\
&= itr\Bigg[\frac{1}{m-l+1}\sum_{j=l}^{m}(\mathrm{Re}[\Lambda_j(d^0)B^0 I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{22} \\
&\quad -E_{22}\,\mathrm{Re}[\Lambda_j(d^0)B^0 I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*])\hat{G}(d^0)^{-1}\Bigg]+o_P(1), \quad \text{and}
\end{aligned}$$

$$\begin{aligned}
s_{2+k}(\theta^0) &= tr\frac{1}{m-l+1}\sum_{j=l}^{m}(\log\lambda_j-\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j) \\
&\quad (E_{kk}\,\mathrm{Re}[\Lambda_j(d^0)B^0 I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]+\mathrm{Re}[\Lambda_j(d^0)B^0 I_y^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{kk})\hat{G}(d^0)^{-1} \\
&= tr\frac{1}{m-l+1}\sum_{j=l}^{m}(\log\lambda_j-\frac{1}{m-l+1}\sum_{j=l}^{m}\log\lambda_j) \\
&\quad (E_{kk}\,\mathrm{Re}[\Lambda_j(d^0)B^0 I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]+\mathrm{Re}[\Lambda_j(d^0)B^0 I_x^{tri}(\lambda_j)B^{0'}\Lambda_j(d^0)^*]E_{kk})\hat{G}(d^0)^{-1},
\end{aligned}$$

for $k=1,2$ and by the same arguments as in the consistency proof. The score vector approximation follows now directly as in Robinson (2008).

The Hessian approximation is also similar to the arguments in Robinson (2008) and therefore omitted here. □

Supplementary Appendix

This Supplementary Appendix contains additional simulation results. The first part shows the results when investigating cross-sectionally correlated errors with $r = 0.5$, the second part shows the results when the DGP is a combination of long-memory process and stationary random level-shift process, the third part gives the results when the DGP is a combination of long memory process and deterministic trend process, and finally the fourth part shows results when combining our two suggested procedures, i.e., we estimate the order of integration based on the estimated fractional cointegration rank instead of assuming it to be known as in Section 6.5.2.

Cross-sectionally Correlated Errors

		Cointegration	TRUE				FALSE			
d	T	Breaks	TRE	RY02	CH06	SRF	TRE	RY02	CH06	SRF
0.2	250	NO	1.00	1.00	0.18	0.10	0.35	0.07	0.08	0.01
		DIS	1.00	0.59	0.15	0.03	0.26	0.03	0.39	0.11
		COB	1.00	1.00	0.60	0.55	0.47	0.53	0.68	0.45
	1000	NO	1.00	1.00	0.46	0.25	0.01	0.00	0.07	0.01
		DIS	1.00	0.50	0.34	0.21	0.01	0.00	0.55	0.54
		COB	1.00	1.00	0.92	0.88	0.03	0.09	0.83	0.77
0.4	250	NO	1.00	1.00	0.62	0.34	0.35	0.08	0.10	0.01
		DIS	1.00	0.83	0.10	0.03	0.26	0.03	0.26	0.03
		COB	1.00	1.00	0.82	0.60	0.46	0.34	0.38	0.15
	1000	NO	1.00	1.00	0.99	0.83	0.01	0.00	0.07	0.01
		DIS	1.00	0.88	0.16	0.16	0.01	0.00	0.30	0.10
		COB	1.00	1.00	1.00	0.95	0.02	0.02	0.43	0.30

Table 6.5: Simulation results fractional cointegration rank — cross-sectionally correlated errors.
In analogy to Table 6.1, the table reports the mean estimated fractional cointegration rank for a bivariate fractionally integrated system. The DGP is based on Equations (6.5) to (6.9) with $r = 0.5$. In case of fractional cointegration, $\beta = (-1, 1)'$ and $b = d$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. Our estimator (TRE) is considered with $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}$, $N = m_1^{-0.2}$, and the univariate estimator by Iacone (2010) to estimate d. For the procedures by Robinson and Yajima (2002) (RY02), Chen and Hurvich (2006) (CH06), and Souza et al. (2018) (SRF), parameter values are chosen according to the authors recommendation.

		Estimator	TMLW				GSE			
			Bias		RMSE		Bias		RMSE	
d	T	Breaks	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
0.2	250	NO	-0.02	-0.02	0.15	0.15	-0.01	-0.02	0.06	0.07
		DIS	0.03	0.03	0.16	0.16	0.20	0.20	0.23	0.23
		COB	0.00	0.00	0.15	0.15	0.13	0.13	0.17	0.17
	1000	NO	-0.01	-0.01	0.07	0.07	-0.01	-0.01	0.04	0.04
		DIS	0.02	0.02	0.08	0.08	0.19	0.19	0.21	0.21
		COB	0.00	0.00	0.07	0.07	0.13	0.13	0.15	0.15
0.4	250	NO	-0.04	-0.04	0.16	0.16	-0.02	-0.02	0.06	0.07
		DIS	0.00	0.00	0.15	0.15	0.09	0.09	0.13	0.13
		COB	-0.02	-0.02	0.15	0.15	0.05	0.05	0.09	0.09
	1000	NO	-0.02	-0.02	0.07	0.07	-0.01	-0.01	0.04	0.04
		DIS	0.00	0.00	0.07	0.07	0.07	0.07	0.09	0.09
		COB	-0.01	-0.01	0.07	0.07	0.04	0.04	0.07	0.07

Table 6.6: Simulation results order of integration — no fractional cointegration but cross-sectionally correlated errors.
In analogy to Table 6.2, the table reports bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally integrated system. The DGP is based on Equations (6.5) to (6.9) with $r = 0.5$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

		Estimator	TMLW						GSE					
			Bias			RMSE			Bias			RMSE		
d	T	Breaks	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$
0.2	250	NO	-0.03	-0.02	-2.58	0.18	0.18	5.28	-0.02	-0.02	-0.48	0.07	0.07	2.05
		DIS	0.12	0.04	-3.34	0.23	0.19	5.95	0.43	0.18	-2.62	0.44	0.22	5.92
		COB	-0.02	0.02	-2.29	0.18	0.18	4.98	0.03	0.18	-0.10	0.08	0.22	0.82
	1000	NO	-0.02	-0.01	-0.90	0.09	0.09	2.89	-0.01	-0.01	-0.05	0.04	0.04	0.38
		DIS	0.10	0.02	-3.09	0.15	0.09	5.51	0.42	0.17	-4.11	0.43	0.19	6.93
		COB	-0.01	0.01	-0.74	0.09	0.09	2.63	0.04	0.16	-0.03	0.06	0.19	0.10
0.4	250	NO	-0.02	-0.03	-1.08	0.18	0.18	3.19	-0.02	-0.02	-0.03	0.08	0.07	0.17
		DIS	0.11	0.00	-2.15	0.23	0.18	4.55	0.42	0.09	-3.09	0.44	0.13	5.73
		COB	-0.01	0.00	-0.93	0.19	0.18	2.97	0.01	0.08	-0.02	0.07	0.12	0.11
	1000	NO	-0.01	-0.02	-0.07	0.09	0.09	0.38	-0.01	-0.01	-0.01	0.04	0.04	0.05
		DIS	0.10	0.00	-0.53	0.15	0.09	1.64	0.42	0.07	-4.53	0.43	0.09	6.54
		COB	-0.01	-0.01	-0.06	0.09	0.09	0.43	0.02	0.06	-0.01	0.04	0.08	0.04

Table 6.7: Simulation results order of integration — fractional cointegration and cross-sectionally correlated errors.
In analogy to Table 6.3, the table reports bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally cointegrated system with cointegration vector $\beta = (1, -1)'$. The DGP is based on Equations (6.5) to (6.9) with $r = 0.5$. Break sizes are random with mean zero and variance one and they occur with probability $5/T$. Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

Stationary Random Level-Shift Process

Now, we present the results when the DGP is a combination of long-memory process and stationary random level-shift process. In analogy to the DGP in Section 6.5, we consider the following process

$$y_{at} = \zeta_a \mu_t + \xi \tilde{\mu}_t + x_t + (1-L)^{-(d-\tilde{d})} u_t \tag{6.29}$$

$$y_{bt} = \zeta_b \mu_t + x_t, \tag{6.30}$$

where

$$\mu_t = (1-\pi_t)\mu_{t-1} + \pi_t \eta_t, \qquad \pi_t \sim B(5/T), \qquad \eta_t \sim N(0,1), \tag{6.31}$$

$$\tilde{\mu}_t = (1-\tilde{\pi}_t)\tilde{\mu}_{t-1} + \tilde{\pi}_t \tilde{\eta}_t, \qquad \tilde{\pi}_t \sim B(5/T), \qquad \tilde{\eta}_t \sim N(0,1), \tag{6.32}$$

$$x_t = (1-L)^{-d} e_t, \qquad \text{and} \qquad \begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right). \tag{6.33}$$

Again, we present results for $d = 0.2, 0.4$, $r = 0, 0.5$, $T = 250, 1000$, $\tilde{d} = 0, d$, the case of joint breaks, and the case of distinct breaks. We omit the case of no breaks as the results are identical to those presented in Section 6.5.

The results concerning the investigation of the fractional cointegration relation can be found in Table 6.8 and the results for the estimation of the order of integration in Tables 6.9 and 6.10. It can be seen that the results are qualitatively similar to those for the nonstationary random level-shift process. The trimmed rank estimator and the trimmed multivariate local Whittle estimator are robust to low-frequency contaminations and provide accurate inference also when the series is contaminated by stationary random level shifts. The non-robust approaches, on the other hand, are seriously distorted and should not be considered for series that potentially exhibit stationary random level shifts.

			Cointegration	TRUE				FALSE			
r	d	T	Breaks	TRE	RY02	CH06	SRF	TRE	RY02	CH06	SRF
0	0.2	250	DIS	0.96	0.26	0.12	0.04	0.00	0.00	0.38	0.06
			COB	1.00	1.00	0.83	0.64	0.00	0.01	0.78	0.30
		1000	DIS	1.00	0.12	0.28	0.30	0.00	0.00	0.48	0.46
			COB	1.00	1.00	0.97	0.93	0.00	0.00	0.89	0.68
	0.4	250	DIS	0.98	0.66	0.08	0.04	0.00	0.00	0.30	0.01
			COB	1.00	1.00	0.94	0.70	0.00	0.00	0.49	0.04
		1000	DIS	1.00	0.80	0.10	0.24	0.00	0.00	0.29	0.05
			COB	1.00	1.00	1.00	0.98	0.00	0.00	0.54	0.13
0.5	0.2	250	DIS	1.00	0.56	0.05	0.01	0.22	0.01	0.27	0.04
			COB	1.00	1.00	0.60	0.49	0.52	0.57	0.68	0.41
		1000	DIS	1.00	0.44	0.22	0.06	0.00	0.00	0.42	0.35
			COB	1.00	1.00	0.93	0.88	0.03	0.08	0.84	0.75
	0.4	250	DIS	1.00	0.85	0.03	0.02	0.25	0.02	0.17	0.01
			COB	1.00	1.00	0.84	0.58	0.48	0.36	0.36	0.10
		1000	DIS	1.00	0.93	0.07	0.07	0.00	0.00	0.18	0.04
			COB	1.00	1.00	1.00	0.95	0.02	0.01	0.41	0.23

Table 6.8: Simulation results fractional cointegration rank — stationary random level-shift process shifts.

In analogy to Table 6.1, the table reports the mean estimated fractional cointegration rank for a bivariate fractionally integrated system. The DGP is based on Equations (6.29) to (6.33). In case of fractional cointegration, $\beta = (-1, 1)'$ and $\tilde{d} = d$. Our estimator (TRE) is considered with $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}$, $N = m_1^{-0.2}$, and the univariate estimator by Iacone (2010) to estimate d. For the procedures by Robinson and Yajima (2002) (RY02), Chen and Hurvich (2006) (CH06), and Souza et al. (2018) (SRF), parameter values are chosen according to the authors recommendation.

			Estimator	TMLW				GSE			
				Bias		RMSE		Bias		RMSE	
r	d	T	Breaks	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
0	0.2	250	DIS	0.02	0.02	0.20	0.20	0.18	0.18	0.21	0.21
			COB	0.01	0.02	0.19	0.20	0.16	0.16	0.19	0.19
		1000	DIS	0.01	0.01	0.09	0.09	0.16	0.16	0.18	0.18
			COB	0.01	0.01	0.09	0.09	0.15	0.15	0.17	0.17
	0.4	250	DIS	-0.01	-0.01	0.20	0.20	0.08	0.08	0.12	0.12
			COB	-0.01	-0.01	0.19	0.20	0.07	0.07	0.11	0.11
		1000	DIS	-0.01	-0.01	0.09	0.09	0.06	0.05	0.08	0.08
			COB	0.00	-0.01	0.09	0.09	0.05	0.05	0.08	0.08
0.5	0.2	250	DIS	0.04	0.04	0.16	0.16	0.20	0.20	0.23	0.23
			COB	0.00	0.00	0.15	0.15	0.13	0.13	0.16	0.16
		1000	DIS	0.02	0.02	0.08	0.08	0.18	0.18	0.20	0.20
			COB	0.00	0.00	0.07	0.07	0.12	0.12	0.14	0.14
	0.4	250	DIS	0.00	0.00	0.15	0.15	0.09	0.09	0.12	0.12
			COB	-0.02	-0.02	0.15	0.15	0.05	0.05	0.09	0.09
		1000	DIS	0.00	0.00	0.07	0.07	0.07	0.07	0.08	0.09
			COB	-0.01	-0.01	0.07	0.07	0.04	0.04	0.06	0.06

Table 6.9: Simulation results order of integration — stationary random level-shift process and no fractional cointegration.
In analogy to Table 6.2, the table reports the bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally integrated system. The DGP is based on Equations (6.29) to (6.33). Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

			Estimator	TMLW						GSE					
				Bias			RMSE			Bias			RMSE		
r	d	T	Breaks	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$
0	0.2	250	DIS	0.10	0.02	0.20	0.23	0.19	2.84	0.38	0.17	1.77	0.39	0.20	4.24
			COB	-0.02	0.02	-0.19	0.19	0.20	2.60	-0.02	0.18	0.00	0.08	0.21	0.37
		1000	DIS	0.07	0.01	0.08	0.12	0.09	1.94	0.37	0.15	1.93	0.38	0.17	4.25
			COB	-0.01	0.01	-0.02	0.09	0.09	1.13	-0.01	0.16	0.00	0.04	0.18	0.06
	0.4	250	DIS	0.10	-0.01	0.12	0.23	0.19	2.28	0.39	0.07	0.61	0.40	0.11	3.23
			COB	-0.02	0.00	-0.09	0.20	0.19	1.78	-0.02	0.07	0.00	0.08	0.11	0.18
		1000	DIS	0.07	-0.01	-0.04	0.13	0.09	0.57	0.39	0.05	0.49	0.40	0.07	2.97
			COB	-0.01	0.00	0.00	0.09	0.09	0.15	-0.01	0.06	0.00	0.04	0.08	0.03
0.5	0.2	250	DIS	0.15	0.04	-3.62	0.24	0.19	6.20	0.43	0.19	-2.52	0.45	0.22	5.99
			COB	-0.02	0.03	-2.12	0.18	0.19	4.79	0.03	0.17	-0.08	0.08	0.20	0.73
		1000	DIS	0.12	0.02	-3.46	0.15	0.09	5.82	0.42	0.17	-4.41	0.43	0.19	7.28
			COB	-0.01	0.01	-0.72	0.09	0.09	2.58	0.04	0.15	-0.03	0.06	0.17	0.09
	0.4	250	DIS	0.14	0.01	-2.43	0.24	0.18	4.78	0.43	0.09	-3.55	0.44	0.12	6.04
			COB	-0.01	0.00	-0.96	0.18	0.18	3.05	0.01	0.08	-0.01	0.07	0.11	0.10
		1000	DIS	0.11	0.00	-0.54	0.15	0.09	1.64	0.42	0.07	-5.26	0.42	0.09	6.97
			COB	-0.01	-0.01	-0.06	0.09	0.09	0.36	0.01	0.05	-0.01	0.04	0.07	0.04

Table 6.10: Simulation results order of integration — stationary random level-shift process and fractional cointegration.
In analogy to Table 6.3, the table reports bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally cointegrated system with cointegration vector $\beta = (1, -1)'$. The DGP is based on Equations (6.29) to (6.33). Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

Deterministic Trend

For the simulation results concerning processes contaminated with deterministic trends, we simulate the following process with common trends

$$y_{1t} = x_t + (1-L)^{-(d-\tilde{d})} u_t + t/T - 1/2 \tag{6.34}$$

$$y_{2t} = x_t + t/T - 1/2, \tag{6.35}$$

$$\text{where} \quad x_t = (1-L)^{-d} e_t, \text{ and } \begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right), \tag{6.36}$$

and the following process with distinct trends

$$y_{1t} = x_t + (1-L)^{-(d-\tilde{d})} u_t + t/T - 1/2 \tag{6.37}$$

$$y_{2t} = x_t + \sin(4\pi t/T), \tag{6.38}$$

$$\text{where} \quad x_t = (1-L)^{-d} e_t, \text{ and } \begin{pmatrix} e_t \\ u_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right). \tag{6.39}$$

Both trends were also considered by Qu (2011) in his simulation.

Again, we present results for $d = 0.2, 0.4$, $r = 0, 0.5$, $T = 250, 1000$, $\tilde{d} = 0, d$, the case of common trends, and the case of distinct trends. We omit the case of no trend as the results are identical to those presented in Section 6.5.

The results concerning the investigation of the fractional cointegration relation can be found in Table 6.11 and the results for the estimation of the order of integration in Tables 6.12 and 6.13. Again, the trimmed rank estimator and the trimmed multivariate local Whittle estimator perform substantially better than the non-robust approaches yielding valid inference also for series that exhibit deterministic trends.

			Cointegration	TRUE				FALSE			
r	d	T	Trend	TRE	RY02	CH06	SRF	TRE	RY02	CH06	SRF
0	0.2	250	DIS	1.00	0.23	0.04	0.10	0.00	0.00	0.58	0.08
			COT	1.00	1.00	0.48	0.27	0.00	0.00	0.34	0.03
		1000	DIS	1.00	0.04	0.56	0.85	0.00	0.00	0.77	0.57
			COT	1.00	1.00	0.91	0.74	0.00	0.00	0.55	0.14
	0.4	250	DIS	1.00	0.90	0.05	0.13	0.00	0.00	0.34	0.02
			COT	1.00	1.00	0.85	0.48	0.00	0.00	0.26	0.01
		1000	DIS	1.00	1.00	0.09	0.79	0.00	0.00	0.32	0.04
			COT	1.00	1.00	1.00	0.93	0.00	0.00	0.23	0.01
0.5	0.2	250	DIS	1.00	0.70	0.00	0.04	0.34	0.00	0.27	0.06
			COT	1.00	1.00	0.20	0.20	0.36	0.14	0.22	0.06
		1000	DIS	1.00	0.51	0.33	0.65	0.01	0.00	0.70	0.51
			COT	1.00	1.00	0.68	0.61	0.01	0.00	0.37	0.22
	0.4	250	DIS	1.00	0.99	0.01	0.07	0.34	0.02	0.14	0.02
			COT	1.00	1.00	0.66	0.40	0.36	0.10	0.14	0.02
		1000	DIS	1.00	1.00	0.05	0.61	0.01	0.00	0.15	0.03
			COT	1.00	1.00	1.00	0.88	0.01	0.00	0.12	0.04

Table 6.11: Simulation results fractional cointegration rank — deterministic trend.

In analogy to Table 6.1, the table reports the mean estimated fractional cointegration rank for a bivariate fractionally integrated system. The DGP is based on Equations (6.34) to (6.39). In case of fractional cointegration, $\beta = (-1, 1)'$ and $\tilde{d} = d$. Our estimator (TRE) is considered with $l = T^{0.5}$, $m = T^{0.75}$, $m_1 = T^{0.7}$, $N = m_1^{-0.2}$, and the univariate estimator by Iacone (2010) to estimate d. For the procedures by Robinson and Yajima (2002) (RY02), Chen and Hurvich (2006) (CH06), and Souza et al. (2018) (SRF), parameter values are chosen according to the authors recommendation.

			Estimator	TMLW				GSE			
				Bias		RMSE		Bias		RMSE	
r	d	T	Trend	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
0	0.2	250	DIS	-0.02	-0.03	0.19	0.20	0.04	0.19	0.08	0.20
			COT	-0.02	-0.02	0.20	0.20	0.04	0.04	0.08	0.08
		1000	DIS	-0.01	-0.02	0.09	0.09	0.05	0.17	0.06	0.17
			COT	-0.01	-0.01	0.09	0.09	0.05	0.05	0.06	0.06
	0.4	250	DIS	-0.03	-0.04	0.20	0.20	0.00	0.08	0.07	0.10
			COT	-0.03	-0.04	0.19	0.20	0.00	0.00	0.08	0.07
		1000	DIS	-0.02	-0.02	0.09	0.09	0.00	0.05	0.04	0.06
			COT	-0.02	-0.02	0.09	0.09	0.00	0.00	0.04	0.04
0.5	0.2	250	DIS	-0.02	-0.02	0.15	0.15	0.10	0.21	0.12	0.22
			COT	-0.02	-0.02	0.15	0.16	0.02	0.02	0.06	0.07
		1000	DIS	-0.01	-0.01	0.07	0.07	0.10	0.18	0.10	0.19
			COT	-0.02	-0.01	0.07	0.07	0.03	0.03	0.05	0.05
	0.4	250	DIS	-0.03	-0.04	0.16	0.15	0.04	0.09	0.08	0.11
			COT	-0.04	-0.04	0.16	0.15	0.00	0.00	0.06	0.06
		1000	DIS	-0.02	-0.02	0.07	0.07	0.03	0.06	0.05	0.07
			COT	-0.02	-0.02	0.07	0.07	0.00	0.00	0.04	0.04

Table 6.12: Simulation results order of integration — deterministic trend and no fractional cointegration.

In analogy to Table 6.2, the table reports bias and RMSE of our trimmed multivariate local Whittle estimator (TRE) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally integrated system. The DGP is based on Equations (6.34) to (6.39). Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

			Estimator	TMLW						GSE					
				Bias			RMSE			Bias			RMSE		
r	d	T	Trend	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$
0	0.2	250	DIS	-0.01	-0.03	-0.28	0.19	0.19	2.83	0.32	0.18	2.23	0.32	0.19	5.23
			COT	-0.02	-0.02	-0.25	0.19	0.19	2.72	-0.02	0.04	0.00	0.08	0.08	0.70
		1000	DIS	-0.01	-0.02	-0.06	0.09	0.09	1.38	0.32	0.15	3.23	0.32	0.15	6.10
			COT	-0.01	-0.01	-0.02	0.09	0.09	1.41	-0.01	0.05	0.00	0.04	0.06	0.10
	0.4	250	DIS	-0.01	-0.04	-0.07	0.19	0.20	1.90	0.34	0.07	-0.05	0.35	0.10	2.68
			COT	-0.02	-0.04	-0.13	0.19	0.19	2.01	-0.02	0.00	0.00	0.08	0.08	0.19
		1000	DIS	-0.01	-0.02	0.00	0.09	0.09	0.21	0.35	0.04	-0.53	0.35	0.06	1.75
			COT	-0.01	-0.02	0.00	0.09	0.09	0.25	-0.01	0.00	0.00	0.04	0.04	0.05
0.5	0.2	250	DIS	-0.02	-0.02	-2.66	0.18	0.18	5.36	0.40	0.21	-7.05	0.40	0.21	8.71
			COT	-0.03	-0.01	-2.68	0.18	0.18	5.40	0.00	0.04	-0.26	0.07	0.08	1.37
		1000	DIS	-0.01	-0.01	-1.07	0.09	0.09	3.17	0.36	0.18	-10.09	0.37	0.18	10.46
			COT	-0.01	-0.01	-0.90	0.09	0.09	2.90	0.01	0.04	-0.05	0.04	0.06	0.22
	0.4	250	DIS	-0.02	-0.04	-1.21	0.19	0.18	3.39	0.40	0.09	-4.73	0.41	0.10	6.31
			COT	-0.03	-0.03	-1.08	0.18	0.18	3.22	-0.01	0.00	-0.03	0.07	0.07	0.15
		1000	DIS	-0.01	-0.02	-0.09	0.09	0.09	0.55	0.38	0.05	-3.97	0.38	0.06	5.20
			COT	-0.01	-0.02	-0.07	0.09	0.09	0.41	0.00	0.00	-0.01	0.04	0.04	0.05

Table 6.13: Simulation results order of integration — deterministic trend and fractional cointegration.

In analogy to Table 6.3, the table reports bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally cointegrated system with cointegration vector $\beta = (1, -1)'$. The DGP is based on Equations (6.34) to (6.39). Moreover, we use $m = T^{0.75}$, $l = 1$ for the standard estimator, and $l = T^{0.5}$ for our procedure.

Estimating d when rk is unknown

In Section 6.5.2, we presented results of the GSE and TMLW estimator when the fractional cointegration rank is known as otherwise differences of the estimates might be caused by differences in the rank estimates. In empirical applications, however, the fractional cointegration is commonly unknown and needs to be estimated beforehand. To show that our procedure is also applicable in this case, Tables 6.14 and 6.15 show bias and RMSE of GSE and TMLW estimator when the fractional cointegration rank is not assumed to be known but estimated beforehand. For the non-robust GSE estimator, we consider estimating the fractional cointegration rank using the estimator by Robinson and Yajima (2002) and for the TMLW estimator we consider our trimmed estimator as presented in Section 6.3. The DGP is again based on Equations (6.5) to (6.9) and parameter values are also chosen as in Section 6.5.

The tables reveal that the results for the TMLW estimator are almost identical to those presented in Section 6.5.2. This is due to the fact that the robust rank estimator correctly identifies the fractional cointegration rank in almost all cases as indicated in Table 6.1. For the GSE estimator, the results change slightly displaying an increase in bias in some situations and a decrease in bias in others. However, the conclusion remains the same. The non-robust GSE estimator is severely biased when the series exhibit distinct or common low-frequency contaminations while the robust TMLW is unbiased in case of common low-frequency contaminations and slightly positively biased in case of distinct low-frequency contaminations. This bias vanishes asymptotically which is not the case for the bias of the GSE estimator.

			Estimator	TMLW				GSE			
				Bias		RMSE		Bias		RMSE	
r	d	T	Breaks	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_1$	$\hat{d}_2$
0	0.2	250	NO	-0.03	-0.02	0.20	0.20	-0.01	-0.02	0.07	0.08
			DIS	0.01	0.01	0.20	0.20	0.18	0.18	0.21	0.22
			COB	0.01	0.01	0.19	0.20	0.16	0.17	0.20	0.20
		1000	NO	-0.02	-0.01	0.09	0.09	-0.01	-0.01	0.04	0.04
			DIS	0.01	0.01	0.09	0.09	0.17	0.17	0.19	0.19
			COB	0.01	0.01	0.09	0.09	0.16	0.16	0.18	0.18
	0.4	250	NO	-0.04	-0.04	0.20	0.20	-0.02	-0.02	0.08	0.08
			DIS	-0.01	-0.01	0.20	0.20	0.08	0.08	0.12	0.12
			COB	-0.01	-0.01	0.20	0.20	0.07	0.07	0.11	0.12
		1000	NO	-0.02	-0.02	0.09	0.09	-0.01	-0.01	0.04	0.04
			DIS	-0.01	-0.01	0.09	0.09	0.06	0.06	0.09	0.09
			COB	-0.01	-0.01	0.09	0.09	0.06	0.06	0.08	0.08
0.5	0.2	250	NO	-0.04	-0.02	0.17	0.17	-0.02	-0.01	0.07	0.07
			DIS	0.01	0.03	0.17	0.17	0.20	0.21	0.23	0.24
			COB	-0.02	0.02	0.18	0.17	0.05	0.16	0.11	0.20
		1000	NO	-0.01	-0.01	0.07	0.07	-0.01	-0.01	0.04	0.04
			DIS	0.01	0.02	0.08	0.08	0.19	0.19	0.21	0.21
			COB	0.00	0.00	0.07	0.07	0.12	0.14	0.14	0.16
	0.4	250	NO	-0.05	-0.03	0.18	0.17	-0.02	-0.01	0.07	0.07
			DIS	-0.01	0.00	0.18	0.17	0.10	0.10	0.13	0.13
			COB	-0.04	0.00	0.18	0.17	0.02	0.07	0.08	0.11
		1000	NO	-0.02	-0.02	0.07	0.07	-0.01	-0.01	0.04	0.04
			DIS	0.00	0.00	0.07	0.07	0.07	0.07	0.09	0.09
			COB	-0.01	-0.01	0.07	0.07	0.04	0.04	0.06	0.07

Table 6.14: Simulation results estimating d when rk is unknown — no fractional cointegration.

In analogy to Table 6.2, the table shows Bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally integrated system with the DGP being based on Equations (6.5) to (6.9). Now, however, the fractional cointegration rank is not assumed to be known but estimated beforehand. Based on this estimate we then estimate the order of integration in a second step. Parameter values are identical to those considered in Tables 6.1 and 6.2.

			Estimator	TMLW						GSE					
				Bias			RMSE			Bias			RMSE		
r	d	T	Breaks	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$	$\widehat{d-\tilde{d}}$	$\hat{d}$	$\hat{\beta}$
0	0.2	250	NO	-0.02	-0.03	-0.28	0.20	0.20	2.81	-0.02	-0.02	0.00	0.08	0.08	1.12
			DIS	0.08	0.02	0.07	0.22	0.19	2.86	0.35	0.21	0.05	0.37	0.25	2.75
			COB	-0.02	0.01	-0.20	0.20	0.19	2.60	-0.02	0.18	-0.01	0.08	0.22	0.37
		1000	NO	-0.01	-0.02	-0.02	0.09	0.09	1.21	-0.01	-0.01	0.00	0.04	0.04	0.16
			DIS	0.06	0.01	0.11	0.12	0.09	1.85	0.36	0.20	-0.09	0.37	0.22	2.34
			COB	-0.01	0.01	-0.04	0.09	0.10	1.11	-0.01	0.17	0.00	0.04	0.20	0.06
	0.4	250	NO	-0.02	-0.04	-0.13	0.20	0.20	2.12	-0.02	-0.01	0.00	0.08	0.08	0.12
			DIS	0.08	-0.01	0.02	0.22	0.19	2.21	0.38	0.09	-0.10	0.40	0.13	2.30
			COB	-0.02	-0.02	-0.12	0.20	0.20	1.89	-0.02	0.08	0.00	0.08	0.13	0.08
		1000	NO	-0.01	-0.02	-0.01	0.09	0.09	0.32	-0.01	-0.01	0.00	0.04	0.04	0.05
			DIS	0.06	-0.01	-0.02	0.12	0.09	0.58	0.39	0.07	-0.07	0.40	0.09	2.17
			COB	-0.01	-0.01	0.00	0.09	0.09	0.18	-0.01	0.06	0.00	0.04	0.09	0.03
0.5	0.2	250	NO	-0.03	-0.02	-2.67	0.19	0.18	5.38	-0.02	-0.02	-0.49	0.08	0.07	2.07
			DIS	0.12	0.04	-3.41	0.24	0.19	6.03	0.41	0.24	-3.48	0.42	0.27	6.14
			COB	-0.02	0.02	-2.29	0.19	0.19	4.99	0.03	0.18	-0.10	0.08	0.22	0.83
		1000	NO	-0.02	-0.01	-0.92	0.09	0.09	2.95	-0.01	-0.01	-0.05	0.04	0.04	0.37
			DIS	0.10	0.02	-3.11	0.15	0.09	5.53	0.40	0.23	-5.01	0.41	0.26	7.29
			COB	-0.01	0.01	-0.75	0.09	0.09	2.67	0.04	0.16	-0.03	0.06	0.19	0.10
	0.4	250	NO	-0.02	-0.03	-1.10	0.19	0.19	3.23	-0.02	-0.01	-0.03	0.08	0.07	0.17
			DIS	0.11	0.00	-2.18	0.23	0.18	4.58	0.42	0.11	-3.52	0.43	0.15	5.73
			COB	-0.02	0.00	-0.98	0.19	0.18	3.09	0.01	0.08	-0.02	0.07	0.12	0.11
		1000	NO	-0.01	-0.02	-0.07	0.09	0.09	0.38	-0.01	-0.01	-0.01	0.04	0.04	0.05
			DIS	0.10	0.00	-0.54	0.15	0.09	1.71	0.42	0.09	-4.63	0.42	0.12	6.47
			COB	-0.01	-0.01	-0.07	0.09	0.09	0.49	0.02	0.06	-0.01	0.04	0.08	0.04

Table 6.15: Simulation results estimating d when rk is unknown — fractional cointegration.
In analogy to Table 6.3, the table shows Bias and RMSE of our trimmed multivariate local Whittle estimator (TMLW) and the standard multivariate local Whittle estimator (GSE) in a bivariate fractionally cointegrated system with the DGP being based on Equations (6.5) to (6.9). Now, however, the fractional cointegration rank is not assumed to be known but estimated beforehand. Based on this estimate we then estimate the order of integration in a second step. Parameter values are identical to those considered in Tables 6.1 and 6.3.

Chapter 7

memochange: an R Package for Estimation and Tests in Persistent Time Series

Co-authored with Kai Wenger.
Published in The Journal of Open Source Software (2019), 4(43), 1820.

https://doi.org/10.21105/joss.01820

Bibliography

Abadir, K., Distaso, W., and Giraitis, L. (2009). Two estimators of the long-run variance: beyond short memory. *Journal of Econometrics*, 150(1):56–70.

Adrian, T. and Franzoni, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance*, 16(4):537–556.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics*, 74(3):487–528.

Aït-Sahalia, Y. and Jacod, J. (2009). Testing for jumps in a discretely observed process. *Annals of Statistics*, 37(1):184–222.

Ait-Sahalia, Y., Kalnina, I., and Xiu, D. (2019). High-frequency factor models and regressions. *Chicago Booth Research Paper*, (19–04).

Aït-Sahalia, Y., Mykland, P., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175.

Alfarano, S. and Lux, T. (2007). A noise trader model as a generator of apparent financial power laws and long memory. *Macroeconomic Dynamics*, 11(S1):80–101.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1):135–167.

Andersen, T. and Benzoni, L. (2009). Realized volatility. In *Handbook of Financial Time Series*, pages 555–575. Springer.

Andersen, T. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3):115–158.

Andersen, T. and Bollerslev, T. (1998a). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905.

Andersen, T. and Bollerslev, T. (1998b). Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*, 53(1):219–265.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T., Bollerslev, T., Diebold, F., and Wu, G. (2005). A framework for exploring the macroeconomic determinants of systematic risk. *American Economic Review: Papers and Proceedings*, 95(2):398–404.

Andersen, T., Bollerslev, T., Diebold, F., and Wu, G. (2006). Realized beta: Persistence and predictability. In *Econometric Analysis of Financial and Economic Time Series*, pages 1–39. Emerald Group Publishing Limited.

Andersen, T., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1):75–93.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Ang, A. and Chen, J. (2007). CAPM over the long run: 1926–2001. *Journal of Empirical Finance*, 14(1):1–40.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006a). The cross-section of volatility and expected returns. *Journal of Finance*, 61(1):259–299.

Ang, A., Piazzesi, M., and Wei, M. (2006b). What does the yield curve tell us about gdp growth? *Journal of Econometrics*, 131(1):359–403.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.

Arteche, J. (2004). Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics*, 119(1):131–154.

Asai, M. and McAleer, M. (2015). Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance. *Journal of Econometrics*, 189(2):251–262.

Atkins, A. and Dyl, E. (1990). Price reversals, bid-ask spreads, and market efficiency. *Journal of Financial and Quantitative Analysis*, 25(4):535–547.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Baillie, R., Bollerslev, T., and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

Bajgrowicz, P. and Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics*, 106(3):473–491.

Bajgrowicz, P., Scaillet, O., and Treccani, A. (2015). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, 62(8):2198–2217.

Bandi, F. and Russell, J. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692.

Bandi, F. and Russell, J. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75(2):339–369.

Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3–18.

Barber, B., Huang, X., and Odean, T. (2016). Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies*, 29(10):2600–2642.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., and Shephard, N. (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., Hansen, P., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, 12(3):1–32.

Barndorff-Nielsen, O. and Shephard, N. (2001). Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241.

Barndorff-Nielsen, O. and Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Barndorff-Nielsen, O. and Shephard, N. (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

Barndorff-Nielsen, O. and Shephard, N. (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. In *Advances in Economics and Econometrics. Theory and Applications*, pages 328–372. Cambridge University Press.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long memory processes: Probabilistic properties and statistical methods*. Springer London, Limited.

Berk, J. and Van Binsbergen, J. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1–23.

Bhandari, L. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, 43(2):507–528.

Bhattacharya, R., Gupta, V., and Waymire, E. (1983). The hurst effect under trends. *Journal of Applied Probability*, 20(3):649–662.

Black, A., Fraser, P., and Power, D. (1992). UK unit trust performance 1980–1989: A passive time-varying approach. *Journal of Banking & Finance*, 16(5):1015–1033.

Blume, M. (1971). On the assessment of risk. *Journal of Finance*, 26(1):1–10.

Boehme, R., Danielsen, B., and Sorescu, S. (2006). Short-sale constraints, differences of opinion, and overvaluation. *Journal of Financial and Quantitative Analysis*, 41(2):455–487.

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

Bollerslev, T., Li, S., and Todorov, V. (2016). Roughing up beta: Continuous versus discontinuous betas and the cross-section of expected stock returns. *Journal of Financial Economics*, 120(3):464–490.

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H. (2014). Stock return predictability and variance risk premia: statistical inference and international evidence. *Journal of Financial and Quantitative Analysis*, 49(3):633–661.

Bollerslev, T. and Mikkelsen, H. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492.

Boyer, B., Mitton, T., and Vorkink, K. (2009). Expected idiosyncratic skewness. *Review of Financial Studies*, 23(1):169–202.

Breidt, J., Crato, N., and De Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(1-2):325–348.

Brennan, M. and Zhang, Y. (2018). Capital asset pricing with a stochastic horizon. *Journal of Financial and Quantitative Analysis, forthcoming.*

Brown, B. (1974). Generalized wald equations in discrete time. *Stochastic Processes and their Applications*, 2(4):349–357.

Brown, K. and Harlow, W. (1988). Market overreaction: magnitude and intensity. *Journal of Portfolio Management*, 14(2):6–13.

Buss, A. and Vilkov, G. (2012). Measuring equity risk with option-implied correlations. *Review of Financial Studies*, 25(10):3113–3140.

Cameron, C., Gelbach, J., and Miller, D. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Chan, H. and Palma, W. (1998). State space modeling of long-memory processes. *Annals of Statistics*, 26(2):719–740.

Chen, W. and Hurvich, C. (2006). Semiparametric estimation of fractional cointegrating subspaces. *Annals of Statistics*, 34(6):2939–2979.

Chernov, M. (2007). On the role of risk premia in volatility forecasting. *Journal of Business & Economic Statistics*, 25(4):411–426.

Christensen, B. and Santucci de Magistris, P. (2010). Level shifts in volatility and the implied-realized volatility relation. *CREATES Research Paper*, (2010–60).

Christensen, K., Oomen, R., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

Christoffersen, P. (2003). *Elements of financial risk management.* Academic Press.

Christoffersen, P. and Diebold, F. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52(8):1273–1287.

Christoffersen, P., Diebold, F., Mariano, R., Tay, A., and Tse, Y. (2007). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *Journal of Financial Forecasting*, 1(2):1–22.

Cochrane, J. H. (2009). *Asset Pricing.* Princeton university press.

Corrado, C. and Truong, C. (2007). Forecasting stock index volatility: comparing implied volatility and the intraday high–low price range. *Journal of Financial Research*, 30(2):201–215.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Pirino, D., and Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Cox, D. and Peterson, D. (1994). Stock returns following large one-day declines: Evidence on short-term reversals and longer-term performance. *Journal of Finance*, 49(1):255–267.

Daniel, K., Mota, L., Rottke, S., and Santos, T. (2018). The cross-section of risk and return. *NBER Working Paper*, (24164).

De Groot, W., Huij, J., and Zhou, W. (2012). Another look at trading costs and short-term reversal profits. *Journal of Banking & Finance*, 36(2):371–382.

Deo, R. and Hurvich, C. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory*, 17(4):686–710.

Diebold, F. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Ding, Z. and Granger, C. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1):185–215.

Dissanayake, G., Peiris, M., and Proietti, T. (2016). State space modeling of gegenbauer processes with long memory. *Computational Statistics & Data Analysis*, 100:115–130.

Engle, R. and Smith, A. (1999). Stochastic permanent breaks. *Review of Economics and Statistics*, 81(4):553–574.

Epps, T. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. and French, K. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fama, E. and MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.

Frazzini, A. and Pedersen, L. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1–25.

Frederiksen, P., Nielsen, F., and Nielsen, M. (2012). Local polynomial whittle estimation of perturbed fractional processes. *Journal of Econometrics*, 167(2):426–447.

Gao, L., Han, Y., Li, S., and Zhou, G. (2018). Market intraday momentum. *Journal of Financial Economics*, 129(2):394–414.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

Graham, J. and Harvey, C. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, 60(2):187–243.

Granger, C. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73(1):61–77.

Granger, C. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.

Granger, C. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

Grundy, B. and Martin, S. (2001). Understanding the nature of the risks and the source of the rewards to momentum investing. *Review of Financial Studies*, 14(1):29–78.

Guégan, D. (2005). How can we define the concept of long memory? An econometric survey. *Econometric Reviews*, 24(2):113–149.

Han, H. and Kristensen, D. (2014). Asymptotic theory for the qmle in garch-x models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics*, 32(3):416–429.

Han, H., Linton, O., Oka, T., and Whang, Y. (2016). The cross-quantilogram: measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics*, 193(1):251–270.

Hansen, P. and Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3(4):525–554.

Hansen, P. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *Journal of Finance*, 64(3):1445–1477.

Hassler, U. and Pohle, M.-O. (2019). Forecasting under long memory and nonstationarity. *Goethe University Working Paper.*

He, Z., Zhu, J., and Zhu, X. (2015). Multi-factor volatility and stock returns. *Journal of Banking & Finance*, 61:132–149.

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). Oxford-man institute's realized library, version 0.2. *Oxford-Man Institute, University of Oxford.*

Hendry, D. and Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business & Economic Statistics*, 25(1):33–51.

Hollstein, F. and Prokopczuk, M. (2016). Estimating beta. *Journal of Financial and Quantitative Analysis*, 51(4):1437–1466.

Hollstein, F., Prokopczuk, M., and Wese Simen, C. (2019a). The conditional Capital Asset Pricing Model revisited: Evidence from high-frequency betas. *Management Science, forthcoming.*

Hollstein, F., Prokopczuk, M., and Wese Simen, C. (2019b). Estimating beta: Forecast adjustments and the impact of stock characteristics for a broad cross-section. *Journal of Financial Markets*, 44:91–118.

Hosking, J. (1981). Fractional differencing. *Biometrika*, 68(1):165–176.

Hou, J. and Perron, P. (2014). Modified local whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, 182(2):309–328.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705.

Hurvich, C., Moulines, E., and Soulier, P. (2005). Estimating long memory in volatility. *Econometrica*, 73(4):1283–1328.

Iacone, F. (2010). Local whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis*, 31(1):37–49.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Kamara, A., Korajczyk, R., Lou, X., and Sadka, R. (2016). Horizon pricing. *Journal of Financial and Quantitative Analysis*, 51(6):1769–1793.

Kara, Y., Boyacioglu, M., and Baykan, Ö. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert Systems with Applications*, 38(5):5311–5319.

Kellard, N., Jiang, Y., and Wohar, M. (2015). Spurious long memory, uncommon breaks and the implied–realized volatility puzzle. *Journal of International Money and Finance*, 56:36–54.

Koijen, R., Lustig, H., and Van Nieuwerburgh, S. (2017). The cross-section and time series of stock and bond returns. *Journal of Monetary Economics*, 88:50–69.

Künsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability*, 23(4):1025–1030.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1):33–59.

LeBaron, B. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*, 1(6):621–631.

LeBaron, B. (2006). Agent-based financial markets: Matching stylized facts with style. In *Post Walrasian Macroeconomics: Beyond the DSGE Model*, pages 221–235. Cambridge University Press Cambridge, UK.

Lee, S. and Mykland, P. (2007). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535–2563.

Leschinski, C. (2017). On the memory of products of long range dependent time series. *Economics Letters*, 153:72–76.

Levi, Y. and Welch, I. (2017). Best practice for cost-of-capital estimates. *Journal of Financial and Quantitative Analysis*, 52(2):427–463.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37.

Linton, O. and Whang, Y. (2007). The quantilogram: with an application to evaluating directional predictability. *Journal of Econometrics*, 141(1):250–282.

Liu, C. and Maheu, J. (2007). Are there structural breaks in realized volatility? *Journal of Financial Econometrics*, 6(3):326–360.

Liu, J., Stambaugh, R., and Yuan, Y. (2018). Absolving beta of volatility's effects. *Journal of Financial Economics*, 128(1):1–15.

Lo, A. (2004). The adaptive markets hypothesis. *Journal of Portfolio Management*, 30(5):15–29.

Lobato, I. (1999). A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, 90(1):129–153.

Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica*, pages 1429–1445.

Malkiel, B. (1973). *A random walk down Wall Street.* WW Norton & Company.

Marinucci, D. and Robinson, P. (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics*, 105(1):225–247.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

McCloskey, A. and Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory*, 29(6):1196–1237.

McGown, K. and Parks, H. (2007). The generalization of faulhaber's formula to sums of non-integral powers. *Journal of Mathematical Analysis and Applications*, 330(1):571–575.

Mincer, J. and Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pages 3–46. NBER.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *Journal of Finance*, 72(4):1611–1644.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783.

Müller, U., Dacorogna, M., Davé, R., Pictet, O., Olsen, R., and Ward, R. (1993). Fractals and intrinsic time: A challenge to econometricians. *39th International AEA Conferenceon Real Time Econometrics, Luxembourg.*

Nielsen, M. (2007). Local Whittle analysis of stationary fractional cointegration and the implied–realized volatility relation. *Journal of Business & Economic Statistics*, 25(4):427–446.

Nielsen, M. and Shimotsu, K. (2007). Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. *Journal of Econometrics*, 141(2):574–596.

Nyberg, H. (2011). Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting*, 27(2):561–578.

Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficients. *Journal of Econometrics*, 13(3):341–363.

Patton, A. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246 – 256.

Patton, A. and Verardo, M. (2012). Does beta move with news? Firm-specific information flows and learning about profitability. *Review of Financial Studies*, 25(9):2789–2839.

Pesaran, H. and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4):461–465.

Qiu, M. and Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. *PloS one*, 11(5).

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics*, 29(3):423–438.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In *Handbook of Economic Forecasting*, volume 2, pages 328–383. Elsevier.

Regnault, J. (1863). *Calcul des chances et philosophie de la bourse*. Mallet-Bachelier.

Robinson, P. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association*, 89(428):1420–1437.

Robinson, P. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23(5):1630–1661.

Robinson, P. (2005). Robust covariance matrix estimation: Hac estimates with long memory/antipersistence correction. *Econometric Theory*, 21(1):171–180.

Robinson, P. (2008). Multiple local Whittle estimation in stationary systems. *Annals of Statistics*, 36(5):2508–2530.

Robinson, P. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, 106(2):217–241.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Ross, S. (2009). *Neoclassical finance*. Princeton University Press.

Scholes, M. and Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics*, 5(3):309–327.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

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

Shimotsu, K. (2007). Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics*, 137(2):277–310.

Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(2):501–540.

Shimotsu, K. (2012). Exact local whittle estimation of fractionally cointegrated systems. *Journal of Econometrics*, 169(2):266–278.

Shimotsu, K. and Phillips, P. (2005). Exact local whittle estimation of fractional integration. *Annals of Statistics*, 33(4):1890–1933.

Sibbertsen, P., Leschinski, C., and Busch, M. (2018). A multivariate test against spurious long memory. *Journal of Econometrics*, 203(01):33–49.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London.

Simes, J. (1986). An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

Souza, I., Reisen, V., Franco, G., and Bondon, P. (2018). The estimation and testing of the cointegration order based on the frequency domain. *Journal of Business & Economic Statistics*, 36(4):695–704.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stambaugh, R. and Yuan, Y. (2016). Mispricing factors. *Review of Financial Studies*, 30(4):1270–1315.

Stock, J. and Watson, M. (2011). Dynamic factor models. In *Oxford Handbook on Economic Forecasting*. Oxford University Press.

Sun, Y. and Phillips, P. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, 115(2):355–389.

Timmermann, A. and Granger, C. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1):15–27.

Vasicek, O. (1973). A note on using cross-sectional information in Bayesian estimation of security betas. *Journal of Finance*, 28(5):1233–1239.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Wenger, K., Leschinski, C., and Sibbertsen, P. (2018). The memory of volatility. *Quantitative Finance and Economics*, 2(1):137 – 159.

Zhang, X. (2006). Information uncertainty and stock returns. *Journal of Finance*, 61(1):105–137.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.

Zhou, G. (2010). How much stock return predictability can we expect from an asset pricing model? *Economics Letters*, 108(2):184–186.