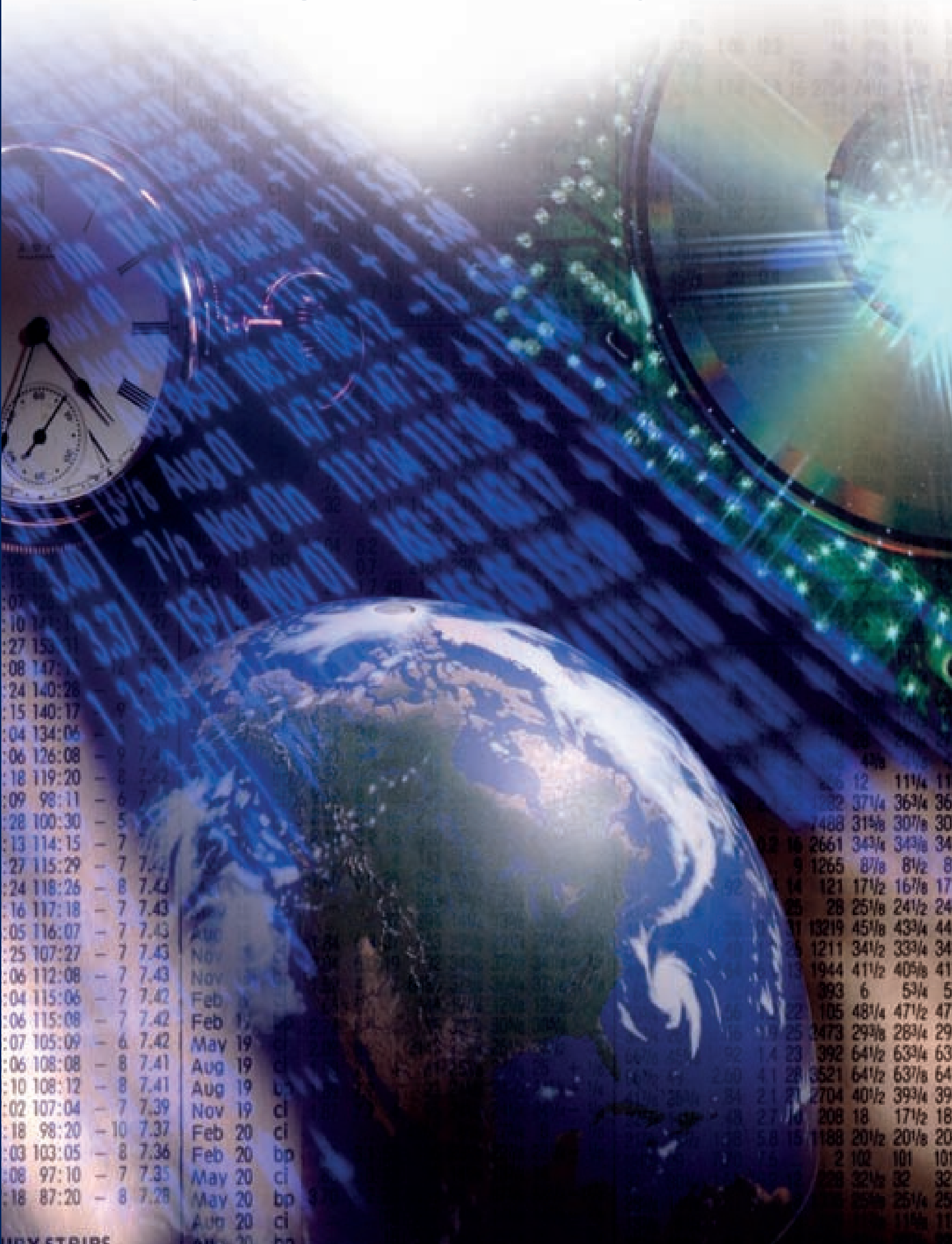
KARIM BANNOUH

Measuring and Forecasting Financial Market Volatility using High-Frequency Data



Measuring and Forecasting Financial Market Volatility using High-Frequency Data

Measuring and Forecasting Financial Market Volatility using High-Frequency Data

Meten en voorspellen van financiële markt volatiliteit met hoge-frequentie data

THESIS

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board
The public defence shall be held on

Friday 11 January 2013 at 09.30 hours
by

Karim Bannouh
born in De Bilt

Doctoral Committee

Promotor:	Prof.dr. D.J.C. van Dijk
Other members:	Prof.dr. H.P. Boswijk Prof.dr. M.J. McAleer Prof.dr. R. Paap
Copromotor:	Dr. M.P.E. Martens

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at Erasmus University Rotterdam
Internet: `http://www.erim.eur.nl`

ERIM Electronic Series Portal: `http://hdl.handle.net/1765/1`

ERIM Ph.D. Series Research in Management, 273
ERIM Reference number: EPS-2013-273-F&A
ISBN 978-90-5892-317-2

Design: B&T Ontwerp en advies `www.b-en-t.nl`
Print: Haveka `www.haveka.nl`

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive®. The ink used is produced from renewable resources and alcohol free fountain solution. Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001. More info: http://www.haveka.nl/greening

Preface

During the completion of my master's thesis I was convinced that after completion I would look for a job in the financial industry. This changed when Dick van Dijk and Martin Martens suggested to stay a few years longer at university as a Ph.D. candidate conducting research on the use of high-frequency data for measuring and forecasting financial market volatility. At the time I knew that both Dick and Martin were active in this relatively new field in financial econometrics. Having greatly benefited from what I learned as Ph.D. candidate I am happy to continue to work at Saemor Capital where portfolio risk management is an integral part of the investment process.

First of all, I am heavily indebted to my (co-)promotors and supervisors Dick van Dijk and Martin Martens for their invaluable guidance and for giving me the opportunity to conduct Ph.D. research on measuring and modeling financial market volatility with high-frequency data. Second, I would like to thank Roel Oomen who co-authored Chapter 4 in this dissertation and shared valuable insights on the subject. Third, I want to thank all committee members, Peter Boswijk, Philip Hans Franses, Frank de Jong, Siem-Jan Koopman, Michael McAleer and Richard Paap.

I would like to express my gratitude to the Erasmus School of Economics (ESE) and the Erasmus Research Institute of Management (ERIM) at Erasmus University Rotterdam (EUR) for giving me the opportunity to participate in their joint doctoral research programme. I thank the secretarial staff at the Econometric Institute and ERIM, in particular Elli Hoek van Dijke, Marisa van Iperen and Olga Rook. I want to thank colleagues at the Econometric Institute, Paul de Boer, Kees Bouwman, Jan Brinkhuis, Herman van Dijk, Dennis Fok, Christiaan Heij, Flavius Frasincar, Lennart Hoogerheide, Erik Kole, Roger Lord, Martijn Mulder, Wing Wah Tam, Albert Wagelmans and Michel van der Wel.

During the years I enjoyed the presence of several fellow Ph.D. candidates, in particular I would like to thank Chen, Kyle, Nuno, Sjoerd and Viorel for the nice atmosphere on the 9th floor. I would also like to thank fellow doctoral candidates at the Erasmus Research Institute of Management and the Tinbergen Institute: Diana Budiono, Peter Exterkate, Kar Yin Lam, Rianne Legerstee, Milan Lovric, Haikun Ning, Arco van Oord, Anne Opschoor, Lenny Pattikawa, Martin Scholtus, Murat Tarakci, Guangyuan Yang and Dan Zhang. I want to thank former colleagues at the ING Model Validation department: Reinhardt Chamonal, David Chkadua, Loredana Mavrodin, Zhaoning Shang, Richard van Tilborgh and Gisella van Vollenhoven. At Saemor and Pelargos Capital, I would like to thank Sven Bakker, Sven Bouman, Hans van der Heiden, Louis Heij, Patrick van de Laar, Rani Piputri, Roelof Salomons and Richard Schreuder.

Bassim, Cem, David, Daniel, Francois, Gabriela, Konrad, Marco, Menno, Nalan, Neeradj, Norman, Rui, Simon and Thijs; thank you for your friendship! I thank my family for their unconditional love and support.

Karim Bannouh,
Rotterdam, 28 November 2012

Contents

Chapter 1

Introduction

1.1 Volatility

Volatility is a measure of the relative degree of change. In Finance volatility is often defined as the dispersion in asset price movements over a period of time. Financial market volatility plays an important role in financial economics and is at the heart of several subjects including asset allocation, market timing, risk management and the pricing of assets and derivatives.

Recent developments in financial markets such as for instance the bursting of the IT bubble, the US subprime mortgage crisis and Europe's ongoing sovereign debt debacles, exemplify the importance of adequate risk measurement and risk management techniques that adapt more rapidly to changing market circumstances than traditional methods do. Coincidentally, measuring and modeling volatility with intraday data has become a rapidly growing field in financial econometrics and applied statistics. In practice traditional portfolio risk models are based on low-frequency asset return data such as for instance sixty monthly observations or more progressively a few years of weekly or daily data. An inherent drawback of the use of risk models based on low-frequency data is that they are severely impacted by structural breaks and these models are not sufficiently adaptive to shifts in financial markets. For this reason the popularity of more accurate short term risk measures and risk models is rapidly increasing in the financial industry. Practically it is by now hard to find a large and respectable vendor of financial risk models that does not offer models that predict risk over relatively short horizons such as monthly, weekly or daily periods. The shift to risk models that utilize high-frequency intraday data is needed to provide practitioners with more accurate short-term risk measures and models. Besides academics certain risk vendors and quant groups within the asset management industry have started studying the practical merits of high frequency data for financial risk management. For instance De Rossi et al. (2012) discuss how low-frequency risk- and portfolio-managers benefit from hedging short-term risk factor exposures using high-frequency data estimates.

A problem that will diminish with time is the relatively short history of available high-frequency data. Although tick-data is available from 1980 onwards for a limited number of assets (FX, bond, stock and commodity futures), this is swamped by over a hundred years of daily data history that is available for thousands of individual securities such as stocks. The adoption of using high-frequency data for financial risk management in practice depends on several other factors as well. Academics can certainly help at several stages to accelerate the transition in the financial industry from low-frequency data risk measurement towards the use of more accurate measures and models based on high-frequency data.

Adequate standards and formal approaches to 'cleaning' high-frequency data are needed.[1] By now considerable progress has been achieved on understanding the theoretical properties of univariate risk measures based on high-frequency data, but the practical relevance of these applications can still be improved. To get regulators to endorse and approve the use of high-frequency data for risk measurement and risk reporting, applications that are appealing to regulators and risk departments of financial institutions could be considered more often. For instance applications dealing with (Conditional) Value-at-Risk measures, forecasts and models are appealing since they are linked to the control and reporting of financial risk, economic capital, regulatory capital and stress testing. In addition, the focus in the multivariate high-frequency literature on volatility estimation needs to move towards more meaningful multivariate dimensions rather than using bi-variate or low-dimensional settings. At the moment the most commonly encountered multivariate dimensions considered in the literature are in the range of two to thirty assets. Financial market participants such as hedge fund and mutual fund managers, however, generally hold substantially more assets. In addition, their investment performance is typically measured relative to market indices with hundreds of underlying assets. For these reasons the number of assets used in multivariate studies on measuring and forecasting financial risk with high-frequency data needs to be dramatically increased. It is surprising that until recently, no academic high-frequency data studies have considered the factor models that are successfully being used in practice and have been documented in

[1] Exceptions that provide the first steps in studying, reporting and formalizing the data cleaning procedures for high-frequency data include Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009).

the literature for decades (see e.g. Ross (1976)). The idea of combining the insights from factor models with the benefits of high-frequency data for accurate estimation of practically relevant vast dimensional covariance matrices remains overlooked.

1.2 Traditional Volatility Models: (G)ARCH and Stochastic Volatility

The vast collection of literature on financial market volatility dynamics was kick-started by Engle (1982) and Bollerslev (1986) who introduced the (Generalized) Autoregressive Conditional Heteroskedasticity – (G)ARCH models. Consistent with economic theory daily asset return predictions exhibit negligible or very little explanatory power. For this reason it is common practice to assume daily stock returns are unpredictable. The volatility of daily returns, however, is conditionally dependent and due to its relatively high persistence it is quite predictable, especially when compared to the predictability of daily stock returns. Mandelbrot (1963) already noticed that periods of high (low) volatility tend to be followed by periods of high (low) volatility. This time-varying volatility clustering and the predictability of volatility processes results in important implications for financial risk measurement and risk management. It is not surprising that the number of econometric contributions on forecasting financial market volatility is vast compared to contributions concerned with forecasting asset returns. The popularity of the (G)ARCH process is partially explained by its ability of incorporating volatility clustering in an intuitive way. In these models expected volatility today is proportional to the previously observed squared return(s). If the squared return was relatively large yesterday, then today's volatility is more likely to be high than is the case if the squared return was relatively small. Since the introduction of (G)ARCH many extensions and modifications have been proposed and the family of (G)ARCH models as such is one of the most important working horses in the field of financial econometrics.

Besides the expanding literature in financial econometrics on (G)ARCH type models, important contributions on understanding and measuring financial market volatility can be ascribed to the class of Stochastic Volatility (SV) models. Although both classes of models are stochastic in nature, the main difference is that in traditional ARCH type models the return process generally depends on one error process and the volatility process in turn depends on this return process. Hence, the volatility process is indirectly driven by the innovations in the return process. In contrast, SV models generally depend on two error processes; one error process in the return equation and an error process in the volatility equation.

Despite the important progress made on understanding, measuring and modeling of financial market volatility using (G)ARCH- and SV-type models, an important problem that remained unsolved for about a decade or two is that the conditional volatility in these models is unobservable and needs to be estimated. The traditional way of estimating volatility is based on using low-frequency asset return data. An important drawback of using data sampled at a low-frequencies, such as daily data, is that although this volatility estimator is unbiased, it is imprecise and the variation in daily volatility estimates is too large for certain applications. This for instance raised the debate on the quality of volatility forecasts generated using standard (G)ARCH-type models. Although these models achieve a high in-sample explanatory power, their out-of-sample forecasting performance is poor when judged on the explanatory power measured in running a regression of a series of realized daily squared returns on daily conditional volatility forecasts.

1.3 Realized Volatility

Merton (1980) notes that the variance of the returns on an asset over an extended period of time can be estimated with high precision if during that period a sufficient number of sub-period returns is available. Because the squared mean return converges to zero as the sampling frequency increases, the variance of the returns over an extended period can be calculated by summing the squared sub-period returns and ignoring the mean return. This is what today is called the concept of realized volatility and this term is interchangeably used with realized variance. Realized volatility type estimators have shown up in several 'early' studies without formal derivations being present at the time.[2]

The concept of realized volatility remained non-formalized until the work of Andersen and Bollerslev (1998), Bardorff-Nielsen and Shephard (2001) and Comte and Renault (1998). Andersen and Bollerslev (1998) show that volatility forecasts generated by (G)ARCH-type models perform satisfactorily after all if the unbiased but noisy daily squared return is replaced by realized volatility when determining the accuracy of volatility forecasts through regressing ex-post realizations on forecasts. Although the daily squared return is accurate on average, it is too noisy causing an underestimation of the explanatory power of potentially accurate volatility forecasts. The formalization of the concept of realized volatility combined with the increasing availability of high-frequency data has spurred the development of a novel and to-

[2]Examples of such studies include French et al. (1987), Schwert (1990), Schwert and Seguin (1990), Hsieh (1991), Zhou (1996) and Taylor and Xu (1997).

day still growing field within financial econometrics that is concerned with accurate measuring, modeling and forecasting of financial market volatility.

One of the most important advantages of using high-frequency data for volatility estimation is the improvement in statistical efficiency that results from the reduction in variance of the realized volatility estimator relative to the variance of the daily squared return estimator. Although in theory the precision of the realized volatility estimator is maximized by sampling asset returns as often as possible, this is not feasible in practice due to the existence of market microstructure frictions. Observed transaction prices at very high frequencies tend to randomly alternate between prices at bid- and ask-quotes. These transient high-frequency price fluctuations are a source of market microstructure noise for volatility estimation since they are independent of the asset's underlying volatility process. Hence, volatility estimates based on returns calculated from transaction prices sampled at very high frequencies are upward biased. The bias caused by bid-ask bounce increases with the sampling frequency. The increase in bias and reduction in variance of the estimator results in a trade-off in terms of statistical efficiency. To strike a balance between bias and precision the originally proposed sampling frequencies were in the 5- to 30-minutes range, motivated by volatility signature plots that display the variance or the mean squared error as a function of the sampling frequency.

Academic interest in the realized volatility literature has shifted to studies that take market microstructure noise into account. This resulted in considerable improvements for practical volatility estimation. On the one hand authors focused on optimizing the sampling frequency in the presence of noise, see e.g. Bandi and Russell (2008b), and bias-correction methods that allow and justify the use of higher sampling frequencies on the other hand, see e.g. Zhang et al. (2005). These recent advances in the academic literature on the use of high-frequency data for improved risk measurement enable practitioners to use risk models based on high-frequency data that can substantially expand their arsenal of risk management tools. Short term risk models based on high-frequency data are more precise, more adaptive and less impacted by structural breaks than their low-frequency data counterparts. Besides improved measurement of portfolio risk, practitioners are now starting to use the merits of high-frequency data at several stages of the integral risk management process. High-frequency data is for example also increasingly being studied by practitioners for improved estimation of dynamic factor exposures and hedging these over short horizons, see e.g. De Rossi et al. (2012).

The literature on measuring and forecasting financial market volatility remains an active field that continues to develop in several directions. Notable is the shift from the initial theoretical working assumptions towards including realistic market microstructure frictions. With the exception of non-trading, most studies that consider the impacts of market microstructure noise on volatility estimation with high-frequency data are confined to univariate settings. Very little is known about the complex structure of multivariate microstructure frictions arising from for instance large trades in basket instruments such as derivatives, market indices and ETFs. Transactions in these instruments are becoming increasingly more relevant than transactions in individual securities. Another potentially important source of market microstructure noise that can affect stocks in particular sectors or countries are recently implemented short-sell restrictions. The common working assumption of independence in microstructure frictions across assets is likely to be challenged. This especially applies to periods surrounding economic news announcements where often entire asset classes move in a certain direction and independence between the arrival of bid- and ask-transactions across underlying individual assets, especially those with similar factor exposures, is not likely to be realistic. The only study that the author is aware of on this topic is the recent work presented in Diebold and Strasser (2012) which opens the door to more work on this challenging but interesting subject that is potentially relevant for multivariate volatility estimation.

Another interesting and promising recently started development in the literature is the increase in high-frequency data studies and applications that are relevant for asset pricing. Recent examples of such studies involve the studying of asset class (variance) risk premia and systemic risks, see e.g. Andersen et al. (2007), Bollerslev et al. (2009), Bollerslev and Todorov (2011), Todorov and Bollerslev (2010) and Wright and Zhou (2009) for work along these lines. Also of importance is the application to asset classes other than stocks and FX rates. For instance, applications in fixed income and commodity markets are rare, which is surprising given the importance of such asset classes. Studies that do look beyond the traditional FX and stock market applications include Andersen et al. (2007), Busch et al. (2010), Duyvesteyn et al. (2011), Fleming et al. (2003), Liu et al. (2012) and Wright and Zhou (2009).

1.4 Parkinson's Range, Realized Range and Market Microstructure Noise

Besides estimating volatility measures using asset returns, it is also possible to use the high-low price range as a measure of return volatility. The daily high and low prices of an asset are often quoted in databases and media such as financial newspapers alongside quantities such as the closing price and the daily return. The daily high-low range is the difference between the highest and lowest price at which transactions are executed and recorded. Parkinson (1980) illustrates that in a world with continuous trading and without market frictions the daily high-low range is a more efficient volatility estimator than the daily squared return since the variance of the range-based estimator is about five times smaller than the variance of the squared return estimator. Garman and Klass (1980) extend the work of Parkinson (1980) by including open and close prices which further increases the efficiency of the daily range-based volatility estimator. The Parkinson (1980) volatility estimator is biased if the underlying diffusion has a drift. Alternative drift-robust daily range estimators are proposed in Rogers and Satchell (1991), Kunitomo (1992) and Yang and Zhang (2000).

Martens and Van Dijk (2007) and Christensen and Podolskij (2007) in independent work propose the realized range as a highly efficient estimator of ex-post volatility. Similar to the concept of realized volatility being based on aggregating squared intraday returns, the realized range is based on aggregating squared intra-period high-low price ranges. The improvement in efficiency of the daily range relative to daily squared return continues to hold when returns and ranges are sampled at higher frequencies. The realized range is upward biased due to bid-ask bounce as is the case for the realized volatility estimator. An advantage of the realized range estimator of Martens and Van Dijk (2007), however, is that it is affected by an opposing source of noise. In the context of the daily range Garman and Klass (1980), Beckers (1983) and Marsh and Rosenfeld (1986) find that the volatility estimates are downward biased due to non-trading. In practice transaction prices are observed in discrete time and occur infrequently and irregularly during a trading day. The 'true' high (low) of the underlying continuous time price process is unlikely to be recorded as a transaction and the observable high (low) price underestimates (overestimates) the unrecorded 'true' high (low). This results in the downward bias in the daily range that carries over to ranges sampled at higher frequencies. The realized volatility estimator is not affected by non-trading, for the realized range, however, bid-ask bounce and non-trading partially offset each other. For the possibly remaining net impact of noise Martens and Van Dijk (2007) propose a multiplicative bias-correction.

The difference between the realized range estimators that appeared in Martens and Van Dijk (2007) and Christensen and Podolskij (2007) is that Martens and Van Dijk (2007) follow Parkinson (1980) by assuming that there is a sufficiently large number of observed transactions in an interval so that the (sum of) squared range(s) is adequately scaled by the variance of the range of a Brownian motion. Christensen and Podolskij (2007), in contrast, propose to determine the scaling parameter by estimating the variance of the range of a discretely observed Brownian motion. Hence, the Christensen and Podolskij (2007) estimator is not downward biased due to non-trading and does not have the advantage of partially offsetting the upward bias due to bid-ask bounce. For this reason Christensen et al. (2009) propose a bias-correction that mitigates the impact of bid-ask bounce.

The proposed biased correction in Christensen et al. (2009) boils down to estimating the bid-ask spread and subtracting it from each intraday range. This methodology works under the implicit assumption that the high (low) is always observed as a transaction executed at the ask (bid) quote. In practice, however, the high (low) price is not always recorded at the ask (bid) quote. Although it is most likely that an intraday range is upward biased, it can also be unbiased when both the high and low price in an interval are recorded at the same type of quote (both are bids or asks). It is also possible that an intraday range is downward biased. Although unlikely, the probability of observing a high price at a bid quote and observing the low price at an ask quote in a specific interval is non-zero. The non-zero probability is important when the intraday sampling frequency is high since this results in a large number of intraday ranges.

In Chapter 2 we relax the implicit assumption in Christensen et al. (2009) of the high (low) price always being a transaction executed at the ask (bid) quote. This results in a novel heuristic approach for correcting the realized range based on simulations, sorting and estimating the probabilities of observing upward, no or downward bias due to bid-ask bounce in a particular intraday interval. This approach results in a further increase in statistical efficiency which is achieved by a further decrease in bias at the cost of a negligible increase in variance. We illustrate the merits of the proposed bias-adjustment for volatility estimation using stochastic volatility simulations and an empirical volatility forecasting application.

1.5 Realized Covariance and Realized Co-Range

The concept of realized volatility is extended to the multivariate case by Andersen et al. (2001), Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004). Similar to the univariate case, realized covariance and realized correlation also benefit from the reduced variance of these estimators. The impacts of market microstructure noise, however, are different. Whereas bid-ask bounce is a dominant source of upward bias for realized volatility estimates, the impact on realized covariance estimates is minor.[3] This is not to say that the impacts of microstructure noise are less complicated in a multivariate setting. Realized covariance estimates based

[3] The impact of bid-ask bounce on microstructure noise partially depends on whether bid- and ask-quoted transactions occur independently across assets or not.

on high-frequency data can be severely biased towards zero since assets trade non-synchronously.

Brandt and Diebold (2006) extend the daily range-based volatility estimator to the multivariate case by forming a portfolio of two assets and using the daily ranges of the portfolio and the two individual assets to estimate the daily covariance; see also Brunetti and Lildholdt (2007). This daily co-range estimator has attractive properties such as the relatively low variance of the co-range estimator.

In Chapter 3 we combine the suggestion of Brandt and Diebold (2006) with the realized range estimator proposed in Martens and Van Dijk (2007) and Christensen and Podolskij (2007) in order to use intraday ranges for covariance estimation resulting in the novel realized co-range estimator. Realized covariances are unaffected by bid-ask bounce under the assumption that bid and ask transactions occur independently across assets. The realized co-range, however, is based on realized range estimates which are impacted by bid-ask bounce. In contrast to the realized covariance, the realized co-range is upward biased due to bid-ask bounce. This source of noise in high-frequency range-based covariance estimates can partially offset the bias towards zero induced by non-trading. Using extensive simulations that incorporate market microstructure noise and an empirical asset allocation application to stocks, bonds and gold futures, we illustrate that the realized co-range has attractive properties as a novel alternative asset return covariance estimator based on high-frequency data.

1.6 Factor Models and High-Frequency Data

For the practically relevant case of portfolios consisting of a large number of assets, several studies have successfully considered factor structures to tackle the 'curse of dimensionality' (see e.g. Chamberlain and Rothschild (1983), Chan et al. (1999) and Jagannathan and Ma (2003)). Recently Fan et al. (2008) illustrate the merits of a linear factor structure using low-frequency daily data. An important problem faced in the high-frequency literature on multivariate volatility estimation is the limitation in asset universe dimension. To guarantee a positive semi-definite covariance matrix most authors have initially avoided applications beyond the bi-variate case (see e.g. Hayashi and Yoshida (2005)) or have used small dimensional applications that cannot work for hundreds of assets. See e.g. Barndorff-Nielsen et al. (2011)) who propose multivariate realized kernels that are implemented based on refresh-time sampling. Refresh-time sampling in the context of realized covariances first appeared in Martens (2006). Although the idea may be useful for small dimensional applications it is suboptimal for vast dimensional applications where substantial differences in trading activity exist across individual securities. In general refresh time sampling leads to the least liquid asset(s) determining the effective sampling frequency. This resulting sampling frequency is much lower than could be achieved for the more liquid assets under consideration. Lowering of the effective sampling frequency in turn reduces the number of assets for which a positive semi-definite covariance matrix can be estimated. Given this problem, which is inherent to refresh-time sampling, Hautsch et al. (2012) propose to improve upon simple refresh-time sampling by first grouping the assets based on their liquidity and estimating covariance matrix series for the individual groups to reduce the loss of data.

In Chapter 4 we propose to combine the insights from the Finance literature on factor models with the increased precision rendered by using high-frequency data. The 'curse of dimensionality' for a long period limited the practical applicability of multivariate volatility estimation with high-frequency data. We introduce a novel approach for accurate measurement and forecasting of the covariance matrix of vast dimensional stock portfolios by combining the use of high- and low-frequency data with a linear factor structure. The factor covariance matrix is estimated using highly liquid exchange traded funds (ETFs) as observable factors. These ETF instruments are essentially free of market microstructure noise which justifies the use of ultra high-frequency data. For the factor loadings, however, we propose the use of daily low-frequency data to circumvent the potentially severe impacts of market microstructure noise for illiquid stocks. Using simulations we illustrate that in a bi-variate setting the realized mixed-frequency factor model (MFFM) compares favorably to (lead-lag adjusted) realized covariance estimators and the 'all overlapping returns' estimator proposed in Hayashi and Yoshida (2005). In an empirical setting the MFFM is successfully implemented by estimating and forecasting the asset return covariance matrix for hundreds of stocks belonging to the S&P500, S&P400 and S&P600 universes.

Chapter 2

Measuring and Forecasting Volatility with the Realized Range in the Presence of Noise and Non-Trading*

2.1 Introduction

Measuring and forecasting the volatility of asset returns plays a key role in various areas of financial economics, including portfolio management, risk management and the pricing of derivatives. The increasing availability of high-frequency asset price data has triggered a vast amount of academic studies proposing volatility estimators that exploit intraday prices to estimate and forecast daily volatility measures.

The realized variance (RV) estimator sums squared non-overlapping intraday returns to estimate the daily variance, see e.g. Andersen et al. (2001). In a frictionless market with continuous trading, RV converges to the integrated variance (IV) as the sampling frequency of the intraday returns increases. In practice, however, high-frequency asset prices are contaminated with market microstructure noise. This causes potentially severe problems in terms of consistent estimation of the daily IV by means of realized measures, see McAleer and Medeiros (2008) for a review. For RV estimators based on intraday returns obtained from transaction prices the dominant source of market microstructure noise is bid-ask bounce. Transactions take place at

*This chapter is based on Bannouh, Martens and Van Dijk (2012).

bid and ask prices causing an upward bias in the RV estimator. The magnitude of the bias increases with the sampling frequency.

A pragmatic solution to circumvent the problems arising from bid-ask bounce is to sample returns more sparsely by using longer intraday intervals; examples include the popular 5- and 30-minute frequencies. While lowering the sampling frequency reduces the bias in RV estimators, it also increases the variance. The use of sparse sampling frequencies aims to strike a balance between these two aspects. More formal approaches to correct for the effects of bid-ask bounce and other types of microstructure noise also exist. Among the most popular bias-correction methods is the two time scales RV (TSRV) estimator of Zhang et al. (2005). In this approach the microstructure noise component is consistently estimated using the highest sampling frequency available and subtracted from a subsampled RV estimator that is estimated using a 'sparse' sampling frequency.

Martens and Van Dijk (2007) and Christensen and Podolskij (2007) propose the realized range (RR) estimator as a more efficient measure of ex-post volatility. The RR estimator replaces the squared intraday returns in the RV estimator by squared intraday ranges. The results of Martens and Van Dijk (2007) illustrate that in a frictionless market the RR estimator is indeed more efficient than the RV estimator when comparing similar sampling frequencies. These results continue to hold in settings where market microstructure noise, in particular bid-ask bounce, is present.

The use of intraday ranges for volatility measurement is further complicated by a different source of market microstructure noise, namely infrequent trading. Trading does not occur continuously, that is, in practice we observe transactions at irregularly spaced points in time, see e.g. Engle (2000) or Griffin and Oomen (2008). For the RV estimator, non-trading increases the variance but does not cause a bias. In contrast, infrequent trading introduces a downward bias in RR estimators as the observed intraday high and low prices are likely to be below and above their 'true' values, respectively.[1] Christensen and Podolskij (2007) propose an adjustment of the standard RR estimator to account for the effects of non-trading.

Returning to the issue of bid-ask bounce, Christensen et al. (2009) propose a 'two time scales' RR (TSRR) estimator that aims to correct the upward bias due to bid-ask bounce along the same lines as the TSRV estimator of Zhang et al. (2005). The two time scales RR is implemented by estimating the bid-ask spread using the highest sampling frequency available and subtracting this quantity from each of the intraday ranges.

[1]Note that a possible advantage of the 'standard' realized range estimator is that the positive bias due to bid-ask bounce and the downward bias due to non-trading offset each other to a certain extent.

In this paper we extend the bias-adjustment for the realized range presented in Christensen et al. (2009) by relaxing the assumption that the observed high (low) price in each intraday interval originates from a transaction taking place at the ask (bid) quote. While this may be the most likely situation, in practice the high (low) price may also be observed as a transaction at the bid (ask) quote, such that an intraday range is not necessarily upward biased. Intuitively, the likelihood of an intraday range being upward biased decreases when the noise-to-volatility ratio becomes smaller or when the trading intensity of the asset becomes lower. We propose a heuristic adjustment of the RR that utilizes simulation-based estimates of the probabilities of an intraday range being upward biased, downward biased or unbiased. For the heuristic adjustment we need three inputs that are readily available from a sample path of tick data for a full trading day for which one wants to estimate the daily volatility. These inputs are estimates of the following quantities: (*i*) the daily range that is unaffected by noise, (*ii*) the non-trading probability and (*iii*) the half-spread. Using these inputs we simulate a geometric Brownian motion with variance (*i*) and implement noise with settings (*ii*) and (*iii*). For the simulated geometric Brownian motions we keep count of how many intraday ranges are upward biased, unbiased or downward biased. By averaging over simulation runs we estimate probabilities for the three cases that can be attached to the ranks of sorted intraday ranges. We apply these probability ranks to the sorted vector of initial high-low ranges for which we are now able to indicate whether an intraday range is expected to be upward biased, unbiased or downward biased.

We study the proposed heuristic bias-adjustment for the realized range estimator in a simulation setting with plausible levels of bid-ask bounce and non-trading. Using Monte Carlo simulations with several different stochastic volatility models as data generating process we find that the heuristically adjusted realized range estimator TSRRh provides volatility estimates that compare favorably, in terms of statistical efficiency, with the (TS)RV and (TS)RR estimators studied in Christensen et al. (2009) and the (TS)RV estimators in Aït-Sahalia and Mancini (2008). In an empirical forecasting application for the relatively liquid IBM stock and Zimmer Holdings (ZMH), a relatively illiquid constituent of the S&P500 belonging to the health care sector, we also find encouraging results. For IBM the heuristically adjusted RR volatility estimator provides more efficient one-step ahead forecasts. For ZMH the TSRRh outperforms (TS)RV and TSRR and competes with the RR estimator.

Our paper is related to several recent articles examining the relative performance of different realized measures in terms of measuring and forecasting the daily integrated variance. Among the studies that focus on out-of-sample predictive ability, Liu et al. (2012) recently consider the model confidence set approach to test for

350 assets, selected from several asset classes, whether alternative volatility forecasts can beat RV forecasts. They conclude that there are better forecasts but that it is difficult to significantly improve upon the RV forecasts. Their study includes the realized range which is implemented in the form proposed by Christensen and Podolskij (2007), which takes non-trading into account but is not unadjusted for other forms of microstructure noise. They find that the realized range forecasts compare favorably, especially for interest rate futures. Aït-Sahalia and Mancini (2008) put forward forecasting results for TSRV and RV measures in the presence of jumps, noise correlated with the efficient price, autocorrelated noise, long-memory in volatility and leverage effects in volatility. In addition they compare TSRV and RV forecasts for the relatively liquid DJIA stocks. They find that TSRV forecasts are more efficient than RV forecasts. Andersen et al. (2011) evaluate out-of-sample volatility forecasts in a simulation setting that uses stochastic volatility diffusions. The resulting efficient price processes are contaminated with microstructure noise. Their analysis is extended in several dimensions such as an implementation where the noise is serially correlated. They find that a combination of the TSRV and a RV estimator constructed by weighting different sampling frequencies performs best. Ghysels and Sinko (2011) evaluate volatility forecasts in the Mixed Data Sampling (MIDAS) framework and include results for iid-distributed noise and dependent noise. Consistent with Aït-Sahalia and Mancini (2008) they find that at high sampling frequencies TSRV forecasts achieve the highest efficiency. Christensen et al. (2009) compare (TS)RV and (TS)RR estimators and find that in the presence of bid-ask bounce TSRR and TSRV compete in terms of statistical efficiency and that TSRR is more efficient when more than 300 observations are available. In an empirical application Christensen et al. (2009) estimate the volatility of two highly liquid IT stocks, Microsoft and INTEL, and find that (TS)RV estimators have a smaller variance than RR. The TSRR they propose, however, has a smaller variance than the (TS)RV estimators.

The remainder of this paper is structured as follows. In Section 2 we develop the heuristic bias-adjustment for the RR estimator and discuss the (two time scales) realized volatility and (two time scales) realized range estimators. The simulation results are discussed in Section 3. Empirical forecasting results are presented in Section 4. We conclude in Section 5.

2.2 Volatility estimators, noise and bias-corrections

2.2.1 Volatility estimators

We assume that the logarithmic asset price P_t follows a driftless diffusion

$$dP_t = \sigma_t dW_t, \tag{2.1}$$

where σ is a strictly positive stochastic volatility process and W_t is a Wiener process. The daily interval is standardized to unity, such that the daily integrated variance (IV) is given by

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds. \tag{2.2}$$

Let $r_{t,j}^{\Delta} = \log P_{t+j\Delta} - \log P_{t+(j-1)\Delta}$ denote the log-return over the j-th intra-day interval of length Δ on day t, for a given interval length $0 < \Delta < 1$ such that we have $J = 1/\Delta$ intervals in a given day.[2] The realized variance estimator is calculated by summing squared intraday returns that are sampled from non-overlapping intervals of length Δ,

$$RV_t^{\Delta} = \sum_{j=1}^{J} r_{t,j}^2. \tag{2.3}$$

The realized range replaces the squared returns in RV by squared intraday ranges,

$$RR_t^{\Delta} = \frac{1}{4\log 2} \sum_{j=1}^{J} (\log H_{t,j} - \log L_{t,j})^2, \tag{2.4}$$

where $H_{t,j} = \sup_{(j-1)\Delta \leq i \leq j\Delta} P_{t+i}$ and $L_{t,j} = \inf_{(j-1)\Delta \leq i \leq j\Delta} P_{t+i}$ denote the high and low prices during the j-th interval on day t. In a frictionless market environment with continuous trading, both RV_t and RR_t are consistent estimates of the integrated variance IV_t when the number of intraday intervals $J \to \infty$. In the constant volatility case $\sigma_t = \sigma$ the variance of RV is $2\sigma^4\Delta^2$ and the variance of RR is approximately[3] $0.407\sigma^4\Delta^2$, which renders the RR about 5 times more efficient.

2.2.2 Market microstructure noise

Market microstructure noise refers to imperfections in the trading process of financial assets causing observed prices to deviate from the underlying 'true' price pro-

[2] For convenience we assume that Δ is such that J is an integer.

[3] The exact variance of the RR is $(\frac{9\zeta(3)}{(4\log 2)^2} - 1)\sigma^4\Delta^2$ where $\zeta(x) = \sum_{m=1}^{\infty} 1/m^x$ is Riemann's zeta function.

cess. Microstructure noise generally implies that realized volatility and realized range measures are inconsistent estimators for the integrated variance, with the impact becoming more pronounced as the sampling frequency increases. We focus on bid-ask bounce and non-trading since these are the two most relevant sources of noise that affect range-based volatility estimates based on high-frequency intra-day transaction prices.

Bid-ask bounce

Observed transactions take place at bid and ask quotes causing negative autocorrelation in high-frequency returns as the observed price jumps transiently from ask to bid and vice versa, see e.g. Roll (1984). Hence, at the micro level bid-ask bounce introduces volatility in the observed price process that is unrelated to the volatility of the 'true' price process. For this reason bid-ask bounce causes an upward bias in high-frequency volatility estimates.

A general representation of bid-ask bounce and the relationship between the 'efficient' price P_t and the 'noisy' transaction price P_t^* is given by:

$$P_t^* = P_t + \omega_t, \tag{2.5}$$

where bid-ask bounce is represented by ω_t which follows an i.i.d. distribution with support on $+\omega$ and $-\omega$, such that ω represents the half-spread.

Infrequent trading

Strictly speaking, non-trading does not fall under the heading of microstructure noise as defined above, in the sense that observed transaction prices are (or can be) equal to the efficient price. As the price process is not observed continuously though, non-trading does affect the RR estimator. As the observed high and low prices in a given intra-day interval are likely to be below and above their 'true' values, respectively, infrequent trading introduces a downward bias in the 'standard' RR estimator in (2.4). Effectively, in the presence of non-trading the scaling parameter $4 \log 2$, which is the variance of a continuously observed Brownian motion, is not appropriate. Following Christensen and Podolskij (2007), we therefore use

$$RR_t^{\Delta} = \frac{1}{\lambda_m} \sum_{j=1}^{J} (\log H_{t,j} - \log L_{t,j})^2, \tag{2.6}$$

where m is the number of observations in an intraday range. The appropriate scaling parameter $\lambda_m = \mathsf{E}[\max_{0 \leq s,t \leq m} (W_{t/m} - W_{s/m})^2]$ is determined through simulating an infrequently observed Brownian motion W and estimating the second moment of its range. Note that this adjustment destroys the possibility that the upward bias due to bid-ask bounce and the downward bias due to infrequent trading (partly) offset each other, necessitating a further adjustment of the RR in Equation (2.6) to account for the effects of microstructure noise.

2.2.3 Correcting for bid-ask bounce

Subsampling aims at at improving the accuracy of realized measures by using multiple intraday sample paths through shifting the point at which a sample starts. Assuming one has access to 1-minute price observations at 9:30, 9:31, 9:32, etc. the standard approach to estimate RV using, for example, 5-minute returns is to use transaction prices at 9:30, 9:35, 9:40 etc. A way to exploit more of the available data is to use a 5-minute price sample consisting of observations 9:31, 9:36, 9:41 etc. This approach provides five different samples giving rise to five different RV estimates. These can be averaged such that more data is used. The number of subsamples one can compute depends on the 'intended' sampling frequency and on the highest sampling frequency available. Assuming that there are S subsamples, the subsampled RV^S estimator is defined as:

$$RV_t^{\Delta,S} = \frac{1}{S}\sum_{s=1}^{S} RV_{t,s}^{\Delta}. \tag{2.7}$$

The two time scales estimator introduced in Zhang et al. (2005) combines the subsampled RV^S estimator at a 'sparse' frequency, e.g. 5-minutes ($n = 78$), with an ultra-high-frequency estimator that uses all of the N observed transaction-based intraday returns to estimate the noise component. At the ultra-high-frequency RV is estimated using all of the $N+1$ observed price ticks in a trading day and is denoted RV^N. This 'all returns' estimator produces a consistent estimate of the quantity $2N\mathsf{E}(\omega^2)$ such that $\mathsf{E}(\omega^2) = RV^N/2N$. Combining the sparsely subsampled RV^S estimator and the 'all returns' estimate to remove the noise results in a consistent estimator of the integrated variance, the two-time-scales realized variance (TSRV) estimator:

$$TSRV_t^{\Delta} = RV_t^{\Delta,S} - \frac{\bar{n}}{N}RV^N, \tag{2.8}$$

where $\bar{n} = \frac{n-S+1}{S}$. A small sample adjustment is applied to adjust for the fact that the number of returns in each of the sub-grids may not be equal:

$$TSRV_t^{\Delta,adj} = \frac{1}{1-\frac{\bar{n}}{N}}TSRV_t^{\Delta}. \tag{2.9}$$

For sufficiently large samples the correction term converges to unity. The TSRV estimator uses all available intraday price observations to estimate the noise component. For the RV subsampler at sparse frequencies, however, TSRV does not necessarily use all of the available data. Range-based volatility estimators by construction use all of the available data to calculate the highs and lows in an interval, and hence, make more efficient use of the high-frequency data to estimate volatility.

Similar to the TSRV estimator, Christensen et al. (2009) propose the use of a bias-correction for the realized range estimator based on two time scales. The bias-correction is derived under the assumption that the noise is represented by bid-ask bounce[4], i.e. an iid-noise distribution centered around zero with support on only two points, see also Equation (2.5) for a general representation. The highest frequency time scale is used to estimate the impact of bid-ask bounce. Specifically, a consistent estimate of the half-spread is obtained using $\hat{\omega} = \sqrt{RV^N/2N}$. This quantity is then used to filter out the bid-ask spread ω in each interval of the sparsely sampled realized range estimator:

$$TSRR_t^{\Delta} = \frac{1}{\tilde{\lambda}_m} \sum_{j=1}^{J} (\log H_{t,j} - \log L_{t,j} - \gamma\hat{\omega})^2, \tag{2.10}$$

where Christensen et al. (2009) use $\gamma = 2$ which is based on the implicit assumption that $H_{t,j}$ is always at the ask-quote and $L_{t,j}$ is always at the bid-quote. The scaling parameter $\tilde{\lambda}_m = \mathsf{E}[\max_{s:\omega_{s/m}=-\omega, t:\omega_{t/m}=\omega}(W_{t/m} - W_{s/m})^2]$ is determined through estimating the variance of the range of a discretely observed Brownian motion that is contaminated with noise.

The TSRR proposed in Christensen et al. (2009) takes into account that observed prices are contaminated by bid-ask bounce and that prices are observed infrequently. The latter is done through the multiplicative scaling parameters λ_m and $\tilde{\lambda}_m$ which take on different values for RR and TSRR due to microstructure noise. Underlying the additive part of the bid-ask correction where $\gamma = 2$, is the implicit assumption that the high is always an ask price and the low is always a bid price. In the presence of plausible levels of bid-ask bounce and non-trading, however, the probabilities of an intraday range being unbiased or downward biased are non-zero. The assumption of all intraday ranges being upward biased only holds when an asset trades very frequently throughout the day and a sufficiently large number of transactions is recorded in each of the intraday sampling intervals. In addition, the noise-to-volatility ratio

[4]It is hard, if not impossible, to derive a bias-adjustment for the RR estimator under noise distributions with unlimited support. Christensen et al. (2009) provide extensions to other microstructure noise distributions with bounded support such as a uniform noise distribution and rounding errors. The focus in their study, however, is also mainly on bid-ask bounce.

should be sufficiently large. For illiquid assets such as stocks that are traded infrequently this assumption may not always hold. This can be exemplified by analyzing an artificial price path where in some specific intraday interval the high and low are equal, i.e. this interval should not contribute to the daily volatility. For the RV and RR estimators this is the case, as both the intraday return and range are zero for this interval and do not contribute to the daily volatility estimates. This specific interval will, however, introduce an upward bias in TSRR of $\frac{4\omega^2}{\bar{\lambda}}$. This upward bias for a specific interval also occurs when the high and low are non-equal but both were recorded at the bid quote (ask quote). For these reasons we relax the assumption that the observed high (low) price always originates from a transaction executed at the ask (bid) quote. Specifically, we use simulation-based estimates of the probabilities that a specific intraday range is unbiased or even downward biased. The underlying idea is that if one would sort all the observed intraday highs (lows), then the highest high (lowest low) is more likely to be at the ask-quote (bid-quote) than is the case for the lowest (highest) observed high (low).

In more detail, we propose the following bias adjustment procedure that is based on simulation and sorting. Given a trading day of tick data that is contaminated by noise and infrequent trading:

1. Estimate the non-trading probability using the number of observed transactions on day t.
2. Use Parkinson (1980)'s daily high-low range estimator to obtain an initial estimate of the volatility for day t.[5]
3. Estimate bid-ask bounce, i.e. $\hat{\omega} = \sqrt{\mathsf{E}(\omega^2)} = \sqrt{RV^N/2N}$.
4. Simulate intraday sample-paths based on a geometric Brownian motion with inputs being the estimated non-trading probability, the initial volatility estimate and the estimated bid-ask spread.
5. Using the bid-ask and non-trading contaminated simulated sample paths, estimate the probability of observing (a) no bias, (b) upward bias and (c) downward bias in the intraday range.[6]
6. Sort the empirical intraday high-low's. Based on the estimated probabilities from the previous step, calculate how many of the intraday ranges are expected to be (a) unbiased, (b) upward biased or (c) downward biased. Use Equation (2.11) and apply (a) $\gamma_j = 0$, (b) $\gamma_j = 2$ and (c) $\gamma_j = -2$ to adjust for (b) upward bias and (c) downward bias.

[5] It is important that this estimator is (almost) not affected by microstructure noise (we will use the daily range, alternatively one can use another (almost) bias-free measure, e.g. the TSRV or the daily squared return).

[6] Case (a) occurs when in an intra-day interval the observed high and low are both executed at a bid price (or both being an ask), (b) occurs when the observed high is an ask-price and the observed low is a bid-price (c) occurs when the high is a bid-price and the low is an ask price.

Hence, our estimator has the same form as the estimator proposed in Christensen and Podolskij (2007) with the difference being that we do not use $\gamma = 2$ to correct each of the intra-day ranges. Instead we propose to use

$$TSRRh_t^{\Delta} = \frac{1}{\tilde{\lambda}} \sum_{j=1}^{J} (\log H_{t,j} - \log L_{t,j} - \gamma_j \hat{\omega})^2, \qquad (2.11)$$

where we use $\gamma_j = 2$ if after sorting and using the simulated probabilities an intraday range is expected to be biased upward (b). Assuming that the J intraday ranges are sorted in a descending manner and the estimated probability of intraday ranges being biased upward is q, then the first Jq intraday ranges are expected to be biased upward. Similarly, assuming that the probability of an intraday being unbiased is estimated to be v, we use $\gamma_j = 0$ (a) for the subsequent Jv intraday ranges and for the remaining[7] $J(1 - q - v)$ intraday ranges $\gamma_j = -2$ (c) is used.[8]

2.3 Monte Carlo Simulation

In the following Monte Carlo simulation experiments we compare ex-post volatility estimates using the (TS)RV and (TS)RR estimators with the newly proposed TSRRh estimator. The estimators are compared in terms of bias, variance and efficiency. We simulate the integrated variance using several stochastic volatility diffusions that were also used in Aït-Sahalia and Mancini (2008), among others. Returns and integrated volatilities are simulated from a Heston Jump-Diffusion, a Fractional Ornstein-Uhlenbeck process and a discrete-time log-volatility model. We simulate 1,000 trading days of 6.5 hours, i.e. 23,401 prices are simulated per day to match a time step of 1 second. Subsequently non-trading is implemented by assuming a trade is observed with probability 0.10 such that on average 2,340 'clean' prices are observed during the day. Microstructure noise is implemented by contaminating the prices with a half-spread of $\omega = 0.025\%$ on the asset price. Bid and ask prices are assumed to occur equally likely. In all experiments we use 100 sub-sample grids to calculate TSRV. For each daily TSRRh estimate 500 simulations are used to estimate

[7] Assuming Jq and Jv are integer.

[8] For the statistical properties of the TSRR(h) estimator we refer to Christensen et al. (2009) who obtain consistency using double asymptotics for the number of intervals and the number of observations per interval. Asymptotically the TSRR and TSRRh estimators share the same statistical properties when microstructure noise is represented by bid-ask bounce as in Equation (2.5) and the number of observations per interval diverges to infinity. The probability of observing upward bias in an intraday range converges to 1. For the TSRRh estimator this results in $\gamma = 2$ and therefore it becomes equivalent to the TSRR estimator. The small sample properties of these volatility estimators under several different stochastic volatility diffusions as data generating process are studied in the following section.

the impact of bid-ask bounce for rank-sorted intraday ranges in order to implement the proposed bias-adjustment as in Equation (2.11).

2.3.1 Heston stochastic volatility jump-diffusion

The data generating process for returns and volatility under the Heston (1993) stochastic volatility jump-diffusion model is specified by

$$\begin{aligned} dP_t &= (\mu - \sigma_t^2/2)dt + \sigma_t dW_{1,t}, \\ d\sigma_t^2 &= -\kappa(\sigma_t^2 - \alpha)dt + \gamma\sigma_t dW_{2,t} + J_t dq_t, \end{aligned}$$

with drift parameter $\mu = 5\%$, a long term average volatility $\alpha = 3.5\%$, and mean reversion parameter $\kappa = 5$. The volatility of volatility parameter $\gamma = 0.5$ facilitates leverage effects as the two Brownian motions are negatively correlated with $\rho = -0.5$. The occurrence of jumps in the volatility process has distribution $q_t \sim Poi(\phi)$ and the jump magnitude follows an exponential distribution $J_t \sim Exp(\zeta)$. Following Aït-Sahalia and Mancini (2008) we set $\lambda = 1/2$ and $\zeta = 0.0007$. Empirical stylized facts are taken into account by the inclusion of jumps in the volatility process and a leverage effect to allow for the empirically plausible negative relation between returns and volatility shifts.

2.3.2 Fractional Ornstein-Uhlenbeck process

Following Aït-Sahalia and Mancini (2008) we simulate IV using a fractional Brownian motion,

$$\begin{aligned} dP_t &= (\mu - \sigma_t^2/2)dt + \sigma_t dW_t, \\ d\sigma &= -\kappa(\sigma_t - \alpha)dt + \gamma dW_{H,t}, \end{aligned}$$

where dW_t is a Wiener process and $dW_{H,t}$ is a fractional Brownian motion with Hurst index $H \in (0,1)$. A fractional Brownian motion is a continuous mean zero Gaussian process with stationary increments and covariance $E(W_{H,t}W_{H,s}) = \frac{1}{2}(s^{2H} + t^{2H} - |s-t|^{2H})$. The covariance structure illustrates that the increments are positively correlated when $\frac{1}{2} < H < 1$ and exhibit long-memory, for $H = \frac{1}{2}$ the increments are independent and correspond to a standard Brownian motion. To simulate the fractional Brownian motion we use the Davies and Harte (1987) algorithm with Hurst effect $H = 0.7$.

2.3.3 Discrete-time log-volatility model

In many applications the logarithm of volatility is used because the logarithm of (realized) volatility is empirically found to be closer to a Gaussian distribution (see e.g. Figure 1 in Andersen et al. (2001)). The discrete time model we use is the model employed in Andersen et al. (2003) and Aït-Sahalia and Mancini (2008). The daily integrated volatility l_t follows an AR(5) process

$$l_t = \frac{1}{2}\log(IV_t) = \phi_0 + \sum_{i=1}^{5} \phi_i l_{t-id} + e_t, \tag{2.12}$$

where IV_t is the daily integrated variance and e_t is white noise. Intraday efficient returns are obtained using $r_t = \sqrt{IV_t} z_t$ with $z_t \sim NID(0,1)$. For the parameters we use those reported by Aït-Sahalia and Mancini (2008), $\phi_0 = -0.0161$, $\phi_1 = 0.35$, $\phi_2 = 0.25$, $\phi_3 = 0.20$, $\phi_4 = 0.10$, $\phi_5 = 0.09$ and $\sigma_e = 0.02$.

2.3.4 Monte Carlo results

Volatility estimation results using Monte Carlo simulations for the three stochastic volatility models[9] discussed above are summarized in Table 2.1. The microstructure noise settings used are a probability equal to 0.10 of observing a trade[10], which results in 2,340 observations per day on average and a half-spread of 0.025% of the asset price.

Under the Heston jump-diffusion the bias for the RV estimator (0.093) is somewhat smaller than would be expected based on using a half-spread of 0.025% of the asset price ($0.0975 = 2*390/5*0.025\%^2$).[11] This is due to the quadratic variation being larger because of jumps in the volatility process. The variance of all the volatility estimators considered under the Heston jump-diffusion models is considerably larger than in models that do not incorporate jumps since the volatility estimators discussed here are not designed to be jump-robust. Theoretically the RR estimator is expected to have a substantially smaller variance than the (TS)RV estimators. It is interesting to compare the competing estimators in the presence of bid-ask bounce, non-trading and jumps in the volatility process. Indeed we find that at the 5-minute

[9]Results for a Brownian motion with constant volatility are similar in the sense that TSRRh improves upon (TS)RV because of having a smaller variance leading to a smaller RMSE. The TSRRh also improves upon (TS)RR because of a smaller bias that comes at the cost of a modest increase in variance. This bias-variance trade-off results in TSRRh having a smaller RMSE than (TS)RR as well. Results are available upon request.

[10]The trading probability is in line with the results presented in Table 1 in Hansen and Lunde (2006).

[11]Errors are multiplied with 10^4 to improve readability.

sampling frequency the variance of RR (0.071) is still more than 3 times smaller than the variance of RV (0.253) and less than half the variance of TSRV (0.170). In terms of RMSE the RR (0.460) performs better than RV (0.511) but in turn it is outperformed by the TSRV (0.414) because the latter is approximately unbiased (−0.038). The bias of the RR (0.374) estimator is substantially larger than the bias in the RV estimator. Bias-correcting the realized range as proposed by Christensen et al. (2009) successfully reduces the bias from 0.374 to −0.263 at the cost of an increase in variance from 0.071 for RR to 0.108 for TSRR. Despite the reduced bias, the TSRR (0.421) still does not improve upon TSRV (0.414). Taking into account that not all intraday ranges are upward biased and that the largest intraday ranges in a day are more likely to be upward biased than the smallest intraday ranges is exemplified by TSRRh (−0.239) having a smaller bias than TSRR (−0.263). As a result the RMSE of TSRRh (0.407) is also smaller than the RMSE (0.414) of the unbiased TRSV estimator. At the 30-minute sampling frequency the impact of noise is substantially smaller as expected[12] and for this reason it is optimal to use the RR without bias-correction.

Across models we find that using 5-minute intervals to estimate daily volatility outperforms the lower 30-minute and daily sampling frequencies in terms of variance and statistical efficiency. Under the fractional Brownian motion model the TSRV estimator minimizes the bias (−0.023) at the 5-minute sampling frequency as was the case under the Heston model. Again the realized range-based estimators achieve a smaller variance than (TS)RV. However, it is also the most biased estimator and for this reason the least efficient with a RMSE of 0.422. The TSRV (−0.023) successfully reduces the bias of RV (0.098) and achieves a RMSE of 0.219. Similarly the TSRR is very successful in reducing the bias of RR (0.395) to −0.138 and also has a smaller RMSE (0.208) than the (TS)RV estimators. By using the informational content contained in the size of the intraday ranges through implementing the TSRRh the bias is further reduced from −0.138 for TSRR, down to −0.120 for TSRRh which results in TSRRh having the smallest RMSE (0.199).

For the discrete-time log-volatility model we find similar results in the sense that at the 5-minute sampling frequency the TSRV estimator minimizes the bias (−0.037) but has a variance (0.048) that is inferior to that of the RR (0.022), TSRR (0.024) and TSRRh (0.025) estimators. The TSRRh (−0.134) is less biased than the TSRR (−0.151) which in turn is less biased than RR (0.380). The result is that, similar to the results under the Heston Jump-Diffusion and the fractional Brownian motion model, the TSRRh at the 5-minute sampling frequency achieves the smallest RMSE in the discrete-time log-volatility model.

[12] For instance the expected RV bias is now only 0.01625.

Table 2.1: Monte Carlo Results for Volatility Estimators under Stochastic Volatility Diffusions

DGP	Heston Jump Diffusion			Fractional Browian Motion			Discrete Log-volatility		
Estimator	Bias	Variance	RMSE	Bias	Variance	RMSE	Bias	Variance	RMSE
RV-5min.	0.093	0.253	0.511	0.098	0.075	0.292	0.084	0.075	0.287
TSRV-5min.	-0.038	0.170	0.414	-0.023	0.047	0.219	-0.037	0.048	0.222
RR-5min.	0.374	0.071	0.460	0.395	0.022	0.422	0.380	0.022	0.408
TSRR-5min.	-0.263	0.108	0.421	-0.138	0.024	0.208	-0.151	0.024	0.217
TSRRh-5min.	-0.239	0.109	0.407	-0.120	0.025	0.199	-0.134	0.025	0.208
RV-30min.	0.011	1.430	1.196	0.017	0.400	0.632	0.000	0.413	0.642
TSRV-30min.	-0.202	0.875	0.957	-0.123	0.243	0.509	-0.139	0.249	0.518
RR-30min.	0.153	0.340	0.603	0.159	0.097	0.349	0.145	0.100	0.348
TSRR-30min.	-0.181	0.344	0.614	-0.100	0.091	0.318	-0.114	0.095	0.328
TSRRh-30min.	-0.181	0.343	0.613	-0.098	0.091	0.317	-0.113	0.095	0.327
RV-390min.	0.038	18.666	4.321	0.009	5.362	2.316	0.006	5.494	2.344
TSRV-390min.	0.035	18.697	4.324	0.007	5.371	2.317	0.004	5.502	2.346
RR-390min.	0.054	3.887	1.972	0.043	1.123	1.061	0.037	1.146	1.071
TSRR-390min.	-0.086	3.714	1.929	-0.056	1.061	1.032	-0.062	1.083	1.043
TSRRh-390min.	-0.097	3.662	1.916	-0.061	1.048	1.025	-0.068	1.070	1.037

Note: The table summarizes the Monte Carlo estimation results for (TS)RV, (TS)RR and TSRRh as ex-post volatility estimators. Volatility and returns are simulated from the Heston Jump-Diffusion, a fractional Brownian Motion and a discrete-time Log-Volatility Model. The bias, variance and RMSE statistics are based on 1,000 daily sample paths with 23,401 prices per day. For each of the 1,000 sample paths the calculation of the TSRRh volatility estimator is based on 500 Monte Carlo simulations that are used for estimating the probabilities of observing bid-ask noise in the intraday ranges. Non-trading and bid-ask bounce are implemented by setting the probability of observing a trade to 0.10 and using a half-spread of 0.025% on the asset price. Results are reported for sampling frequencies of 5, 30 and 390 minutes.

2.4 Empirical application

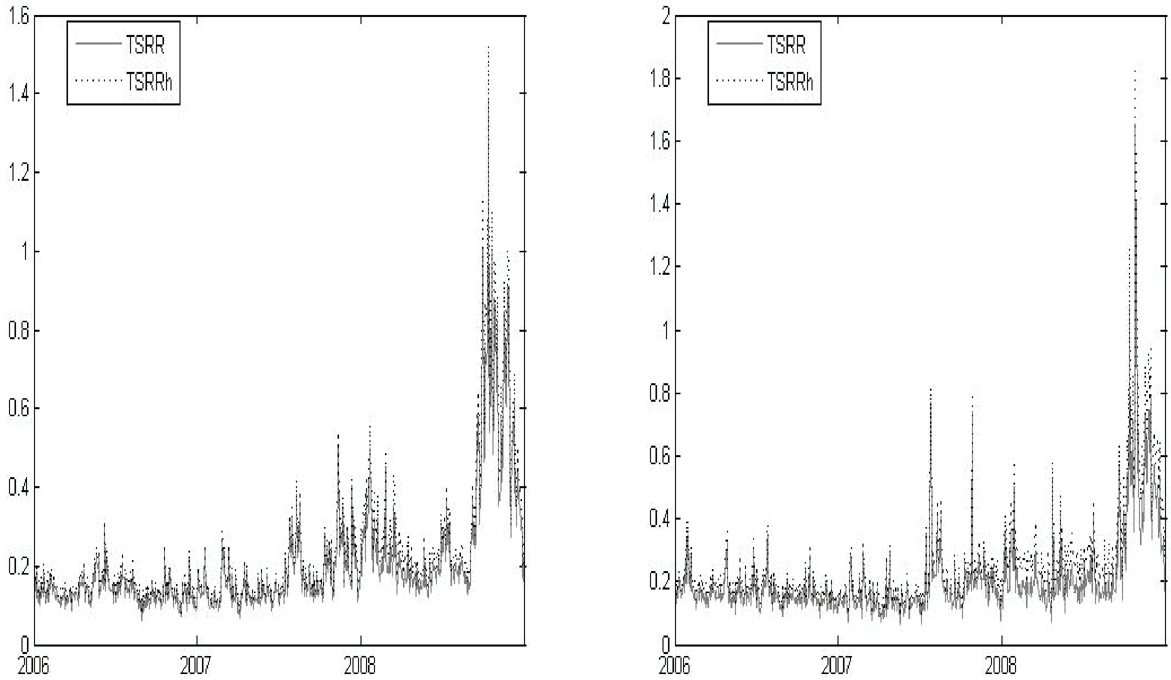
For a relatively liquid (IBM) and illiquid (Zimmer Holdings, ZMH) stock we obtain intraday transaction prices and quotes from the TAQ database for the 1/1/2006 – 12/31/2008 period. The data are cleaned following the procedures documented in Barndorff-Nielsen et al. (2009) with the exception that we do not use moving-average rules to judge the adequacy of observed transactions.[13] Using the cleaned data we estimate the bid-ask spreads, following Roll (1984), to be 2.13 basis points (bps) for IBM and 4.93 bps for ZMH. The daily and intra-daily variation in bid-ask spreads through our sample period is, however, quite substantial. This particularly applies to the financial market turmoil in 2008. The trading probabilities are estimated to be 0.084 for ZMH and 0.201 for IBM on a 1-second time-grid.[14]

Figure 2.1 plots annualized volatility estimates for IBM and ZMH. The sample period 2006–2008 is interesting since it contains the relatively tranquil period before the 2008 crisis, the height of the financial market turmoil in the second half of 2008 when stock market volatility peaked and the reversion towards normal volatility levels at the end of 2008. Consistent with the simulation results, the two time scales realized range and the heuristically adjusted two time scales realized range render very similar volatility dynamics. Since the heuristical adjustment does not always assume that an intraday range is upward biased, the empirical volatility estimates are slightly higher than the two time scales realized range volatility estimates. Note that these differences are more emphasized when volatility is relatively high (2008) and the stock is relatively illiquid (ZMH), as expected based on the discussion in Section 2.2.3 and the simulation results in Section 2.3.

We evaluate the out-of-sample forecasting performance of the heuristically bias-adjusted RR, (TS)RV and (TS)RR estimators. For each realized measure we use an AR(1) model (with intercept) to construct one day ahead volatility forecasts, using a rolling window of one year to estimate the AR(1) coefficients. The out-of-sample period is 1/1/2007–12/31/2008. We compare volatility forecasts using the commonly used 5-minute sampling frequency. This choice is motivated by the Monte-Carlo

[13] Transactions and quotes are cleaned as follows: 1: Delete observations not originating from the NYSE 2: Delete all implausible data, e.g. negative quotes/prices those equal to 0, 0.01 or e.g. 999.9., observations associated with a negative spread (ask<bid) etc. 3: Delete observations with sale condition other than "E"/"F". 4: Delete observations with time stamps outside the 9:30–16:00 hours. 5: Delete all corrected observations ($corr \neq 0$) 6: When multiple transaction prices have the same time stamp use the median, do the same for bid-quotes and ask-quotes. 7: Delete transactions that traded more than a spread size outside the bid-ask spread.

[14] For IBM the number of observed transactions before data cleaning procedures is substantially larger with 29,923 observations per day. We follow the convention to limit ourselves to the 1-second time grid, as described in the footnote above, we take the median of those transactions and this dramatically reduces the resulting number of transactions that are used to estimate the volatility.



(a) IBM 5-minute sampling

(b) ZMH 5-minute sampling

Figure 2.1: This figure illustrates annualized volatility estimates for (a) IBM and (b) Zimmer Holdings over the sample period 1/1/2006 – 12/31/2008 using 5-minute sampling frequencies for the two time scales realized range in Equation (2.10) **and the heuristically adjusted two time scales realized range in Equation** (2.11)**.**

results described in Section 2.3. We report Mincer-Zarnowitz and encompassing regression results to evaluate the predictive accuracy. In the Monte Carlo simulation we illustrated that for several stochastic volatility models the TSRRh is a highly efficient volatility estimator in the presence of bid-ask bounce and non-trading. Since for empirical data the integrated variance is unknown we compare the volatility forecasts using forecast comparison regressions rather than bias, variance and RMSE.

We run Mincer-Zarnowitz and encompassing regressions to evaluate the competing forecasts and following Aït-Sahalia and Mancini (2008) we use the two-time-scales realized variance TSRV as the ex-post volatility measure. Hence, the Mincer-Zarnowitz regressions are of the form

$$TSRV_t = \alpha + \beta x_{t|t-1} + \varepsilon_t, \tag{2.13}$$

where $x_{t|t-1}$ is the volatility forecast for day t conditional on the data available at day $t-1$.[15] In the encompassing regressions the realizations are regressed on two competing forecasts (being, e.g., the realized range and realized volatility forecast),

$$TSRV_t = \alpha + \beta_1 x_{1,t|t-1} + \beta_2 x_{2,t|t-1} + \varepsilon_t. \tag{2.14}$$

Equation (2.14) is a pair-wise encompassing forecast regression in the standard form.[16] For these regressions we report the coefficient estimates and their corresponding t-statistics based on Newey-West HAC robust standard errors (20 lags).

2.4.1 Empirical forecast results

Table 2.2 summarizes the Mincer-Zarnowitz regression results for volatility forecasts based on the (TS)RV, (TS)RR and TSRRh estimators. We find for both stocks that the differences in forecast accuracy are small due to the high correlation between volatility forecasts. For the relatively liquid IBM stock, we find that the realized variance forecasts have a Mincer-Zarnowitz R^2 of 49.6%. The two-time-scales realized volatility manages an R^2 of 50.8%. It slightly underperforms the unadjusted realized range forecasts which explain 50.9% of the variation in the ex-post TSRV estimates. This finding is quite remarkable, in the sense that the TSRV serves as proxy for the integrated variance in the Mincer-Zarnowitz regressions. Forecasts based on the bias-adjusted realized range proposed by Christensen et al. (2009) achieve an R^2

[15] The joint null hypothesis for x being unbiased and efficient is given by $H0: \alpha = 0$ and $\beta = 1$.

[16] The null hypothesis of forecasts x_1 encompassing forecasts x_2 is given by $H0: \beta_1 = 1 \bigcap \beta_2 = 0$ and the alternative hypothesis is $H1: \beta_1 \neq 1 \bigcup \beta_2 \neq 0$. Similarly the competing non-nested null hypothesis of forecasts x_2 encompassing forecasts x_1 is given by $H0: \beta_2 = 1 \bigcap \beta_1 = 0$ and the alternative hypothesis is $H1: \beta_2 \neq 1 \bigcup \beta_1 \neq 0$.

of 50%, hence the bias-adjusted realized range performs slightly worse compared to its unadjusted counterpart. Consistent with the volatility estimation results in the Monte Carlo simulations, the empirical forecasts based on the heuristically adjusted realized range outperform the forecasts based on other estimators as the TSRRh achieves an R^2 of 51.0%.

Table 2.2: Mincer-Zarnowitz Forecast Regressions

	RV	TSRV	RR	TSRR	TSRRh	RV	TSRV	RR	TSRR	TSRRh
	Panel A: IBM 5m					*Panel C: IBM 5m with outlier correction*				
α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
tstat	1.311	1.471	1.531	1.523	1.398	1.494	1.642	1.668	1.660	1.539
β	1.231	1.222	1.010	1.196	0.914	1.164	1.147	0.946	1.123	0.858
tstat	23.801	19.853	18.282	19.448	17.546	21.980	16.745	11.509	12.612	11.182
R^2	0.496	0.508	0.509	0.500	0.510	0.685	0.687	0.686	0.681	0.689
	Panel B: ZMH 5m					*Panel D: ZMH 5m with outlier correction*				
α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
tstat	-0.928	-0.879	-0.242	-1.022	-0.988	-1.076	-1.030	-0.372	-1.230	-1.150
β	1.736	1.820	1.562	2.054	1.250	1.691	1.779	1.522	2.003	1.220
tstat	9.124	7.295	10.921	9.995	8.812	10.019	7.740	12.244	11.040	9.517
R^2	0.331	0.328	0.335	0.328	0.333	0.433	0.432	0.437	0.430	0.436

Note: The table summarizes the results of Mincer-Zarnowitz forecast regressions with and without an outlier-correction applied to 10/10/2008. The (TS)RV, (TS)RR and TSRRh forecasts are generated using 5-minute sampling frequencies and a AR(1) process that is dynamically re-estimated using a moving window with window length 250 days. The imperfect volatility proxy used is the TSRV at the 5-minute sampling frequency.

For the relatively illiquid stock, Zimmer Holdings (ZMH), we find that the R^2's are substantially lower than for IBM volatility forecasts. Interestingly, the advantage of a bias-correction almost vanishes. This may be due to the fact that most corrections, in contrast to TSRR(h), are derived under continuous-time assumptions that do not hold for illiquid stocks. For example, the standard realized volatility has a Mincer-Zarnowitz R^2 of 33.1%, being somewhat higher than that of the TSRV (32.8%). Again we expected the latter to actually have a small advantage since it is the ex-post quantity used to evaluate the forecasts. Unreported simulation results indicate that TSRV does not outperform the standard RV estimator due to the noise estimate $RV^N/2N$ being inaccurate when N is small in practice, whereas in the theory outlined by Zhang et al. (2005) it is assumed that $N \to \infty$. When N is large we can assume that the volatility signal in RV^N is dwarfed by the noise signal. It is easy to see, however, that when N is small the volatility signal in RV^N increases. For this reason it causes a downward bias due to overcorrecting for noise.[17] The (TS)RV and TSRR forecasts are outperformed by the unadjusted realized range ($R^2 = 33.5\%$) and the novel heuristic adjustment ($R^2 = 33.3\%$). The bid-ask adjustment of Christensen et al. (2009) is at par with the two-time-scales estimator (32.8%). The heuristic bias-adjustment for the realized range (33.3%) outperforms (TS)RV and TSRR. Hence, for the relatively illiquid ZMH stock we find that bias-adjustments do not pay-off in terms of forecasting performance, it is in this case better to just use the RR volatility estimator without applying a bias-correction and if we insist on using intraday data, then the TSRRh is preferred based on its forecast regression R^2.

Table 2.3 summarizes the results for the encompassing forecast regressions in Equation (2.14). Here we compare directly with each other the AR(1) forecasts of the various volatility measures at the 5-minute frequency. We find that for IBM the forecasts obtained from the TSRV estimator encompass those from the unadjusted RV estimator, as expected based on the results in Aït-Sahalia and Mancini (2008). The coefficient on TSRV (1.570) in column 1 of panel A, is statistically significant ($t =$ 3.500) whereas the coefficient on RV is negative (-0.357) and statistically insignificant ($t =$ -0.846). Similarly, the results in column 2 of panel A indicate that the unadjusted realized range encompasses the unadjusted realized variance with coefficients being 1.090($t =$ 1.692) and -0.100($t =$ -0.134), respectively. Adding RR or TSRV forecasts to unadjusted RV forecasts results in the same R^2 of 50.9%. When we add the forecasts based on the TSRR estimator to RV forecasts (panel A, column 3) we find that the R^2 shrinks to 50.1% and both coefficients are statistically insignificant. However, adding the forecasts based on the heuristic bias-adjustment for realized range (TSRRh) to unadjusted RV forecasts actually improves the R^2 to 51.0% with its coefficient being 0.992($t =$ 1.709) and the coefficient on RV being -0.107($t =$ -0.145). In addition we report encompassing regression results for all other (bi-variate) forecast combinations and find that adding the unadjusted RR forecasts to the TSRV forecasts results in similar and statistically insignificant coefficients being 0.562($t =$ 0.733) and 0.546($t =$ 0.599), respectively, and an R^2 of 51.0%. Hence, combining RV and TSRRh forecasts results in the same R^2 as combining TSRV and RR. When we add the TSRR forecasts to TSRV forecasts we again find statistically insignificant coefficients being 1.633($t =$ 1.053) on TSRV and -0.408($t =$ -0.269) on TSRR. In contrast, we find that TSRV 6.143($t =$ 4.891) and TSRRh -6.139($t =$

[17] See e.g. also Zhang et al. (2005) or Aït-Sahalia and Mancini (2008) who report a very small negative bias in TSRV in a setting where 23,401 transactions per day are observed, if we move to more realistic settings and the number of observations decreases, this negative bias becomes more pronounced. Of course, using a lower sampling frequency for TSRV could reduce the impact of non-trading.

-3.857) compete, having statistically significant coefficients of similar absolute size but opposite signs, due to a high correlation between the forecasts. Running an encompassing regression for TSRR and TSRRh forecasts results in both forecasts being statistically significant and opposite signs with TSRR having a coefficient of -3.219(t = -2.332) and 3.346(t = 3.334) for TSRRh. Looking at the 10 possible forecast combinations the optimal combination found for the IBM data is that of RR and TSRR forecasts with an R^2 of 52.9%.

A similar analysis for the relatively illiquid stock (ZMH) illustrates that in contrast to the IBM results now RV 1.652(t = 1.561) forecasts outperform TSRV 0.090(t = 0.084) forecasts (column 1, panel B). The RV forecasts 0.254(t = 0.327), however, are outperformed by the RR 1.336(t = 2.321) forecasts (column 2, panel B), as expected. However, the RV forecasts 1.719(t = 2.007) almost reduce the coefficient on TSRR 0.021(t = 0.024) to zero (column 3, panel B). Hence, whereas the bias-adjustments worked for the relatively liquid IBM data this is not the case for the illiquid ZMH data. We find similar results when we add the TSRV or TSRR to RR forecasts, that is, the unadjusted RR forecasts are better than the TSRV and TSRR. In the direct competition between TSRV and TSRRh both coefficients are insignificant for IBM data with TSRRh having a somewhat larger weight of 0.554 than the weight of 0.486 on TSRV (column 7, panel A). For ZMH data (column 7, panel B), however, TSRRh forecasts encompass TSRV forecast with TSRRh having a statistically significant coefficient of 1.288(t = 2.358) whereas the coefficient on TSRV is negative -0.056 and insignificant. Similar results are found after applying an outlier correction to 10/10/2008 (column 7, panel D). For the out-of-sample period 2007–2008, the TSRRh forecasts are preferred over TSRR and TSRV forecasts.

Table 2.3: Empirical Encompassing Forecast Regressions

Panel A: IBM 5m

	RV+TSRV	RV+RR	RV+TSRR	RV+TSRRh	TSRV+RR	TSRV+TSRR	TSRV+TSRRh	RR+TSRR	RR+TSRRh	TSRR+TSRRh
α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
tstat	1.538	1.909	1.802	1.683	1.725	1.624	1.542	1.547	1.281	1.207
β_1	-0.357	-0.100	0.349	-0.107	0.546	1.633	0.486	6.143	-0.588	-3.219
tstat	-0.846	-0.134	0.368	-0.145	0.599	1.053	0.569	4.891	-0.260	-2.332
β_2	1.570	1.090	0.861	0.992	0.562	-0.408	0.554	-6.139	1.446	3.346
tstat	3.500	1.692	0.914	1.709	0.733	-0.269	0.842	-3.857	0.700	3.334
R^2	0.509	0.509	0.501	0.510	0.510	0.509	0.511	0.529	0.510	0.521

Panel B: ZMH 5m

	RV+TSRV	RV+RR	RV+TSRR	RV+TSRRh	TSRV+RR	TSRV+TSRR	TSRV+TSRRh	RR+TSRR	RR+TSRRh	TSRR+TSRRh
α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
tstat	-0.932	-0.347	-0.947	-1.009	-0.092	-1.034	-0.939	1.787	-0.496	-0.851
β_1	1.652	0.254	1.719	0.522	-0.259	0.910	-0.056	5.556	1.078	-0.634
tstat	1.561	0.327	2.007	0.622	-0.282	1.193	-0.073	7.458	1.059	-0.420
β_2	0.090	1.336	0.021	0.878	1.780	1.035	1.288	-5.319	0.391	1.630
tstat	0.084	2.321	0.024	1.524	2.424	1.351	2.358	-5.442	0.448	1.684
R^2	0.331	0.335	0.331	0.334	0.335	0.329	0.333	0.342	0.335	0.334

Panel C: IBM 5m with outier correction

	RV+TSRV	RV+RR	RV+TSRR	RV+TSRRh	TSRV+RR	TSRV+TSRR	TSRV+TSRRh	RR+TSRR	RR+TSRRh	TSRR+TSRRh
α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
tstat	1.709	2.067	1.954	1.861	1.878	1.868	1.675	1.661	1.434	1.358
I(20081010)	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
tstat	54.862	32.956	39.821	36.919	74.747	100.894	82.796	53.829	47.287	54.283
β_1	0.387	0.500	0.908	0.371	0.652	1.449	0.453	4.370	-2.669	-2.780
tstat	0.653	0.555	0.827	0.441	0.837	1.163	0.695	4.416	-0.992	-1.826
β_2	0.769	0.545	0.250	0.588	0.412	-0.300	0.522	-4.093	3.270	2.959
tstat	1.240	0.699	0.226	0.881	0.597	-0.238	0.973	-3.275	1.359	2.711
R^2	0.688	0.688	0.685	0.690	0.688	0.687	0.690	0.695	0.691	0.697

Panel D: ZMH 5m with outier correction

	RV+TSRV	RV+RR	RV+TSRR	RV+TSRRh	TSRV+RR	TSRV+TSRR	TSRV+TSRRh	RR+TSRR	RR+TSRRh	TSRR+TSRRh
α	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
tstat	-1.110	-0.426	-1.139	-1.156	-0.357	-1.186	-1.121	1.734	-0.713	-0.969
I(20081010)	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
tstat	72.364	83.896	72.779	76.846	111.750	81.674	75.200	101.525	100.563	79.768
β_1	1.160	0.177	1.514	0.245	0.059	1.147	0.085	5.165	0.815	-0.815
tstat	1.572	0.256	2.133	0.333	0.075	1.741	0.124	6.646	0.796	-0.541
β_2	0.563	1.365	0.213	1.046	1.472	0.719	1.162	-4.850	0.570	1.709
tstat	0.711	2.632	0.303	1.954	2.495	1.202	2.431	-4.676	0.643	1.756
R^2	0.434	0.437	0.433	0.437	0.437	0.432	0.436	0.443	0.437	0.437

Note: The table summarizes the results for the encompassing forecast regressions in equation (2.14). For the (TS)RV, (TS)RR and TSRRh volatility estimators based on the 5-minute frequency, forecasts are generated using an AR(1) process that is dynamically re-estimated using a moving window with window length 250 days. These forecasts are then compared pairwise by regressing the imperfect volatility proxy TSRV sampled at the 5-minute frequency on the two forecasts under consideration. The left upper segment labeled RV+TSRV, for example, shows the coefficients of the RV forecasts (β_1) and TSRV forecasts (β_2) with their respective t-statistics, as well as the regression R-squared. Panel A and B show the results for IBM and ZMH, respectively. In panels C and D we repeat the analysis with an outlier correction applied to 10/10/2008.

2.4.2 Outlier correction

During our out-of-sample period, which contains the height of the recent financial crisis and the beginning of its aftermath, several trading days exhibited extremely high volatility and can be regarded as outliers. It is interesting to analyze how an outlier correction would influence the results. There is a vast literature on how to adjust for outliers, such as truncating values that are more than several standard deviations away from the (local) average of the volatility process or incorporating dummy variables etc. Because there are several ways to go and we do not want to alter the empirical data too much we will only incorporate a dummy for 10/10/2008 which was found to be an outlier using several approaches and analyze how this alters the results discussed above.[18]

The Newey-West t-statistic for the dummy variable on 10/10/2008 is larger than 70 for all estimators when using Mincer-Zarnowitz regressions and for the IBM data the t-stat is 145 for the dummy when using RV forecasts. Note the huge increase in the Mincer-Zarnowitz R^2's for IBM and ZMH by explicitly incorporating this outlier. For the IBM data the average R^2 shifts 18.1% in absolute terms and 35.8% in relative terms and for ZMH the shifts are 10.3% and 31.0%, respectively.

For the IBM data the conclusions do not change in the sense that if we rank the forecasts on the Mincer-Zarnowitz regression R^2 the TSRRh ($R^2 = 68.9\%$) forecasts still slightly outperform (TS)RV and (TS)RR forecasts. Similarly, for the ZMH data we again find that the unadjusted RR has the largest R^2 being 43.7% and if we insist on using a bias-adjusted estimator the TSRRh achieves the best result with $R^2 = 43.6\%$.

For the encompassing regressions we find that for the IBM forecasts the TSRRh are not rendered obsolete by the other forecasts. The ZMH results illustrate that in the encompassing regressions with outlier correction the TSRRh performs satisfactorily as it outperforms the (TS)RV and TSRR and it competes with the unadjusted realized range. Hence, including an outlier dummy for the most severe outlier in our sample does not alter the main conclusions.

[18] For example, the RV on 10/10/2008 is more than 8 standard deviations away from the unconditional average.

2.5 Conclusion

We have proposed a novel heuristic bias-correction for realized range-based volatility estimates. For the heuristic adjustment we use three inputs that are easily and accurately estimated from high-frequency data. The needed inputs are estimates of the following quantities: (i) the daily range that is unaffected by noise, (ii) the non-trading probability and (iii) the half-spread. Using these inputs we simulate a geometric Brownian motion with variance (i) and implement noise with settings (ii) and (iii). For the simulated Brownian motions we keep count of how many intraday ranges are upward biased (most likely), unbiased or downward biased (least likely). By averaging over simulation runs we estimate probabilities for the three cases that can be attached to the ranks of sorted intraday ranges. We apply these probability ranks to the sorted vector of initial high-low ranges for which we are now able to indicate whether an intraday range is expected to be upward biased, unbiased or downward biased.

Using three stochastic volatility models for the integrated volatility, which can include jumps, leverage effects and dependence in the increments of a Brownian motion, we find that in the presence of bid-ask bounce and non-trading, volatility estimates based on the new heuristically bias-adjusted realized range estimator (TSRRh) are more efficient than estimates based on the realized variance, realized range and their two time scales adjusted counterparts.

In an empirical setting we evaluated out-of-sample volatility forecasts using Mincer-Zarnowitz and encompassing forecast regressions. For the relatively liquid IBM stock we find that the heuristically bias-adjusted realized range estimator (TSRRh) compares favorably to forecasts based on the (TS)RV and (TS)RR estimators. For the relatively illiquid Zimmer Holdings stock (ZMH), we find that TSRRh improves upon (TS)RV and TSRR forecasts and is on par with the RR estimator. Since the heuristic adjustment does not assume that every observed intraday range is always upward biased due to microstructure frictions, this essentially leads to a 'smaller' bias-adjustment and therefore higher volatility estimates. 'Overcorrecting' for microstructure noise frictions for illiquid assets that have relatively small noise-to-volatility ratios, can lead to unwanted underestimation of volatilities and risk measures such as Value-at-Risk, that use volatility estimates as inputs.

Chapter 3

Range-based Covariance Estimation using High-Frequency Data: The Realized Co-Range*

3.1 Introduction

This paper develops a novel estimator of the daily quadratic covariation between asset returns, based on high-frequency intraday price ranges. This so-called realized co-range estimator combines two recent ideas that have revived the use of high-low ranges for estimating the volatility and covariance of asset returns. First, it employs the realized range, independently introduced by Martens and Van Dijk (2007) and Christensen and Podolskij (2007), to estimate daily volatility by means of the sum of squared intraday price ranges. Second, it adopts the suggestion of Brandt and Diebold (2006) to use range-based volatility estimates of a portfolio and the individual assets to estimate their covariance.

The increasing availability of high-frequency asset price data has led to a rapidly expanding literature on the use of intraday prices to measure, model and forecast daily volatility, see Andersen et al. (2006a) and McAleer and Medeiros (2008) for recent reviews. Based on the theoretical results of Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2001, 2003), in the absence of microstructure noise the

*This chapter is based on the article by Bannouh, Van Dijk and Martens (2009).

sum of squared intraday returns, called realized variance, is a highly efficient estimator of the daily quadratic variation. The benefits of high-frequency data continue to hold in a multivariate context as intraday returns provide more accurate estimates of the daily covariance between asset returns. Barndorff-Nielsen and Shephard (2004) show that the realized covariance, that is, the sum of cross-products of intraday returns, converges in probability to the quadratic covariation. The economic value of using realized covariances in a volatility timing strategy has been explored by Fleming et al. (2003) and De Pooter et al. (2008), who find that a risk-averse investor is willing to pay between 50 and 200 basis points per annum to switch from covariance measurement based on daily data to intraday data.

Intraday price ranges have only recently been considered for the purpose of estimating daily volatility. This might appear surprising, given that it has been known since Parkinson (1980) that the high-low range is considerably more efficient as an estimator of volatility than the squared return, with a variance that is five times smaller. Martens and Van Dijk (2007) and Christensen and Podolskij (2007) exploit this result for developing an estimator of daily volatility based on intraday ranges, which is more efficient than the realized variance (sampled at the same frequency) by the same 5:1 ratio. A plausible reason for ignoring the range in the context of high-frequency data is that the extension to the multivariate case, that is, to estimate the covariation between asset returns, was an unresolved challenge until the recent proposal by Brandt and Diebold (2006), see also Brunetti and Lildholdt (2007). This exploits the fact that the covariance between two assets can be expressed in terms of their individual variances and the variance of a portfolio of the two assets. The range-based covariance estimator is then obtained by using daily price ranges to estimate these variances.

The main contribution of this paper is to combine these two ideas to provide an intraday range-based covariance estimator. In particular, we employ the realized range of Martens and Van Dijk (2007) and Christensen and Podolskij (2007) for estimating the daily volatilities that enter the co-range estimator of Brandt and Diebold (2006), which results in the novel realized co-range estimator. Given the relative efficiency of the realized range estimator, we expect the realized co-range also to be more efficient than the realized covariance.

Market microstructure effects pose a serious challenge to the use of high-frequency data. In the univariate case, the most important effect is due to bid-ask bounce, which renders the standard realized variance estimator biased and inconsistent. This has led to several proposals for bias-corrected realized volatility estimators on the one hand, and for determining the optimal sampling frequency for the standard realized variance estimator on the other hand.[1] As discussed in Martens and Van Dijk (2007) and Christensen and Podolskij (2007), the realized range estimator in addition suffers from infrequent trading. This causes a downward bias as the observed minimum and maximum price over- and underestimate the true minimum and maximum, respectively. In the multivariate case, the greatest concern for realized covariance estimators is the presence of non-synchronous trading. As a result of assets trading at different times, estimates of their covariance will be biased towards zero. This so-called Epps (1979) effect becomes worse with increasing sampling frequency. The impact of microstructure noise on the realized covariance estimator has recently received a considerable amount of attention, see Hayashi and Yoshida (2005), Sheppard (2006), Griffin and Oomen (2011), Zhang (2011), Voev and Lunde (2007), and Bandi et al. (2008), among others.

In this paper we propose the use of an additive bias-correction for the realized co-range, where we add the average difference between the covariance estimates based on daily ranges and on intraday ranges over the previous Q trading days to the standard realized co-range estimate. The main advantage of this additive bias-correction is that it deals with the "net" bias that arises due to different possible microstructure effects. This contrasts to other bias-corrections that have been put forward in the context of range-based volatility estimators, which correct only for a single source of bias such as infrequent trading, see Rogers and Satchell (1991), or bid-ask bounce, see Christensen et al. (2009).

We assess the performance of the realized co-range estimator by means of extensive simulation experiments and an empirical application. In the simulations we start from an idealized continuous-time setting without microstructure noise, where we find that the realized co-range outperforms the returns-based realized covariance estimator, as expected. In more realistic settings that incorporate bid-ask bounce, infrequent trading and non-synchronous trading, we find that the impact of the different microstructure effects is reduced successfully by using the additive bias-correction. The bias-corrected realized co-range is more efficient than the bias-corrected realized covariance estimator for plausible levels of noise, as is the case for the daily co-range compared to the daily covariance estimator.

In the empirical application we focus on the economic value of high-frequency intraday ranges for estimating covariances. Specifically, we adopt the framework developed by Fleming et al. (2001, 2003) and use the realized co-range in a dynamic

[1] The choice of sampling frequency reflects the trade-off between accuracy, which is theoretically optimized using the highest possible frequency, and microstructure noise, which calls for lowering the data frequency. See Oomen (2005), Zhang et al. (2005), Aït-Sahalia et al. (2005), Bandi and Russell (2006, 2008a), Hansen and Lunde (2006) and Awartani et al. (2009) among others, for recent discussions.

volatility timing strategy for constructing mean-variance efficient portfolios consisting of futures on stocks, bonds and gold. Sampling at the popular 5-minute frequency, we find that the bias-corrected realized co-range and realized covariance estimators provide similar results in terms of portfolio return and risk, before transaction costs. At first sight, this indicates that both estimators render similar (co)variance dynamics. Closer inspection reveals that the correlation estimates obtained from the realized co-range are less 'noisy' than those resulting from the realized covariance. In the volatility timing strategy this causes less variation in the realized co-range portfolio weights, which implies lower turnover and, hence, lower transaction costs. Taking transaction costs into account, we find that a mean-variance investor would be willing to pay 60 basis points per annum to switch from the realized covariance to the realized co-range when the decay parameter of the used exponential weighting scheme is estimated using a maximum-likelihood procedure. A sensitivity analysis of the decay parameter, which determines how much weight is put on recent estimates for predicting covariances, illustrates that the realized co-range outperforms the realized covariance also in terms of risk-return characteristics and therefore gross Sharpe ratios when more weight is put on the most recent estimate.

The remainder of this paper is organized as follows. In Section 3.2 we discuss the realized (co-)variance and realized (co-)range estimators. In Section 3.3 we use Monte Carlo simulations to analyze the properties of the realized co-range and realized covariance estimators in the presence of noise. In Section 3.4 we consider the empirical application to volatility timing strategies. We conclude in Section 3.5.

3.2 Volatility and covariation estimators

The traditional way to estimate daily volatility *ex post* is by means of the daily squared return. Although this estimator is unbiased (in the absence of drift in the asset price), it also is very noisy in the sense that it has a high variance. In order to improve accuracy, high-frequency intra-day returns may be used. Dividing day t into M non-overlapping intervals of length $\Delta = 1/M$, the realized variance estimator is given by

$$RV_t^{\Delta} \equiv \sum_{m=1}^{M} (\log P_{t,m} - \log P_{t,m-1})^2, \tag{3.1}$$

where $P_{t,m}$ is the last observed transaction price during the m-th interval on day t. In the absence of noise and under weak regularity conditions for the stochastic log-price process, the realized variance is a consistent estimator of the daily integrated variance, see Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003), among others.

Further efficiency gains can be achieved by exploiting the superior properties of the range as a volatility proxy compared to squared returns. In particular, Martens and Van Dijk (2007) and Christensen and Podolskij (2007) define the realized range as the sum of intraday price ranges, that is,

$$RR_t^{\Delta} = \sum_{m=1}^{M} \frac{1}{4\log 2}(\log H_{t,m} - \log L_{t,m})^2, \tag{3.2}$$

where the high price $H_{t,m}$ and the low price $L_{t,m}$ are defined as the maximum and minimum, respectively, of all transaction prices observed during the m-th interval on day t. The scaling factor $1/(4\log 2)$ is the second moment of the range of a Brownian motion. Note that the realized range exploits the complete price path in the intra-day intervals, while the realized variance only uses the first and last price observations. For this reason, the realized range achieves a lower variance than the realized variance based on the same sampling frequency. Specifically, assuming that the asset price P_t follows a geometric Brownian motion with constant instantaneous variance σ^2, the variance of the realized range estimator is equal to $0.407\sigma^4\Delta^2$, compared to $2\sigma^4\Delta^2$ for the realized variance. Hence, the variance of the realized range is approximately five times smaller than the variance of the realized variance estimator. Christensen and Podolskij (2007) show that the realized range remains consistent and relatively efficient in case volatility is time-varying, requiring only mild assumptions on the stochastic volatility process σ_t.

3.2.1 Realized covariance and realized co-range

The intraday return-based realized variance in (3.1) provides an efficient estimator for the variances of individual asset returns. Similarly, the realized covariance between assets i and j can be obtained by summing cross-products of intraday returns,

$$RCV_t^{\Delta} = \sum_{m=1}^{M} r_{i,t,m} r_{j,t,m}, \tag{3.3}$$

where $r_{i,t,m} = log(P_{i,t,m}/P_{i,t,m-1})$ is the continuously compounded return on asset i during the m-th interval on day t. Barndorff-Nielsen and Shephard (2004) study the properties of the realized covariance, showing that it is consistent for the daily integrated covariation under mild regularity conditions.

Brandt and Diebold (2006) introduce a simple but effective way to estimate the covariance by combining range-based estimates of the variances of two individual assets and of a portfolio composed of these assets. Consider a portfolio of assets i and j with weights λ_i and $\lambda_j = 1 - \lambda_i$, with asset returns denoted as r_i and r_j. The variance of the portfolio return $r_p \equiv \lambda_i r_i + \lambda_j r_j$ is given by

$$\mathrm{Var}[r_p] = \lambda_i^2 \mathrm{Var}[r_i] + \lambda_j^2 \mathrm{Var}[r_j] + 2\lambda_i\lambda_j \mathrm{Cov}[r_i, r_j],$$

such that, after rearranging

$$\mathrm{Cov}[r_i, r_j] = \frac{1}{2\lambda_i\lambda_j}\left(\mathrm{Var}[r_p] - \lambda_i^2 \mathrm{Var}[r_i] - \lambda_j^2 \mathrm{Var}[r_j]\right). \tag{3.4}$$

The daily co-range estimator of Brandt and Diebold (2006) is obtained by using the daily high-low range of the corresponding prices of the portfolio and the individual assets as estimators of three variances on the right-hand side of (3.4). Due to the fact that the range-based variance estimates make use of the complete price path during the day, the daily co-range estimator is relatively efficient compared to the cross-product of daily returns. In fact, Brandt and Diebold (2006) find that, in the absence of noise, the efficiency of the daily co-range is between that of the realized covariance in (3.3) computed using 3-hour and 6-hour intraday returns. Furthermore, the daily co-range turns out to be robust to the effects of microstructure noise such as bid-ask bounce and non-synchronous trading, which severely affect the realized covariance.

We combine the idea of using intraday ranges for estimating daily volatilities, with the idea of estimating the daily covariance from estimates of the volatilities of the individual assets and of the portfolio. Specifically, using the realized range defined in (3.2) for estimating the three variances on the right-hand side of (3.4), we obtain the realized co-range estimator

$$RCR_t^{\Delta} = \frac{1}{2\lambda_i\lambda_j}\left(RR_{p,t}^{\Delta} - \lambda_i^2 RR_{i,t}^{\Delta} - \lambda_j^2 RR_{j,t}^{\Delta}\right), \tag{3.5}$$

where $RR_{p,t}$ is the realized range of the portfolio, and $RR_{i,t}$ and $RR_{j,t}$ are the realized ranges of the individual assets. Each realized range is estimated using (3.2).[2] It is

[2] Note that (3.4) can also be used to estimate the daily covariance with realized variances, by using RV as defined in (3.1) to estimate the variances on the right-hand side. However, this yields exactly the realized covariance as given in (3.3), as

$$\frac{1}{2\lambda_i\lambda_j}\left(\sum_{m=1}^{M}(\lambda_i r_{i,m} + \lambda_j r_{j,m})^2 - \lambda_i^2\sum_{m=1}^{M} r_{i,m}^2 - \lambda_j^2 \sum_{m=1}^{M} r_{j,m}^2\right) = \sum_{m=1}^{M} r_{i,m} r_{j,m}.$$

important to note that the high (low) price of asset i in a given intraday interval will probably be obtained at a different point in time than the high (low) price of asset j. The high-low range of the portfolio then is not the same as the weighted sum of the individual ranges. Therefore, it is necessary to construct a portfolio price path at the highest possible sampling frequency and estimate the realized range of the portfolio, $RR_{p,t}$, using (3.2).

3.2.2 Bias-correction

As discussed in the introduction, market microstructure effects hamper the use of high-frequency data for estimating daily variances and covariances. First, both the realized variance in (3.1) and the realized range in (3.2) suffer from an upward bias due to the presence of bid-ask bounce. For example, when trading is continuous, the observed high price in a given interval is an ask and the observed low price is a bid with probability close to 1. The realized range therefore overestimates the true daily variance by an amount equal to the squared spread s^2 times the number of intraday intervals M. Second, while infrequent trading does not affect the realized variance, it leads to a downward bias in the realized range. When the continuous underlying price process is only observed at discrete points in time, the observed high price during a given intraday interval underestimates the true maximum. Similarly, the observed low price overestimates the true minimum. A correction for the infrequent trading bias in range-based volatility estimators has been proposed by Rogers and Satchell (1991). Christensen and Podolskij (2007) directly account for infrequent trading in their realized range estimator by replacing the scaling factor $1/(4\log 2)$ in (3.2) by the second moment of the range of a Brownian motion when it is observed infrequently. Martens and Van Dijk (2007) suggest to deal with the "net" bias due to the combined effects of bid-ask bounce and infrequent trading on the realized range by applying a multiplicative bias-correction, see also Fleming et al. (2003). Specifically, the scaled realized range is defined by

$$RR_{S,t}^{\Delta} = \left(\frac{\sum_{q=1}^{Q} RR_{t-q}}{\sum_{q=1}^{Q} RR_{t-q}^{\Delta}} \right) RR_t^{\Delta}, \tag{3.6}$$

where $RR_t \equiv RR_t^1$ is the daily range. Hence, the multiplicative correction factor is the ratio of the average daily range estimator and the average of the realized range over the past Q days.

Although Martens and Van Dijk (2007) demonstrate that the multiplicative correction in (3.6) is quite effective in removing the bias in the realized range, here we consider an alternative, *additive* correction. This is motivated by observing that the presence of market microstructure effects is often represented by assuming that the observed log price $\log P_t$ is equal to the efficient log price $\log P_t^*$ plus an additive noise term ε_t:

$$\log P_t = \log P_t^* + \varepsilon_t, \tag{3.7}$$

where ε_t is assumed to have zero mean and variance σ_ε^2. In this set-up, the realized variance based on observed returns converges to the true integrated variance *plus* a bias term determined by the noise variance σ_ε^2 and the covariance between $\log P_t^*$ and ε_t (which often is assumed to be zero, but see the discussion in Hansen and Lunde (2006)). This suggests that we may use an *additive* bias-correction, and define a corrected realized variance estimator as

$$RV_{C,t}^{\Delta} = RV_t^{\Delta} + \frac{1}{Q}\left(\sum_{q=1}^{Q} RV_{t-q} - \sum_{q=1}^{Q} RV_{t-q}^{\Delta}\right), \tag{3.8}$$

where $RV_t \equiv RV_t^1$ is the daily squared return. As discussed in Christensen et al. (2009), deriving consistent estimators of the integrated variance based on intraday high-low ranges is difficult, if not impossible in the presence of general microstructure noise as in (3.7). For that reason, we adopt a pragmatic approach and consider a realized range estimator with an additive bias-correction of the form (3.8), that is,

$$RR_{C,t}^{\Delta} = RR_t^{\Delta} + \frac{1}{Q}\left(\sum_{q=1}^{Q} RR_{t-q} - \sum_{q=1}^{Q} RR_{t-q}^{\Delta}\right). \tag{3.9}$$

For covariance estimators based on intraday data, the most important microstructure effect is the occurrence of non-synchronous trading. Using returns or ranges over fixed intraday intervals results in covariance estimates that are biased towards zero. This so-called Epps (1979) effect becomes worse with increasing sampling frequency, and in the limit the standard realized covariance and realized co-range estimators converge to zero. Most of the recent proposals for alternative high-frequency covariance estimators are in fact attempts to fix the bias due to non-synchronous trading, see Hayashi and Yoshida (2005) and Sheppard (2006), among others. Griffin and Oomen (2011), Zhang (2011) and Voev and Lunde (2007) consider the combination of non-synchronous trading and additive microstructure noise. Here we limit ourselves to implementing the additive bias-correction discussed above for the realized co-range and realized covariance.

3.3 Monte Carlo simulation

In this section we investigate the performance of the realized co-range estimator in a controlled environment by means of Monte Carlo simulations. Throughout we compare the realized co-range estimator with the realized covariance estimator. Of particular interest are the effects of bid-ask bounce, infrequent trading and non-synchronous trading on the two estimators and the usefulness of the additive bias-correction described in the previous section.

3.3.1 Simulation design

We simulate prices for two correlated assets for 24-hour days, assuming that trading takes place around the clock. For each day t, the initial prices for both assets are set equal to 1, and subsequent log prices for asset $i = 1, 2$ are simulated using

$$\log P^*_{i,t+k/K} = \log P^*_{i,t+(k-1)/K} + \varepsilon_{i,t+k/K}, \qquad i = 1, 2, \;\; k = 1, 2, \ldots, K, \qquad (3.10)$$

where K is the number of prices per day. We simulate 100 prices per second, such that $K = 8,640,000$, where price observations are equidistant and occur synchronously for the two assets. The shocks $\varepsilon_{i,t+k/K}$ are serially uncorrelated and normally distributed with mean zero and variance $\sigma_i^2/(D \cdot K)$, where D is the number of trading days in a year, which we set equal to 250. The annualized standard deviations σ_i of the log price processes are set equal to 0.20 and 0.40 (20% and 40%) for assets 1 and 2, respectively. The shocks $\varepsilon_{1,t+k/K}$ and $\varepsilon_{2,t+k/K}$ are contemporaneously correlated with correlation coefficient ρ, which we set equal to 0.50, resulting in a covariance between the asset returns of 0.04.

We consider the effects of bid-ask bounce by assuming that transactions take place either at the ask price or at the bid price, which are equal to the true price plus and minus half the spread, respectively. Hence, the actually observed price $P_{i,t+k/K}$ is equal to $P^*_{i,t+k/K} + s/2$ (ask) or $P^*_{i,t+k/K} - s/2$ (bid), where s is the bid-ask spread and $P^*_{i,t+k/K}$ is the true price obtained from (3.10). We assume that bid and ask prices occur equally likely and that the occurrence of bid and ask prices is independent across assets. Infrequent trading is implemented by imposing that, given the price path obtained from (3.10), the probability to actually observe the price $P^*_{i,t+k/K}$ is equal to $p_{\text{obs}} = 1/(100\tau)$. Put differently, the price of each asset is observed on average only every τ seconds. Price observations for the two assets occur independently, such that in addition we observe prices non-synchronously.

Computational details

We assess the potential merits of using intraday ranges for measuring (daily) co-movement by computing both the realized co-range and the realized covariance. To do so we divide the trading day into Δ-minute intervals, which is referred to as the Δ-minute frequency below, where we vary Δ among $1, 2, 3, 4, 5, 10, 15, 20, 30, 45, 60$, and 1440. For example, when $\Delta = 5$ we divide the 24-hour day into 288 five-minute intervals. For the realized covariance at this sampling frequency the cross-products of five-minute returns are summed to obtain the realized covariance at that frequency, as in (3.3). The realized co-range is computed using (3.5) as follows. For the two assets the high and low prices are computed per five-minute interval and the resulting five-minute squared ranges are summed to obtain the realized ranges for the day, as in (3.2). To obtain the realized range of a portfolio consisting of the two assets, we first compute the intraday prices of an equally-weighted portfolio setting $\lambda_i = 0.5$, $i = 1, 2$, and assuming continuous rebalancing throughout the day.[3] Note that in the case of synchronous price observations for the two assets we can compute exact portfolio prices at each instant. In case of non-synchronous trading the portfolio price is updated each time a new price for one of the two assets occurs, combining this with the most recently observed (hence stale) price for the other asset. Second, the portfolio prices are used to compute the corresponding realized range in the usual way. This is then combined with the realized ranges for the two individual assets using (3.5) to compute the realized co-range. We also consider the bias-corrected versions of the realized covariance and the realized co-range, computed according to (3.8) and (3.9), respectively.[4]

In our experiments the characteristics of bid-ask bounce and infrequent trading are identical for all trading days, such that in principle we could use a large number of trading days Q to compute the additive adjustment factor to fully explore the merits of the bias-adjustment procedure. In practice, however, the characteristics of microstructure noise are likely to change over time and a smaller value of Q seems more appropriate. We therefore set $Q = 66$ throughout the simulations. The sensitivity of the results with respect to the value of Q is discussed in more detail below. For each selected frequency we compute the mean and Root Mean Squared Error (RMSE) for the various estimators of the assets' covariation, taken over 5000 simulated trading days.

[3]We perform a sensitivity analysis on the portfolio weights by experimenting with $\lambda_1 = 0.1, 0.3, 0.5, 0.7, 0.9$. We find that the choice of portfolio weights has only minor influence on the efficiency (RMSE) of the co-range estimator. Detailed simulation results are available upon request.

[4]We also considered an alternative bias-correction for the realized co-range, by computing it according to (3.5) but using the scaled realized ranges as defined in (3.9). This "indirect" bias-correction results in qualitatively and quantitatively similar results (which are available upon request) as the "direct" bias-correction reported in this section.

3.3.2 Simulation results

Table 3.1 reports results for experiments with (close to) continuous trading where all $K = 8,640,000$ prices on a given day are observed. In order to establish a benchmark, panel A shows results for the ideal situation where the 'clean' prices $P^*_{i,t+k/K}$ are observed. Panel B considers the effects of bid-ask bounce, where we set the spread s equal to 0.0005 (or 0.05% of the starting price of 1) as in Brandt and Diebold (2006).[5]

As expected, in Panel A the RMSE of the realized co-range is always substantially lower than that of the realized covariance at the same frequency. In fact, for all but the very highest sampling frequencies the ratio of the RMSE's is close to $\sqrt{5}$, which is the ratio of the standard deviations of the daily squared returns and daily ranges. Hence, the same efficiency factor seems to apply to the intraday range- and return-based measures of covariation examined here. The slight loss in relative efficiency of the realized co-range at the highest sampling frequencies is due to the downward bias it experiences when the underlying price ranges are computed over very short intervals, as shown in the second column of Table 3.1. This is inherent to the nature of the high-low range: In case the price path is not observed continuously (in this case we observe 'only' 6000 prices per minute) the observed minimum and maximum over- and underestimate the true high and low prices, respectively, such that the observed range underestimates the true range. We investigate the effects of infrequent trading in more detail below. Finally, we mention that also the daily co-range suggested by Brandt and Diebold (2006), reported in the last row of this table ($\Delta = 1440$ minutes), achieves an RMSE that is substantially lower at 4.044 compared to 8.830 for the cross-product of daily returns.

The results in panel B demonstrate that, as expected, in the presence of bid-ask bounce the realized co-range suffers from a pronounced upward bias, which gets worse as the sampling frequency increases. With continuous price observations the observed range for the individual assets will overestimate the true ranges by a quantity close to the spread, as the maximum price will be an ask and the minimum price will be a bid with probability close to 1. For the equally-weighted portfolio, the true range is overestimated by the bid-ask spread as well when trading is continuous. Hence, the net effect on the realized co-range in (3.5) is an upward bias. The realized covariance is not affected by bid-ask spread, which also is conform expectations. In this particular parameter configuration the realized co-range outperforms the realized covariance up to the 45-minute frequency. For higher sampling frequencies the RMSE

[5] Results for other magnitudes of the bid-ask spread are summarized at the end of this section, with details being available upon request.

Table 3.1: Realized co-range and realized covariance with continuous trading and bid-ask bounce

Frequency (minutes)	RCR_t^{Δ} Mean	RCR_t^{Δ} RMSE	$RCR_{C,t}^{\Delta}$ Mean	$RCR_{C,t}^{\Delta}$ RMSE	RCV_t^{Δ} Mean	RCV_t^{Δ} RMSE	$RCV_{C,t}^{\Delta}$ Mean	$RCV_{C,t}^{\Delta}$ RMSE
Panel A: No bid-ask bounce ($s = 0$)								
1	3.928	0.130	3.945	0.505	3.999	0.229	4.000	1.045
2	3.946	0.160	3.944	0.513	3.996	0.322	3.999	1.067
3	3.955	0.190	3.944	0.526	3.990	0.402	4.000	1.112
4	3.961	0.218	3.944	0.539	3.988	0.469	3.999	1.125
5	3.963	0.236	3.945	0.545	3.995	0.513	4.002	1.147
10	3.965	0.335	3.944	0.591	3.963	0.733	3.996	1.259
15	3.957	0.422	3.943	0.647	3.961	0.931	4.000	1.404
20	3.959	0.469	3.943	0.684	3.958	1.008	3.999	1.440
30	3.961	0.595	3.942	0.769	3.968	1.302	4.000	1.645
45	3.953	0.724	3.940	0.868	3.976	1.607	3.995	1.903
60	3.951	0.846	3.937	0.969	3.994	1.850	3.986	2.092
1440	3.928	4.044	3.928	4.044	3.985	8.830	3.985	8.830
Panel B: Bid-ask bounce ($s = 0.0005$)								
1	14.615	10.617	4.154	0.576	3.992	0.359	4.017	1.134
2	10.881	6.886	4.153	0.580	3.996	0.428	4.015	1.155
3	9.395	5.402	4.153	0.594	3.986	0.472	4.019	1.179
4	8.556	4.565	4.153	0.610	3.988	0.542	4.013	1.204
5	8.000	4.012	4.153	0.617	3.983	0.583	4.018	1.222
10	6.694	2.725	4.153	0.668	3.955	0.778	4.011	1.326
15	6.141	2.198	4.152	0.730	3.962	0.964	4.019	1.448
20	5.828	1.907	4.152	0.769	3.942	1.037	4.015	1.489
30	5.463	1.611	4.151	0.857	3.967	1.329	4.016	1.679
45	5.165	1.413	4.149	0.953	3.975	1.640	4.008	1.950
60	4.993	1.356	4.145	1.057	4.004	1.858	4.003	2.118
1440	4.135	4.128	4.135	4.128	3.982	8.871	3.982	8.871

Note: The table summarizes the results of simulating 5000 days of 8,640,000 (log) prices (100 prices per second) from a bivariate normal distribution with mean zero, variance 4 and 16 and correlation 0.5, such that the true covariance is equal to 4. Panel A reports results in case all prices are observed without distortion due to bid-ask bounce. Panel B reports results in case all prices are observed, but are converted to bid and ask prices (with equal probability) by either subtracting or adding half the spread $s = 0.0005$ (on a starting price of 1). The occurrence of bid and ask prices for the two assets is independent. For each day the realized co-range (RCR_t^{Δ}), the bias-corrected realized co-range ($RCR_{C,t}^{\Delta}$), the realized covariance (RCV_t^{Δ}), and the bias-corrected realized covariance ($RCV_{C,t}^{\Delta}$) are computed for various sampling frequencies shown in column 1. $RCR_{C,t}^{\Delta}$ and $RCV_{C,t}^{\Delta}$ are obtained from (3.9) and (3.8) with $Q = 66$ (with RR and RV replaced by RCR and RCV).

of the realized covariance is smaller. Bias-correcting the realized co-range works remarkably well, in the sense that the bias is removed completely and the RMSE values are almost brought back to the original level observed for $RCR^{\Delta}_{C,t}$ in the ideal case without bid-ask bounce. For sampling frequencies of 10 minutes or lower, the RMSE of the bias-corrected realized co-range is smaller than the RMSE of the realized covariance.

Table 3.2 shows the results when infrequent trading occurs, such that for both assets the price is observed on average only every $\tau = 12$ seconds.[6] Again panels A and B show results for experiments without and with bid-ask bounce, respectively.

The results in panel A show that for the realized co-range the RMSE first decreases when increasing the sampling frequency up to 20 minutes. It increases again for higher frequencies because the larger bias due to non-trading (and hence underestimating the range for each intraday interval) then outweighs the reduction in the standard deviation of the estimates. The realized covariance estimator is not affected by infrequent trading but does suffer from non-synchronous trading in terms of a bias towards zero.[7] As a result, at the 15-minute frequency the realized co-range still is a more accurate measure of co-movement than the corresponding realized covariance, but at higher frequencies the realized covariance has a lower RMSE than the realized co-range.[8] Bias-correcting the realized co-range in this case eliminates the bias to a large extent but not completely, due to the fact that the daily co-range also is somewhat biased downward due to the infrequent and non-synchronous trading. The bias-adjustment does reduce the RMSE of the realized co-range considerably, such that $RCR^{\Delta}_{C,t}$ is more accurate than the realized covariance for all sampling frequencies except 2-5 minutes.

In case bid-ask bounce and infrequent and non-synchronous trading are jointly present in panel B of Table 3.2, we find that the realized co-range still suffers from an upward bias, but it is of a considerably smaller magnitude than in the case of bid-ask bounce only due to the off-setting negative bias induced by infrequent and non-synchronous trading. As a result, the realized co-range now has a lower RMSE than the realized covariance at all sampling frequencies. The overall minimum RMSE for the realized co-range is obtained at the 2-minute sampling frequency and equals 0.205. For the realized covariance the optimal frequency is the 3-minute frequency for which the RMSE is 0.551. Note that in this particular setting, where the different

[6] Results for other trading frequencies are summarized at the end of this section, with details being available upon request.

[7] This becomes evident from unreported results from simulations with infrequent but simultaneous trading for the two assets. Detailed results are available upon request.

[8] Of course the exact frequency at which one estimator improves over the other will depend on the trading intensity. For example, when transaction prices are observed once per second on average the realized co-range improves over the realized covariance up to the five-minute frequency.

Table 3.2: Realized co-range and realized covariance with infrequent trading and bid-ask bounce

Frequency (minutes)	RCR_t^{Δ} Mean	RCR_t^{Δ} RMSE	$RCR_{C,t}^{\Delta}$ Mean	$RCR_{C,t}^{\Delta}$ RMSE	RCV_t^{Δ} Mean	RCV_t^{Δ} RMSE	$RCV_{C,t}^{\Delta}$ Mean	$RCV_{C,t}^{\Delta}$ RMSE
Panel A: No bid-ask bounce ($s = 0$)								
1	2.017	1.985	3.866	0.547	3.203	0.828	3.979	1.210
2	2.470	1.535	3.866	0.557	3.596	0.521	3.980	1.235
3	2.696	1.314	3.866	0.564	3.731	0.482	3.979	1.254
4	2.836	1.179	3.866	0.572	3.791	0.506	3.979	1.277
5	2.937	1.084	3.866	0.583	3.833	0.550	3.980	1.300
10	3.211	0.849	3.866	0.626	3.914	0.751	3.979	1.399
15	3.339	0.770	3.865	0.667	3.942	0.927	3.979	1.493
20	3.418	0.739	3.866	0.703	3.955	1.045	3.979	1.574
30	3.516	0.748	3.866	0.780	3.961	1.294	3.977	1.743
45	3.592	0.813	3.866	0.877	3.960	1.590	3.978	1.967
60	3.641	0.893	3.864	0.973	3.959	1.818	3.979	2.158
1440	3.862	4.173	3.862	4.173	3.976	9.192	3.976	9.192
Panel B: Bid-ask bounce ($s = 0.0005$)								
1	3.861	0.210	3.956	0.555	3.213	0.869	3.973	1.237
2	4.082	0.205	3.956	0.565	3.599	0.583	3.973	1.257
3	4.129	0.250	3.956	0.574	3.732	0.551	3.972	1.280
4	4.142	0.278	3.956	0.584	3.793	0.574	3.972	1.301
5	4.143	0.301	3.957	0.596	3.836	0.608	3.973	1.322
10	4.139	0.387	3.956	0.645	3.917	0.794	3.971	1.425
15	4.126	0.460	3.956	0.689	3.944	0.956	3.972	1.512
20	4.114	0.514	3.956	0.727	3.962	1.078	3.971	1.598
30	4.100	0.625	3.956	0.809	3.963	1.312	3.970	1.757
45	4.079	0.752	3.956	0.908	3.955	1.603	3.971	1.978
60	4.069	0.864	3.955	1.007	3.954	1.829	3.971	2.169
1440	3.952	4.208	3.952	4.208	3.968	9.183	3.968	9.183

Note: The table summarizes the results of simulating 5000 days of 8,640,000 (log) prices (100 prices per second) from a bivariate normal distribution with mean zero, variance 4 and 16 and correlation 0.5, such that the true covariance is equal to 4. Subsequently with probability $p_{\text{obs}} = 1/(100\tau)$ we observe a price and with probability $1 - p_{\text{obs}}$ we do not, such that the price is observed on average only every τ seconds. The table reports results for $\tau = 12$. Panel A reports results in case prices are observed without distortion due to bid-ask bounce. Panel B reports results in case the observed prices are converted to bid and ask prices (with equal probability) by either subtracting or adding half the spread $s = 0.0005$ (on a starting price of 1). The occurrence of price observations and bid and ask prices for the two assets is independent. For each day the realized co-range (RCR_t^{Δ}), the bias-corrected realized co-range ($RCR_{C,t}^{\Delta}$), the realized covariance (RCV_t^{Δ}), and the bias-corrected realized covariance ($RCV_{C,t}^{\Delta}$) are computed for various sampling frequencies shown in column 1. $RCR_{C,t}^{\Delta}$ and $RCV_{C,t}^{\Delta}$ are obtained from (3.9) and (3.8) with $Q = 66$ (with RR and RV replaced by RCR and RCV).

biases affecting the realized co-range approximately cancel out, bias-adjustment is not attractive. Although the mean of $RCR^{\Delta}_{C,t}$ is closer to the true covariance of 4, the variance increases considerably due to the bias-adjustment such that the RMSE increases compared to the 'raw' realized co-range RCR^{Δ}_{t}. Finally, in Table 3.2, we also observe that bias-correcting the realized covariance is never worthwhile, as apparently the reduction in bias does not outweigh the increase in variance, such that the RMSE of $RCV^{\Delta}_{C,t}$ is always substantially higher than the RMSE of the 'raw' realized covariance RCV^{Δ}_{t}.

3.3.3 Sensitivity analysis: Trading frequency, spread size, and bias correction

Whether or not the realized co-range improves upon the realized covariance, and whether or not bias-adjustment is appropriate of course depends on the (relative) magnitudes of the market microstructure frictions. For that reason, Table 3.3 provides an overview of the different covariance estimators for different trading frequencies τ, ranging from 2 to 60 seconds, and different bid-ask spreads s, ranging from 0 to 0.00075. For each combination of τ and s, the table shows the optimal sampling frequency for the four covariance estimators, along with the corresponding RMSE and mean. The covariance estimator that achieves the lowest RMSE for a given combination of τ and s is shown in italics.

The patterns that can be observed in Table 3.3 are as expected. First, in the absence of bid-ask bounce (first column, $s = 0$), the realized covariance estimator performs best when prices are observed relatively frequently, up to once per 15 seconds on average. Under these circumstances, the downward bias due to non-synchronous trading affects the realized covariance less than the realized co-range and the bias-correction cannot compensate for this. However, when trading is less frequent, the adjusted realized co-range achieves the lowest RMSE.

Second, in the presence of bid-ask bounce, the unadjusted realized co-range performs best for those combinations of τ and s for which the downward bias due to infrequent trading and the upward bias due to bid-ask bounce approximately cancel. This is the case for s equal to 0.00025 and τ equal to 3 to 6 seconds, for s equal to 0.0005 and τ equal to 10 to 20 seconds, and for s equal to 0.00075 and τ equal to 20 to 40 seconds. Hence the higher the spread (upward bias) the less frequent trading should be (downward bias) to have the two biases offset each other. Note that in financial markets there is indeed a strong negative relationship between spread and trading frequency: The less an asset trades, the higher the spread. We would there-

Table 3.3: Realized co-range and realized covariance with infrequent trading and bid-ask bounce

Trading frequency (seconds)	$s=0$ SF	Mean	RMSE	$s=0.00025$ SF	Mean	RMSE	$s=0.0005$ SF	Mean	RMSE	$s=0.00075$ SF	Meean	RMSE
RCR_t^{Δ}												
2	10	3.652	0.480	8	4.276	0.420	36	4.753	1.029	72	5.016	1.425
3	12	3.613	0.531	*1*	*4.133*	*0.179*	36	4.626	0.934	72	4.916	1.351
4	15	3.600	0.571	*2*	*3.879*	*0.199*	36	4.524	0.866	72	4.834	1.291
6	20	3.576	0.629	*6*	*3.728*	*0.379*	18	4.524	0.721	72	4.700	1.202
10	20	3.464	0.705	12	3.592	0.549	*5*	*4.349*	*0.440*	36	4.730	1.016
12	24	3.458	0.738	15	3.557	0.604	*2*	*4.082*	*0.205*	36	4.621	0.937
15	24	3.401	0.779	16	3.472	0.674	*4*	*3.865*	*0.272*	24	4.592	0.828
20	30	3.387	0.830	24	3.456	0.743	10	3.720	0.447	*1*	*4.109*	*0.238*
30	40	3.358	0.906	32	3.383	0.847	20	3.564	0.643	*4*	*3.872*	*0.292*
40	40	3.271	0.965	40	3.345	0.917	24	3.429	0.765	*10*	*3.686*	*0.483*
50	48	3.253	1.019	40	3.263	0.975	32	3.387	0.851	15	3.546	0.625
60	48	3.192	1.061	45	3.222	1.026	40	3.355	0.917	20	3.456	0.727
$RCR_{C,t}^{\Delta}$												
2	1	3.908	0.542	1	3.961	0.545	1	4.059	0.571	2	4.164	0.630
3	1	3.902	0.542	1	3.948	0.544	1	4.041	0.566	2	4.144	0.623
4	1	3.897	0.543	1	3.938	0.542	1	4.026	0.562	2	4.128	0.617
6	1	3.887	0.544	1	3.921	0.543	1	4.002	0.559	2	4.100	0.607
10	1	3.871	0.546	1	3.899	0.545	1	3.968	0.552	*1*	*4.059*	*0.589*
12	1	3.866	0.547	1	3.891	0.547	1	3.956	0.555	*1*	*4.045*	*0.589*
15	1	3.857	0.548	1	3.881	0.548	1	3.942	0.553	*1*	*4.026*	*0.582*
20	*1*	*3.843*	*0.550*	*1*	*3.864*	*0.548*	1	3.921	0.549	1	4.001	0.570
30	*1*	*3.824*	*0.556*	*1*	*3.842*	*0.551*	*2*	*3.890*	*0.558*	1	3.960	0.556
40	*1*	*3.808*	*0.559*	*1*	*3.820*	*0.555*	*1*	*3.861*	*0.550*	1	3.923	0.553
50	*1*	*3.791*	*0.563*	*1*	*3.804*	*0.557*	*1*	*3.841*	*0.550*	*1*	*3.898*	*0.550*
60	*1*	*3.776*	*0.565*	*1*	*3.787*	*0.561*	*1*	*3.821*	*0.555*	*1*	*3.874*	*0.552*
RCV_t^{Δ}												
2	*1*	*3.866*	*0.268*	*1*	*3.867*	*0.296*	*1*	*3.868*	*0.393*	*2*	*3.942*	*0.556*
3	*1*	*3.799*	*0.306*	1	3.798	0.331	*1*	*3.798*	*0.418*	*2*	*3.903*	*0.560*
4	*1*	*3.733*	*0.353*	1	3.732	0.374	*2*	*3.866*	*0.449*	*2*	*3.869*	*0.568*
6	*2*	*3.797*	*0.388*	2	3.798	0.411	*2*	*3.798*	*0.476*	*2*	*3.798*	*0.590*
10	*3*	*3.776*	*0.457*	*3*	*3.774*	*0.477*	3	3.772	0.532	3	0.583	0.626
12	*3*	*3.731*	*0.482*	*3*	*3.732*	*0.499*	3	3.732	0.551	3	3.733	0.642
15	*3*	*3.666*	*0.521*	*3*	*3.664*	*0.536*	3	3.663	0.585	4	3.747	0.670
20	4	3.657	0.576	4	3.656	0.589	4	3.657	0.631	4	3.657	0.704
30	6	3.661	0.657	6	3.662	0.666	6	3.663	0.701	6	3.665	0.762
40	6	3.548	0.720	6	3.550	0.728	6	3.553	0.759	6	3.556	0.815
50	8	3.578	0.787	8	3.581	0.796	8	3.585	0.826	8	3.588	0.877
60	8	3.491	0.835	8	3.495	0.843	8	3.499	0.870	8	3.503	0.918
$RCV_{C,t}^{\Delta}$												
2	1	3.981	1.208	1	3.978	1.212	1	3.976	1.237	1	3.974	1.301
3	1	3.981	1.208	1	3.979	1.212	1	3.978	1.236	1	3.976	1.298
4	1	3.982	1.208	1	3.980	1.212	1	3.979	1.235	1	3.978	1.296
6	1	3.981	1.209	1	3.978	1.213	1	3.976	1.239	1	3.973	1.305
10	1	3.980	1.209	1	3.976	1.210	1	3.972	1.232	1	3.968	1.292
12	1	3.979	1.210	1	3.976	1.212	1	3.973	1.237	2	3.970	1.301
15	1	3.979	1.211	1	3.976	1.216	1	3.974	1.241	1	3.973	1.305
20	1	3.977	1.208	1	3.975	1.213	1	3.973	1.239	1	3.972	1.300
30	1	3.975	1.206	1	3.972	1.211	1	3.969	1.234	1	3.966	1.288
40	1	3.977	1.213	1	3.975	1.218	1	3.974	1.238	1	3.972	1.283
50	1	3.976	1.216	1	3.974	1.221	1	3.973	1.239	1	3.971	1.277
60	1	3.974	1.217	1	3.972	1.222	1	3.970	1.238	1	3.969	1.272

Note: The table provides results for the simulation experiment with infrequent and non-synchronous trading and bid-ask bounce for different trading frequencies τ, ranging from 2 to 60 seconds, and different bid-ask spreads s, ranging from 0 to 0.00075. See Table 3.2 for details of the simulation set-up. For each combination of τ and s, the table shows the optimal sampling frequency for the four covariance estimators, along with the corresponding RMSE and mean. The covariance estimator that achieves the lowest RMSE for a given combination of τ and s is shown in italics.

fore expect that (close to) offsetting biases will arise in actual data, and this will favor the realized co-range.

Third, when trading occurs relatively frequently, such that the upward bias dominates, the realized covariance is preferred. On the other hand, when trading occurs relatively infrequently, such that the downward bias dominates, the adjusted realized co-range renders the lowest RMSE values. Note that for the largest spread considered here, $s = 0.00075$, the adjusted realized co-range in fact also outperforms the realized covariance for 'frequent' trading at once per 10-15 seconds. Unreported results for even higher values of s (0.001) show that the bias-corrected realized co-range in that case dominates for all trading frequencies.

Fourth, Table 3.3 shows that for the bias-corrected realized covariance and realized co-range, apart from a few exceptions, the highest possible sampling frequency leads to the lowest RMSE. This is not surprising given that sampling more frequently leads to a lower standard deviation of the covariance estimators, which in the absence of bias also implies a lower RMSE. For the unadjusted realized covariance, we observe that less frequent trading always leads to less frequent sampling to achieve the lowest possible RMSE. For the realized co-range this only holds in the absence of bid-ask bounce. When bid-ask bounce does occur next to infrequent and non-synchronous trading, it sometimes pays off to sample more frequently, such that the positive and negative biases approximately cancel.

Finally, as noted in the previous section, the number of trading days Q used to compute the correction factor for $RCR^{\Delta}_{C,t}$ in (3.9) is a crucial choice to be made. If the trading intensity and the spread are constant over time, Q may be set large in order to gain accuracy. On the other hand, when the magnitude of these microstructure frictions varies over time, only the recent price history should be used and Q should be set fairly small. Figure 3.1 shows the RMSE of the bias-corrected realized co-range for the experiment with infrequent trading as a function of Q. The RMSE monotonically declines as Q increases, but the largest gains occur up to $Q = 100$, beyond which the RMSE more or less stabilizes. Hence, our choice of $Q = 66$ does not seem unreasonable. Also note that the reduction in RMSE due to increasing Q is largest for higher sampling frequencies.

3.4 Empirical application

In this section we study the empirical usefulness of the realized co-range by evaluating its performance in a dynamic volatility timing strategy, adopting the framework developed by Fleming et al. (2001, 2003). We consider an investor who uses conditional mean-variance analysis for constructing a portfolio with minimum variance

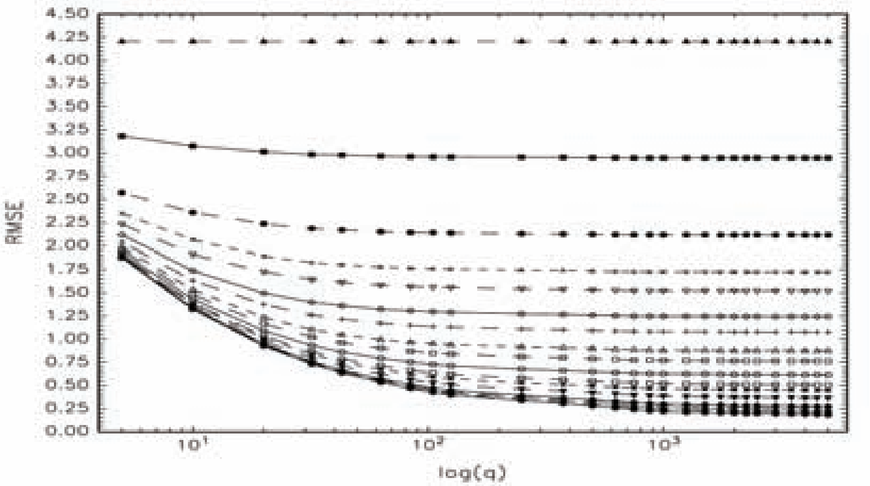


Figure 3.1: RMSE of the bias-corrected realized co-range $RCR^{\Delta}_{C,t}$ obtained from (3.9) as a function of Q, for sampling frequencies as listed in the first column of Table 3.2 augmented with 90, 120, 180, 240, 360 and 720 minutes. The sampling frequency increases from the top line to the bottom.

given a specific target return. The portfolio is dynamic in the sense that optimal weights are re-computed daily. The investor follows a volatility-timing strategy, as the portfolio weights are based on a forecast of the conditional covariance matrix while expected returns are held constant. We assess the merits of using the realized co-range to construct these forecasts, relative to the realized covariance. We also include their daily counterparts in the comparison.

3.4.1 Data

Following Fleming et al. (2001, 2003), we consider portfolios consisting of stocks, bonds and gold. We assume that the investor trades futures contracts to construct her portfolio, to avoid short-selling restrictions and to save on transaction costs. We obtain intraday transactions data for futures contracts on stocks (S&P500, Chicago Mercantile Exchange, with trading hours from 8:30 am - 3:15 pm), US T-bonds (Chicago Board of Trade, 8:00 am - 2:00 pm) and gold (New York Mercantile Exchange, 7:20 am - 1:30 pm[9]), for the period from January 3, 1984 to December 31, 2006.[10]

[9] All trading hours are standardized to the CST zone.

[10] The data is obtained from Tick Data, `http://www.tickdata.com/`.

We exclude all days on which any of the three markets is closed, leaving a total of 5,592 days of high-frequency data on which the three contracts traded simultaneously. Our sample period includes the October 19, 1987 stock market crash and September 11, 2001. The bid-ask spreads on the three futures contracts were much higher in the days following the October 1987 crash, and for this reason we follow Fleming et al. (2003) by replacing the high-frequency covariance estimators by their daily counterparts for the period October 19 to 30, 1987. We exclude the days following September 11, 2001 as markets were closed. The very large negative overnight return for the September 11-17 period is excluded as well.

At any given day we use prices of the nearby futures contract in each market, rolling to the second nearby contract when the nearby contract enters its final month for gold and bonds and on the 11th trading day in the final month for stocks. We assume that the investor updates her portfolio daily at 1:30 pm, motivated by the trading hours for gold futures. For the daily return series that is used to evaluate the investment strategy, we use the last transaction prices occurring before that time. In case a contract is rolled forward on day $t+1$ we use the price at 1:30 pm on day t of the 'new' contract to compute the daily return for day $t+1$.

We adopt the popular five-minute frequency for computing the realized (co-)ranges and realized (co-)variances. The five-minute returns that are used in the latter estimator are obtained from the last transaction prices in each intraday interval. For the variance estimates for day t based on the realized variance and realized range we use all intraday returns and prices, respectively, between 1:30 pm on day $t-1$ and 1:30 pm on day t. For the covariance estimates, we adopt the following two-step procedure of Fleming et al. (2003). First, we construct estimates of volatilities and covariances with intraday data for the common trading hours from 8:30 am - 1:30 pm, from which we back out estimates of the correlations among the three assets. Second, we convert these correlations back into covariances by using the realized variances and realized ranges for the complete trading day.

The high-frequency data only covers the part of the day during which futures markets are open. Fleming et al. (2003) and De Pooter et al. (2008) add the cross-product of overnight returns to the realized covariance estimate in order to obtain a measure of the covariation during a complete 24-hour day. Both studies find that incorporating overnight returns adds information and improves the performance of the volatility timing strategy. We choose not to include the overnight returns, however, as adding the same overnight returns to the realized co-range and realized covariance would diminish the difference between these estimators and, presumably, any differences in their overall performance. Since both estimators aim to estimate the integrated covariation they are already expected to behave similarly to a large extent.

Table 3.4: Summary statistics

Panel A: Mean, standard deviation, skewness and kurtosis

	Stocks	Bonds	Gold
Mean	0.072	0.044	−0.011
St.Dev.	0.166	0.104	0.150
Skewness	−0.720	0.464	−0.229
Kurtosis	17.417	16.290	10.182

Panel B: Correlations

	Stocks	Bonds	Gold
Stocks	1		
Bonds	0.101	1	
Gold	−0.107	−0.070	1

Panel C: Realized variance and realized range

	$\sqrt{RV_s}$	$\sqrt{RR_s}$	$\sqrt{RV_b}$	$\sqrt{RR_b}$	$\sqrt{RV_g}$	$\sqrt{RR_g}$
Mean	0.104	0.069	0.078	0.053	0.109	0.072
St.Dev.	0.053	0.032	0.032	0.018	0.055	0.033
Skewness	2.658	2.519	1.912	1.784	2.207	2.059
Kurtosis	16.818	15.161	9.667	9.692	13.059	12.168

Panel D: Realized correlations

	$\rho_{s,b}(RCV)$	$\rho_{s,b}(RCR)$	$\rho_{g,s}(RCV)$	$\rho_{g,s}(RCR)$	$\rho_{g,b}(RCV)$	$\rho_{g,b}(RCR)$
Mean	0.112	0.128	−0.056	0.003	0.006	0.070
St.Dev.	0.388	0.314	0.179	0.128	0.183	0.122
Skewness	−0.501	−0.593	−0.445	−0.513	0.136	0.033
Kurtosis	2.229	2.568	3.555	3.878	3.461	3.597

Note: This table summarizes the data statistics of the daily returns, realized volatility and realized range estimators and the correlations implied by the intraday estimators for the stocks (s), bond (b), and gold (g) futures. $\sqrt{RV_x}$ represents the annualized realized volatility sampled at the 5-minute frequency and $\sqrt{RR_x}$ is the realized range. The realized correlations implied by the realized covariance are denoted $\rho_{x,y}(RCV)$ and $\rho_{x,y}(RCR)$ is the correlation implied by the realized co-range.

Table 3.4 displays summary statistics for the annualized daily returns and the high-frequency volatility and correlation estimates. For all three assets we find that the mean of the realized range is smaller than the mean of the realized variance, suggesting that infrequent trading is more important than bid-ask bounce for these futures contracts, such that the realized range is biased downward. As expected, the standard deviation of the realized range turns out to be substantially smaller than that of the realized variance. This carries over to the multivariate case, where the standard deviation of the correlations estimated by means of the realized co-range is smaller than the standard deviation of the correlations implied by the realized covariance. This indicates that the realized co-range is less noisy than the realized covariance. Graphical support for this result is provided in Figure 3.2. Although both estimators render similar correlation patterns, the daily correlations based on the realized covariance are more volatile and show much more spikes than the correlations based on the realized co-range.

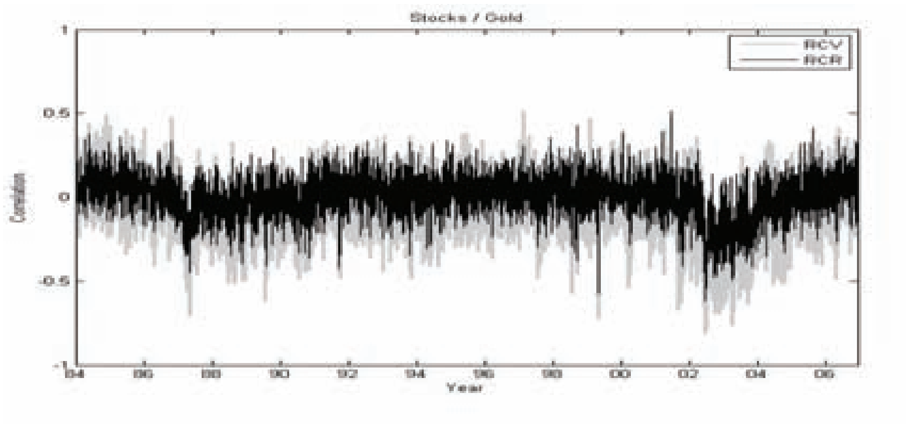
3.4.2 Volatility timing

The investor uses conditional mean-variance analysis for forming a portfolio with minimum variance given a specific target return. The portfolio weights w_t for day t follow from solving the standard quadratic programming problem, where the portfolio variance $\sigma_p^2 = w_t' \Sigma_t w_t$ is minimized subject to a target portfolio return $\mu_p = w_t' \mu_t$, with μ_t the (3×1) vector of conditional expected returns for the stocks, bonds and gold futures and Σ_t the (3×3) conditional covariance matrix. The resulting minimum-variance weights are given by

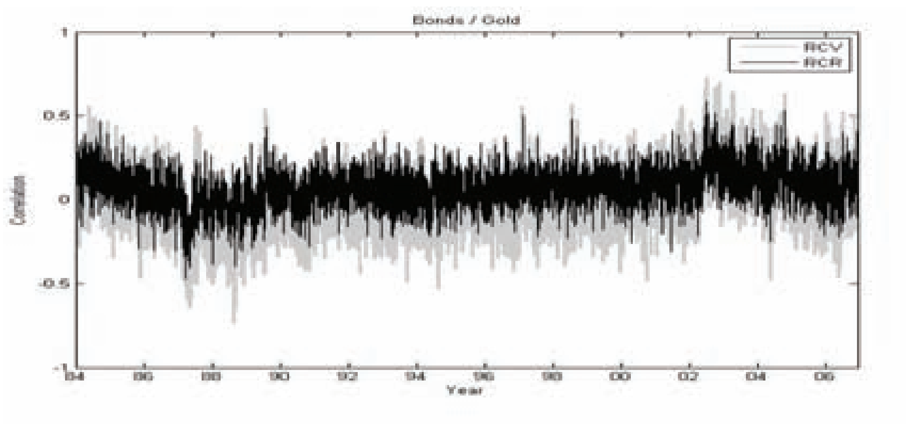
$$w_t = \frac{\mu_p \Sigma_t^{-1} \mu_t}{\mu_t' \Sigma_t^{-1} \mu_t}. \tag{3.11}$$

In general, these unrestricted portfolio weights do not add up to one. We include a risk-free asset (cash) with weight $1 - w_t' \iota$, where ι is a vector of ones, which makes the portfolio fully invested. Given that we want to focus on differences in the investment strategy's performance through the dynamics of Σ_t, we keep the expected returns constant by setting $\mu_t = \mu$, where μ is taken to be the vector of full-sample mean returns, see line 1 of Table 3.4. Again following Fleming et al. (2003), we set the annualized target portfolio return $\mu_p = 10\%$ and assume a constant risk-free rate $r_f = 6\%$.

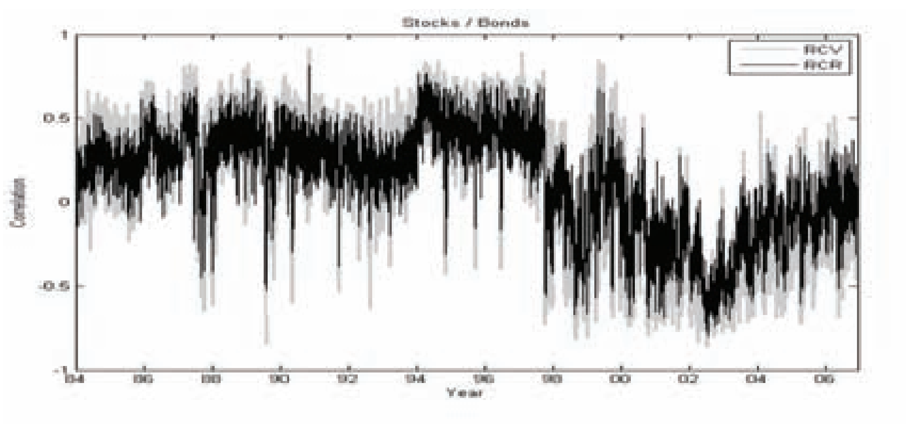
Implementing the volatility timing strategy requires an estimate of the conditional covariance matrix Σ_t in (3.11). Following Fleming et al. (2003), we use backward-looking 'rolling' estimators using an exponential weighting scheme, motivated by the



(a) Gold and stocks



(b) Gold and bonds



(c) Stocks and bonds

Figure 3.2: This figure illustrates plots of the realized correlations between gold and stocks (a), gold and bonds (b) and stocks and bonds (c). The correlations

work of Foster and Nelson (1996) and Andreou and Ghysels (2002). The general expression for the rolling daily conditional covariance estimator for day t is given by

$$\widehat{\Sigma}_t = \exp(-\alpha)\widehat{\Sigma}_{t-1} + \alpha \exp(-\alpha) V_{t-1}, \tag{3.12}$$

where α is the decay parameter and V_{t-1} is an estimate of the realized covariance matrix on day $t-1$. We expect that the decay parameter decreases with the level of noise in V_{t-1}. Our main interest lies in the performance of the volatility timing strategy when using the realized (co-)range in (3.2) and (3.5) to construct V_{t-1}, compared to using the realized (co-)variance in (3.1) and (3.3). In order to reduce the effects of microstructure noise, we follow Fleming et al. (2003) and employ the bias-corrected versions of these estimators, as discussed in Section 3.2.2. For the additive bias-correction we set $Q = 22$, that is, we use a shorter history compared to the simulations in Section 3.3. This is motivated by the fact that the rolling estimator in (3.12) 'smooths' the realized estimators, which also reduces the effects of noise to a certain extent. To gauge the benefits of using intraday data, we also include estimators of V_{t-1} based on daily (close-to-open) returns and daily (co-)ranges.

The decay parameter α in (3.12) is estimated by maximizing the log-likelihood function corresponding with the model

$$r_t = \mu + \widehat{\Sigma}_t^{\frac{1}{2}} z_t \tag{3.13}$$

where μ is the vector of daily expected returns, $\widehat{\Sigma}_t$ is the conditional covariance matrix obtained from (3.12), and $z_t \sim NID(0, I)$. The likelihood function is maximized using the complete sample period, as in Fleming et al. (2003), although we use the first year of our sample as a burn-in period for the (co)variance dynamics.[11] The maximum-likelihood procedure results in a decay parameter of 0.082 for the daily co-range and 0.0298 for the daily covariance. For the intraday based covariation estimators we find decay parameters equal to 0.064 and 0.062 for the bias-corrected realized co-range and realized covariance, respectively. The daily co-range has a much larger decay parameter than the daily covariance estimator, due to the fact that the co-range is less noisy.

The daily portfolio weights (3.11) are based on the one day ahead forecast of the bias-corrected rolling covariance matrix estimator in (3.12). The daily realized returns of the portfolio are obtained as $w_t' r_t$, where r_t is the vector of daily returns. Below we report the annualized average portfolio returns, volatility and the Sharpe Ratio (SR). Furthermore we also keep track of the turnover of the portfolio using

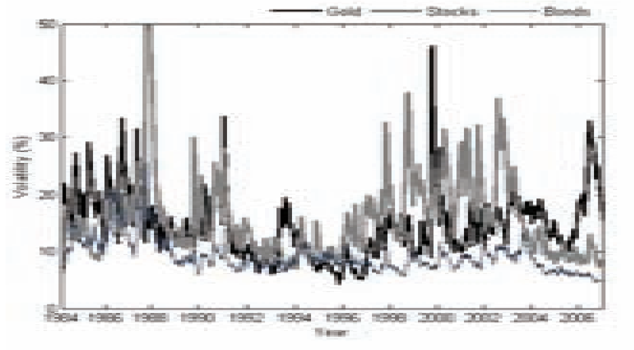
[11] This burn-in period is excluded from all performance evaluations below.

$TO_t = |w_t - w_{t-1}|'\iota$ to provide insight into the transaction costs arising from daily rebalancing the portfolio. In addition, we compute the break-even costs, that is, the level of transaction costs that would reduce the profitability of the investment strategy to zero.[12]

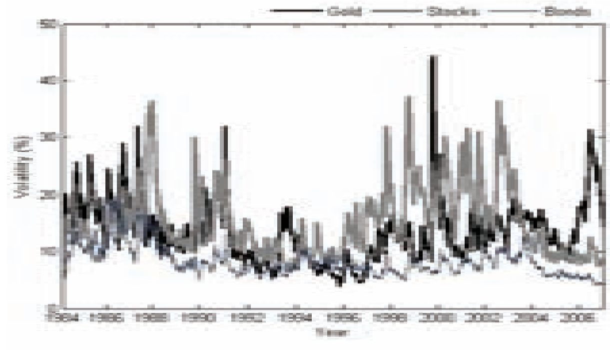
3.4.3 Empirical results

The rolling estimators based on the realized range and realized variance estimators, as well as their daily counterparts, result in similar volatility dynamics, as visualized in Figure 3.3. The main difference appears to be that the range-based estimators in panels (a) and (c) show higher levels of stock volatility during periods of turmoil such as October 1987 and the Russia crisis in 1998.

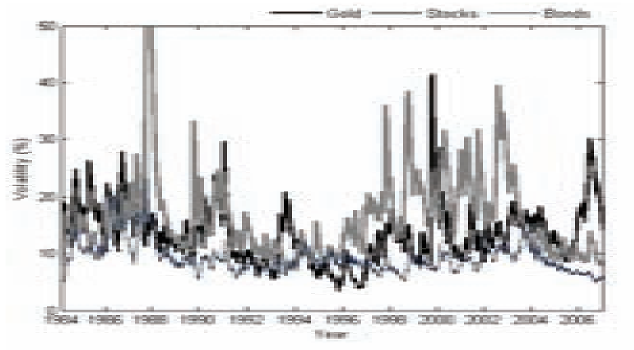
[12]Here we deviate from Fleming et al. (2003) who use a quadratic utility function to assess the economic value of volatility timing using realized covariances compared to daily covariances. The reason we do not use such utility function is because it is based on the mean and volatility of the portfolio returns. Tables 3.5 and 3.6 illustrate that the mean portfolio return and the volatility of the portfolio are very similar for the realized covariance and realized co-range. However, the levels of noise in the correlations implied by the realized covariance and realized co-range are different, see Figure 3.4 and Table 3.4, and for this reason there exists a substantial difference in turnover in both portfolio weights, see Table 3.5. Therefore we choose to compare both estimators in terms of break-even transaction costs.



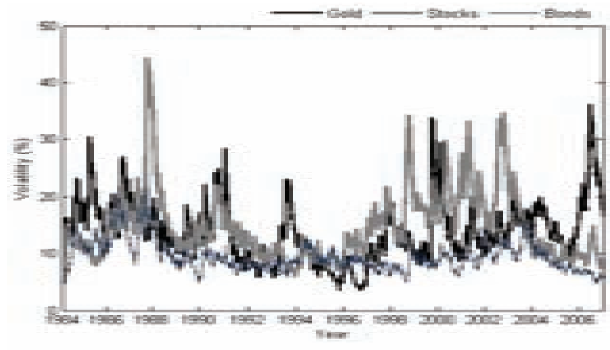
(a) Realized co-range



(b) Realized covariance



(c) Daily co-range



(d) Daily covariance

Figure 3.3: This figure illustrates plots of the volatility estimates obtained from the rolling covariance estimator Σ_t in (3.12) with inputs for V_{t-1} being the realized co-range (a), realized covariance (b), daily co-range (c) and the daily covariance estimator (d).

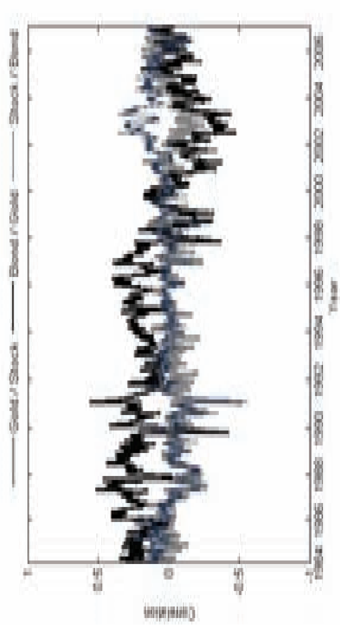
Larger differences between range-based and return-based rolling estimators are found in the behavior of the correlations, shown in Figure 3.4. Although the general correlation patterns appear to be similar, the correlations based on the realized covariance or daily covariance are larger in absolute value and tend to fluctuate more widely than the range-based correlations, in agreement with the differences in realized correlations shown in Figure 3.2. The relative stability of the range-based correlations may in fact be an advantage. Not only is an investor likely to be reluctant to base her investment decisions on very volatile correlation estimators, a more stable correlation estimate will also result in less day-to-day fluctuations in the portfolio weights, and hence lower transaction costs.

The results in Table 3.5 illustrate that over the whole sample (1985-2006) the performance of the volatility timing strategy based on the bias-corrected rolling realized co-range and realized covariance estimators is very close. The realized co-range earns a slightly higher average return of 9.4% compared to 9.3% for the realized covariance. As the portfolio volatilities are the same at 7.2% annually, this results in Sharpe ratios that are almost identical and equal to 1.31 and 1.30, respectively.

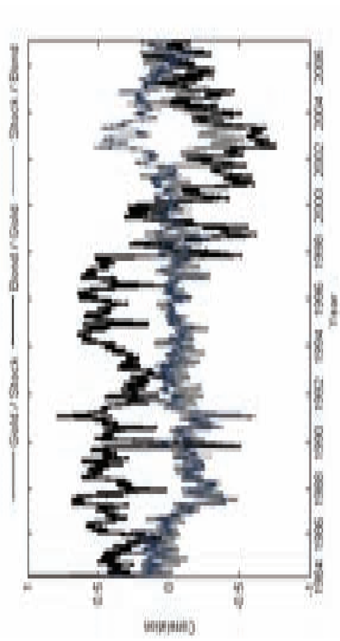
Although the strategies based on the realized co-range and realized covariance result in a similar performance before transaction costs, the dynamics of the underlying portfolios are quite different. The portfolio weights displayed in Figure 3.5 demonstrate that the realized co-range yields weights that are much less volatile than the realized covariance. The turnover generated by the realized co-range portfolio is equal to 5.9, almost 40% smaller than that of the realized covariance at 9.5. This results in break-even transaction cost levels of 159.8 and 98.4 basis points for the realized co-range and for the realized covariance, respectively. Hence, the realized co-range outperforms the realized covariance substantially by more than 60 basis points.

Table 3.5 illustrates that the main results continue to hold across three-year subperiods. The realized co-range and realized covariance provide similar risk-return characteristics and therefore identical Sharpe ratios. The turnover in the realized co-range portfolio weights is substantially smaller than that of the realized covariance in each of the subsamples considered. This results in a better performance of the realized co-range in terms of break-even transaction costs.

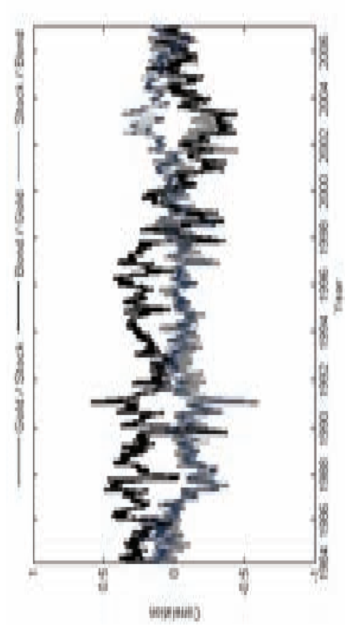
Consistent with other high-frequency data studies, we find that the use of intraday data leads to better estimates than daily data. This holds for both return- and range-based estimators. When we compare the performance of the daily covariance with the realized covariance using 5-minute intervals we see that more precise (co)variance estimates increase the average return by 20 basis points from 9.1% to 9.3%. The risk of the portfolio, as measured by the volatility of the portfolio returns, decreases from



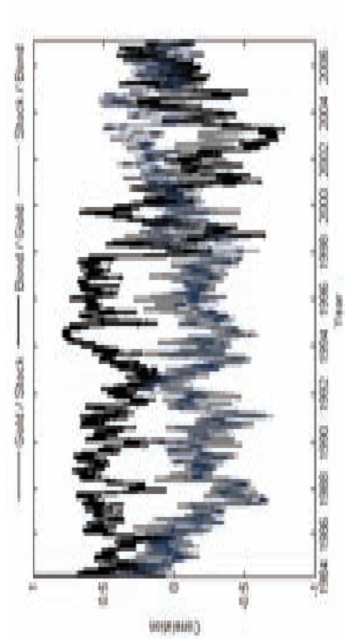
(a) Realized co-range



(b) Realized covariance



(c) Daily co-range



(d) Daily covariance

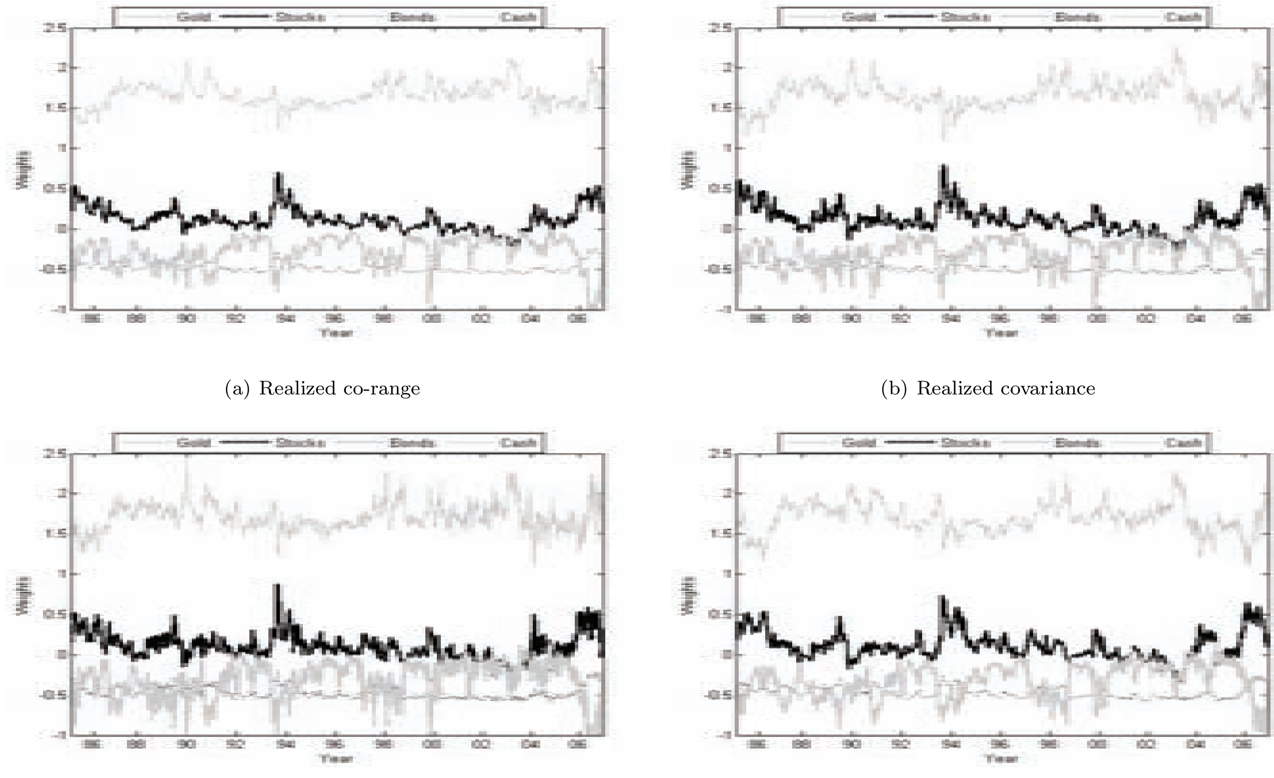
Figure 3.4: This figure illustrates plots of the daily correlations implied by the rolling covariance estimator Σ_t in (3.12) with inputs for V_{t-1} being the realized co-range (a), realized covariance (b), daily co-range (c) and the daily covariance estimator (d).

Table 3.5: Volatility timing strategy

	Realized Co-Range (5 min)					Realized Covariance (5 min)				
	Mean	St.Dev.	SR	TO	BETC	Mean	St.Dev.	SR	TO	BETC
1985-2006	0.094	0.072	1.306	5.9	159.8	0.093	0.072	1.291	9.5	98.4
1985	0.000	0.061	-0.007	2.9	-1.4	0.004	0.061	0.058	4.6	7.8
1986-1988	0.110	0.078	1.406	5.5	199.8	0.110	0.079	1.381	9.1	121.0
1989-1991	0.132	0.070	1.879	6.4	207.4	0.130	0.070	1.856	9.6	135.1
1992-1994	0.092	0.053	1.746	5.7	161.4	0.087	0.053	1.642	9.2	94.3
1995-1997	0.161	0.047	3.412	4.5	360.6	0.161	0.047	3.403	7.3	221.4
1998-2000	0.132	0.073	1.822	5.7	233.2	0.130	0.073	1.787	8.8	147.8
2001-2003	0.014	0.085	0.169	5.3	27.1	0.017	0.085	0.203	8.9	19.3
2004-2006	0.038	0.084	0.446	8.3	45.5	0.036	0.084	0.432	13.6	26.8

	Daily Co-Range					Daily Covariance				
	Mean	St.Dev.	SR	TO	BETC	Mean	St.Dev.	SR	TO	BETC
1985-2006	0.088	0.072	1.224	17.1	51.6	0.091	0.074	1.234	10.2	89.0
1985	-0.006	0.060	-0.092	8.3	-6.7	-0.001	0.064	-0.020	5.0	-2.6
1986-1988	0.107	0.079	1.362	14.6	73.7	0.123	0.080	1.534	8.8	139.3
1989-1991	0.120	0.069	1.742	17.2	69.5	0.119	0.070	1.695	9.8	121.4
1992-1994	0.087	0.052	1.658	14.6	59.4	0.096	0.054	1.779	9.3	102.6
1995-1997	0.156	0.047	3.300	13.1	118.9	0.153	0.049	3.150	8.0	190.7
1998-2000	0.121	0.073	1.657	15.4	78.9	0.120	0.075	1.598	9.3	128.4
2001-2003	0.021	0.085	0.242	20.6	10.0	0.020	0.086	0.230	11.5	17.2
2004-2006	0.028	0.085	0.333	24.5	11.6	0.026	0.086	0.303	14.7	17.7

Note: This table summarizes the performance statistics of the volatility timing strategy based on the bias-corrected rolling estimators constructed from 5 minute return data (realized covariance) and 5 minute range data (realized co-range). The decay parameter α is estimated in a maximum-likelihood procedure. The annualized average daily portfolio return is denoted by Mean and St.Dev. represents the corresponding annualized volatility of daily portfolio returns. TO is the annualized average absolute daily change in portfolio weights. BETC represents the annualized break-even transaction costs (in basis points), which would reduce the profitability of the investment strategy to zero.



(a) Realized co-range

(b) Realized covariance

(c) Daily co-range

(d) Daily covariance

Figure 3.5: This figure illustrates the daily minimum-variance portfolio weights of a dynamic volatility timing strategy. The portfolio weights follow from 3.11 with Σ_t obtained from (3.12) with inputs for V_{t-1} being the realized co-range (a), realized covariance (b), daily co-range (c) and the daily covariance estimator (d).

7.4% to 7.2%, which is equal to the risk of the daily co-range portfolio. The turnover decreases from 10.2 to 9.5, leading to slightly higher break-even transaction costs, an increase of almost 10 basis points per annum.

For the co-range the use of intraday data is even more profitable.[13] The daily co-range achieves an average portfolio return of 8.8%, compared to 9.4% for the realized co-range. The return difference of about 60 basis points comes at the same level of portfolio risk, such that the Sharpe ratio increases from 1.23 to 1.31 when switching from the daily co-range to the realized co-range. The most notable difference occurs in terms of turnover, which is substantially larger for the daily co-range (17.1) than for the realized co-range. This leads to much higher transaction costs for the daily co-range, resulting in a large difference in break-even transaction costs of more than 100 basis points (51.6 compared to 159.8).

Concluding, the use of the realized co-range in a volatility timing investment strategy has appealing economic advantages, especially when transaction costs are taken into account. The more precise covariance estimates that are obtained with high-frequency intraday ranges result in a more stable portfolio with smaller turnover and, thus, higher break-even transaction costs.

3.4.4 Sensitivity analysis for the decay parameter

The empirical results described above are based on decay parameters α in (3.12) that maximize the log-likelihood function of the GARCH-type model in (3.13). For several reasons it is interesting to examine the sensitivity of the volatility timing results for the choice of α. First, the results of Fleming et al. (2003) and De Pooter et al. (2008) illustrate that the decay parameters that maximize the statistical fit do not necessarily provide the best economic performance in a volatility timing strategy. Second, in the forecasts obtained from (3.12), the decay rate α determines the weight put on the 'backward looking' rolling estimator, which equals $\exp(-\alpha)$, and the weight put on the innovation term, $\alpha \exp(-\alpha)$. When α is small this indicates that the estimator requires a high degree of 'smoothing' because it is noisy, whereas a large α indicates less smoothing. Hence when we use a small decay parameter for the realized covariance and realized co-range, they are expected to provide similar results as the noise is 'smoothed' out to a large extent.[14] The maximum-likelihood estimates for the decay parameters are relatively small, which implies a high degree of

[13]This result is corroborated by Table 3.6, which summarizes the results for a range of decay parameters. The higher profitability of the realized co-range is not caused by the fact that it has a smaller decay parameter than the daily co-range. When we fix the decay parameter to be equal, the realized co-range also outperforms the daily co-range.

[14]We thank Torben Andersen for providing this insight.

smoothing and therefore small differences in performance. Pinning down α at larger values for both estimators is a good test case as it allows us to assess the quality of the innovation term V_{t-1} in (3.12). Doing so we expect to find larger differences in the average returns and levels of risk for the volatility-timing portfolios based on the realized covariance or realized co-range. Third, we expect the decay parameter to have an important impact on turnover as a large decay parameter implies a large weight on the most recent innovation and therefore more turnover in the portfolio weights. Fixing the decay rates enables a fairer comparison in terms of the break-even transaction costs.

Table 3.6 summarizes the empirical results for the competing rolling variance-covariance estimators over the 1985-2006 period for a grid of decay parameters ranging from 0.01 to 0.175. Before taking into account turnover and transaction costs the performance of the realized co-range and realized covariance is similar for small and intermediate decay rates ($\alpha = 0.01$ to 0.10). For example, using a decay rate of 0.05 results in identical levels of risk (7.2%) and return (9.4%) and therefore approximately equal Sharpe ratios (1.302 and 1.308).

However, for larger decay rates with more weight being put on the most recent estimate, the performance of the realized covariance deteriorates while the performance of the realized co-range remains robust. For example, when we use a decay rate of 0.15 the Sharpe ratio of the realized covariance decreases to 1.15 while the realized co-range still achieves a Sharpe ratio equal to 1.29. At the daily sampling frequency the co-range also seems more robust to the choice of decay parameter than the daily covariance estimator. Although the performance of the daily co-range worsens for large decay rates, the decline in performance is substantially less than that of the daily covariance. Before taking into account turnover and transaction costs, the daily co-range outperforms the daily covariance for the whole range of decay rates, and it competes with the realized covariance when using decay rates larger than or equal to 0.15. The realized co-range outperforms the daily co-range regardless of the decay rate. The difference is more pronounced for large decay parameters. The risk-return characteristics of the realized co-range are substantially better as expressed by the difference in Sharpe ratio, which increases with the weight put on the most recent covariance estimate.

The portfolios based on the realized co-range exhibit substantially lower turnover than the portfolios based on the realized covariance, with the difference ranging from 33% for a decay parameter of 0.01 up to 62% for a decay parameter of 0.15. Using the realized covariance instead of the daily covariance reduces turnover by about 35% for large decay rates and up to 55% for small and intermediate decay parameters. A similar comparison of the realized co-range with its daily counterpart shows a

Table 3.6: Volatility timing strategy: Sensitivity analysis for decay rate

	Realized Co-Range (5 min)					Realized Covariance (5 min)				
α	Mean	St.Dev.	SR	TO	BETC	Mean	St.Dev.	SR	TO	BETC
0.010	0.095	0.074	1.279	1.1	846.0	0.095	0.074	1.286	1.7	567.8
0.025	0.095	0.073	1.304	2.6	369.2	0.095	0.073	1.310	4.0	236.2
0.050	0.094	0.072	1.308	4.8	198.1	0.094	0.072	1.302	7.8	121.3
0.075	0.094	0.072	1.303	6.8	137.4	0.093	0.072	1.284	11.4	81.1
0.100	0.093	0.072	1.297	8.8	105.7	0.091	0.072	1.264	15.1	60.5
0.125	0.093	0.072	1.292	10.8	86.3	0.090	0.072	1.243	18.9	47.7
0.150	0.093	0.072	1.288	12.7	73.2	0.096	0.083	1.153	33.7	28.6
0.175	0.093	0.072	1.285	14.5	63.7	0.086	0.101	0.854	37.4	23.0

	Daily Co-Range					Daily Covariance				
α	Mean	St.Dev.	SR	TO	BETC	Mean	St.Dev.	SR	TO	BETC
0.010	0.093	0.074	1.263	2.2	431.5	0.093	0.074	1.257	3.4	273.0
0.025	0.092	0.073	1.271	5.4	172.1	0.091	0.073	1.245	8.5	107.0
0.050	0.090	0.072	1.251	10.6	85.2	0.088	0.074	1.181	17.1	51.2
0.075	0.089	0.072	1.229	15.7	56.4	0.084	0.075	1.120	25.6	32.8
0.100	0.087	0.072	1.211	20.8	42.1	0.081	0.076	1.064	34.2	23.7
0.125	0.086	0.072	1.195	25.7	33.6	0.079	0.077	1.014	42.8	18.4
0.150	0.086	0.072	1.182	30.6	28.0	0.076	0.079	0.969	51.3	14.9
0.175	0.085	0.073	1.170	35.4	24.0	0.074	0.080	0.929	59.8	12.4

Note: This table displays the performance statistics of the sensitivity analysis for decay rates in the volatility timing strategy based on the bias-corrected rolling estimators constructed from 5 minute return data (realized covariance) and 5 minute range data (realized co-range). α is the decay parameter. The annualized average daily portfolio return is denoted by Mean and St.Dev. represents the corresponding annualized volatility of daily portfolio returns. TO is the annualized average absolute daily change in portfolio weights. BETC represents the annualized break-even transaction costs (in basis points), which would reduce the profitability of the investment strategy to zero.

reduction in turnover of about 50% to almost 60%, for small and large decay rates, respectively.

From the sensitivity analysis we conclude that the realized co-range is more robust to the choice of decay parameter α in (3.12) than the realized covariance, the daily covariance and the daily co-range. The smaller turnover and better break-even transaction costs of the realized co-range do not depend on the decay rate but they are caused by more precise estimates of the covariance matrix. When pinning down the decay rate, the realized co-range never has a worse performance than the realized covariance before measuring turnover and costs. If we do take into account turnover and compute break-even transaction costs, the realized co-range outperforms the realized covariance and the two daily estimators. When the weight on recent estimates is large, the realized co-range also achieves a better performance than the realized covariance in terms of risk and returns and therefore results in higher Sharpe ratios in the volatility timing strategy.

3.5 Conclusion

We have extended range-based covariance estimation with a novel high-low range estimator based on intraday data. The covariance between two assets is backed out from their individual variances and the variance of a portfolio of the two assets, where the realized range is used to estimate each of these variances. In case of continuous trading and no market frictions, the realized co-range provides a considerably more accurate measure of covariation than the realized covariance, which uses cross-products of intraday returns, due to the relative efficiency of range-based volatility measures when using the same sampling frequency.

The realized co-range continues to have attractive properties in the presence of market microstructure noise due to bid-ask bounce, infrequent trading and non-synchronous trading. A key advantage of the co-range is that the upward bias due to bid-ask bounce and the downward bias due to infrequent and non-synchronous trading partially offset each other. Although the realized covariance is also downward biased due to non-synchronous trading, it is not biased due to bid-ask bounce. In simulation experiments we indeed find that the realized co-range improves upon the realized covariance for empirically plausible levels of bid-ask bounce and non-synchronous trading. In case the different biases do not offset each other, bias-correcting the realized co-range with the recent historical average (relative) level of the daily co-range is an effective procedure to restore the efficiency of the realized co-range.

In the empirical study for S&P500, bond and gold futures, we find that in Fleming *et al.*'s (2003) volatility timing strategy the realized co-range and realized covariance provide similar results before taking into account transaction costs. The level of noise in the correlations implied by the realized co-range is substantially smaller than that of the realized covariance, leading to smaller variation in the portfolio weights whilst still providing a similar risk-return profile. After taking into account transaction costs, the realized co-range outperforms the realized covariance by about 60 basis points per annum when the decay parameters in the exponentially weighted rolling estimators are estimated by maximum-likelihood. The reason behind the similar risk-return profile is that for forecasting covariances both the realized co-range and the realized covariance use small decay parameters putting considerable weight on older covariance estimates to smooth noise. If larger decay parameters are used, putting more emphasis on the covariance estimates of the most recent day, the risk-return profile of the co-range becomes superior.

Several interesting directions for future research emerge from our study. First, it will be interesting to study the theoretical properties of the co-range. Second, alternative estimators based on intraday highs and lows could be explored, such as the generalized range estimator of Dobrev (2007). Third, an empirical application to individual stocks could be considered, to examine the properties of the realized co-range for assets exhibiting more noise, and for higher dimensional problems. Fourth, in the context of such large-scale covariance matrices, factor models may be considered to alleviate the adverse effects of non-synchronous trading. Fifth and finally, the impact of jumps on the realized co-range should be investigated. The impact of jumps on the realized range has been examined by Christensen and Podolskij (2006), who consider a bipower range estimator as a jump-robust alternative to the 'standard' realized range. Also, Dobrev's (2007) generalized range estimator is robust to jumps. In case of covariance estimation, a crucial question is whether the jumps are idiosyncratic or common across assets. Bollerslev et al. (2008) provide empirical evidence for US stocks that both types of jumps are relevant. In case of fully idiosyncratic jumps, jump-robust range estimators may be used to estimate the covariance due to the continuous part of the price processes. In case of correlated (or partially common) jumps, 'standard' range-based estimators may be used to estimate the 'total' covariance. A key issue then may also be to separately identify the covariances due to the continuous and jump parts of the price processes.

Chapter 4

Realized Mixed-Frequency Factor Models for Vast-Dimensional Covariance Estimation*

4.1 Introduction

Accurate measures and forecasts of asset return covariances are important for financial risk management and portfolio management. Recent academic research in these areas has focused on two different issues. First, intraday data has been shown to render more precise measures and forecasts of daily asset return volatilities and covariances. Second, for the practically relevant case of portfolios consisting of a large number of assets, factor structures have been found useful to tackle the "curse of dimensionality". In this paper we put forward a novel approach for accurate measurement and forecasting of the covariance matrix of vast dimensional portfolios by combining the use of high *and* low-frequency data with a linear factor structure. Specifically, we introduce a "mixed-frequency" factor model (MFFM), where high-frequency data on relatively liquid factors is used for precise estimation of the factor covariance matrix and idiosyncratic risk whereas the factor loadings are estimated from low-frequency data.

*This chapter is based on the article by Bannouh, Martens, Oomen and Van Dijk (2012).

In recent years, a substantial body of literature has emerged on the use of high-frequency data for obtaining more accurate measures and forecasts of financial risk, see e.g. Andersen et al. (2006a) and McAleer and Medeiros (2008) for recent reviews. For the multivariate case, Barndorff-Nielsen and Shephard (2004) introduced the realized covariance, summing the cross-products of intraday returns. Market microstructure, however, poses two challenges: First, transactions take place against bid and ask prices, causing overestimation of the volatility. Second, non-synchronous trading of stocks biases covariance estimates towards zero. Several covariance estimators have been proposed that are robust to microstructure frictions. Focussing on the bi-variate case, Bandi and Russell (2005) illustrate how to choose the optimal sampling frequency for the realized covariance in the presence of microstructure noise. Hayashi and Yoshida (2005) propose an "all overlapping" returns estimator that is robust to non-synchronous trading. Griffin and Oomen (2011), Martens (2006), and Voev and Lunde (2007) provide further insights into the properties of the Hayashi-Yoshida and lead-lag adjusted realized covariance estimators in the presence of non-trading and microstructure noise. Zhang (2011) extends the two-scales estimator of Zhang et al. (2005) to covariance estimation. Moving beyond a bi-variate setting, Barndorff-Nielsen et al. (2011) introduce multivariate realized kernels which deliver consistent and positive semi-definite covariance matrix estimates. For these multivariate kernels, refresh time-sampling discards a substantial part of the available high-frequency data, although Hautsch et al. (2012) propose a block approach to reduce this problem. None of the aforementioned approaches, however, can empirically cope with a universe consisting of hundreds or even thousands of stocks that make up most stock market indices used to benchmark fund managers.

Recently Fan et al. (2008) revisited the use of factor models for covariance estimation in case of a large number of assets, in order to reduce the dimensionality of the problem. They show that the factor model approach improves over the sample covariance matrix (based on daily data) in particular when the portfolio optimization problem requires the inverse of the covariance matrix. The reason is that in the factor model approach only the factor covariance matrix needs to be inverted, which typically is of much lower dimension. In addition, using the covariance matrix based on a factor structure reduces the problem of error maximization for portfolio construction applications, see for example Jagannathan and Ma (2003).

With the MFFM we introduce an innovative methodology that exploits the advantages of both high-frequency data and the factor model approach: It enables more efficient estimation of covariances whilst still being able to cope with a very large number of stocks. The covariance matrix based on the factor model requires three estimates: The covariance matrix of the factor returns, the factor loadings, and the stock-specific variances. Without compromising the consistency and positive-definiteness of the resulting covariance matrix we can choose different sampling frequencies for each of these three estimates.

First, in the MFFM approach we use realized covariances obtained from high-frequency intraday returns to estimate the daily factor covariance matrix. This is motivated by the fact that nowadays highly liquid financial contracts such as index futures and exchange-traded funds (ETFs) are available as proxies for the most commonly used factors. This further increases the added value of high-frequency data because microstructure frictions are relatively small. For this reason the factor covariance matrix can be estimated with high precision from intraday data. Second, we estimate the factor loadings using daily data for the reason that single-day betas based on high-frequency data are very noisy due to the non-synchronicity between factor returns and stock returns, see for example Andersen et al. (2006b), Todorov and Bollerslev (2010) or Hansen et al. (2010) for related discussions.

Finally, although intraday data is also available for individual stocks, these are generally less liquid than index futures and ETFs. Hence, we can use intraday data for stock-specific variances, but possibly at a lower frequency than the one used for the factor covariance matrix.

We provide theoretical, simulation-based and empirical evidence that the MFFM offers a useful approach for estimating vast dimensional covariance matrices. In the theoretical part of this paper we show that, assuming i.i.d. microstructure noise and a Poisson arrival process for non-synchronous trading, the covariance estimates of the MFFM are unbiased and we obtain a closed form expression for the variance of these covariance estimates. Based on analytical expressions for the variance of the estimators, we show that the MFFM improves substantially in terms of efficiency over that of the popular Hayashi and Yoshida (2005), realized covariance and lead-lag adjusted realized covariance estimators.

Next, we use Monte Carlo simulations to show that the MFFM estimator is also superior to the realized covariance estimator, when we relax several of the assumptions underlying the theoretical results and move from the bi-variate case to a realistic setting of 500 assets.

We empirically evaluate the MFFM estimator by comparing its performance to the (sample) realized covariance and a factor model based on daily returns. We consider three stock universes: The S&P 500 (large caps, most liquid), the S&P 400 (mid caps), and the S&P 600 (small caps, illiquid). To the best of our knowledge, we are the first in the literature to consider such high dimensional problems involving high-frequency data. Of course, in the empirical case unlike for the theory and simulations we do not know the true covariances. For this reason we analyze two empirical applications. First, we use Mincer-Zarnowitz and forecast encompassing regressions to obtain insights in the ability of the MFFM to forecast the volatility of vast dimensional portfolios out-of-sample. Second, we evaluate the performance of minimum tracking error portfolios.[1] We find that in each of the three S&P universes the out-of-sample MFFM portfolio volatility forecasts improve upon realized covariance and daily factor model forecasts when we rank the forecasts on their Mincer-Zarnowitz R^2. Using encompassing regressions, in which we add the realized covariance or daily factor model forecasts to MFFM we find that the coefficient on realized covariance and the daily factor model is negative. Adding these forecasts to the MFFM forecasts improves the MFFM forecasts only marginally. When the objective is to track a benchmark using out-of-sample covariance matrix forecasts, the MFFM provides smaller tracking errors and much smaller portfolio turnover than the realized covariance. Conventional factor models based on daily data manage to achieve a similar tracking error as the MFFM, but only if a long historical data period is used. This is due to the fact that it needs a substantial amount of smoothing, whereas the MFFM can manage the same performance with a very short span of historical data. In addition, the portfolio turnover of the daily factor model is about three times larger than the MFFM turnover. For forecasting portfolio volatility and for minimizing the tracking error we find that differences between the MFFM and realized covariance increase as we move from the most liquid stock universe to the least liquid universe, as expected.

In recent work Hansen et al. (2010) and Noureldin et al. (2012) advocate the use of high-frequency data in a parametric GARCH framework. Related to our idea of using a mixed-frequency sampling approach for modeling vast dimensional covariance matrices several authors have recently implemented subcases and modifications of the mixed-frequency (factor model) methodology. Kyj et al. (2009) study a single-factor model, which is a special case of the MFFM, to forecast covariance matrices in the absence of noise and non-trading. Halbleib and Voev (2011) propose to use mixed-frequency sampling for predicting covariance matrices by using high-frequency data for realized volatilities and low-frequency data for correlations. Hence, without using a factor structure, by using mixed-frequency sampling they successfully circumvent the issue of non-trading for estimating correlations. Combining the Hautsch et al. (2012) blocking and regularization kernel estimator with the MFFM, Hautsch et al. (2011) propose to select factors in a data driven way where mixed-sampling frequencies can be used for volatilities, correlation eigenvalues and eigenvectors. In contrast

[1] Chan et al. (1999) show that differences between covariance estimators are small for minimum variance portfolios because the market factor dominates.

to our study they use a multi-time-scale approach for reducing the impacts of noise, non-trading and estimation error, rather than studying these frictions explicitly.

The remainder of this paper is structured as follows. In Section 2 we derive the theoretical properties of the MFFM and provide a theoretical comparison with the bi-variate Hayashi and Yoshida (2005), realized covariance and lead-lag estimators. Section 3 contains an extensive Monte Carlo study in which we replicate the S&P500 universe to evaluate the realized covariance and the MFFM covariance matrix estimates. In Section 4 we study the empirical performance of the MFFM and compare it to the realized covariance and a factor model based on daily data. We conclude in Section 5.

4.2 The Mixed-Frequency Factor Model

Consider a linear factor structure for the return on asset i, that is

$$r_i = \mu_i + \beta_i' f + \varepsilon_i \tag{4.1}$$

where f is a $K \times 1$ vector of common factors, β_i is a $K \times 1$ vector of factor loadings measuring the exposure to f, and ε_i is the idiosyncratic component. We assume that $\mathsf{E}[f] = 0$ and $\mathsf{E}[\varepsilon_i] = 0$, such that μ_i is the expected return. Furthermore, we assume that the idiosyncratic component is orthogonal to the common factors, i.e. $\varepsilon_i \perp f$. Under these assumptions the covariance between asset i and asset j can be expressed as

$$\gamma_{ij} \equiv \mathsf{Cov}[r_i, r_j] = \beta_i' \Lambda \beta_j + \sigma_{ij} \tag{4.2}$$

where $\Lambda = \mathsf{E}[ff']$ is the factor covariance matrix and $\sigma_{ij} = \mathsf{E}[\varepsilon_i \varepsilon_j]$ is the covariance between the assets' idiosyncratic components. Throughout, we consider a "strict" factor structure in the spirit of Ross (1976), i.e. we assume that the factor structure exhausts the dependence among the assets so that $\sigma_{ij} = 0$ for $i \neq j$. Approximate factor models where σ_{ij} can be non-zero but small are considered in Chamberlain and Rothschild (1983); Ingersoll (1984) and Connor and Korajczyk (1994).

Using hats to denote estimates of unknown quantities, the covariance estimator is given by

$$\widehat{\gamma}_{ij} = \widehat{\beta}_i' \widehat{\Lambda} \widehat{\beta}_j \qquad \text{for } i \neq j. \tag{4.3}$$

The properties of this generic covariance estimator are characterized in the theorem below, where we use the notation $\widehat{X} = X + X^{\varepsilon}$.

Theorem 4.2.1 *Assuming (i)* $\mathsf{E}[\sigma_{ij}] = 0$ *for* $i \neq j$*, (ii)* $\mathsf{E}[\beta^\varepsilon] = 0$*, (iii)* $\mathsf{E}[\Lambda^\varepsilon] = 0$*, and (iv)* $\beta^\varepsilon \perp \Lambda^\varepsilon$ *element-by-element, then*

$$\mathsf{E}[\widehat{\gamma}_{ij}] = \gamma_{ij}, \tag{4.4}$$

for $i \neq j$ *with*

$$\begin{aligned}\mathsf{V}[\widehat{\gamma}_{ij}] &= \beta_i' \Lambda \Sigma_{\beta,j} \Lambda \beta_i + \beta_j' \Lambda \Sigma_{\beta,i} \Lambda \beta_j + tr(\Sigma_{\beta,i} \Lambda \Sigma_{\beta,j} \Lambda) \\ &\quad + g(\beta_i \beta_i', \beta_j \beta_j', \Phi) + g(\beta_i \beta_i', \Sigma_{\beta,j}, \Phi) + g(\beta_j \beta_j', \Sigma_{\beta,i}, \Phi) + g(\Sigma_{\beta,i}, \Sigma_{\beta,j}, \Phi),\end{aligned} \tag{4.5}$$

where $\Sigma_{\beta,i} = \mathsf{V}[\widehat{\beta}_i]$ *and* $\Phi = \mathsf{E}[vech(\Lambda^\varepsilon) vech(\Lambda^\varepsilon)']$ *and*

$$g(A, B, \Phi) = \sum_{m,n,p,q}^{N} A_{mp} B_{nq} \Phi_{f(p,n), f(q,m)},$$

and $f(p,q) = N(\min\{p,q\} - 1) + \frac{1}{2}(\min\{p,q\} - \min\{p,q\}^2) + \max\{p,q\}$*.*

Proof See Appendix 4.A. ■

It is useful to note that the assumptions in this Theorem are not unreasonable for the mixed-frequency approach developed in this paper. Specifically, we propose to estimate betas using low-frequency data, such that it is plausible to assume that betas are unbiased, whereas the factor covariance matrix is estimated from high-frequency data. The factors are essentially free of microstructure noise since the ETFs we propose as factors are very liquid, see Table 4.1. This justifies the assumption that the factor covariance estimates are unbiased and that possible sources of noise in low-frequency betas and factors observed at high sampling frequencies are independent[2].

The linear factor decomposition of asset returns in (4.1) has a long and established history in the theoretical and empirical finance literature. Three types of factor models can be distinguished, depending on how the factors f and the associated exposures β are constructed. Specifically, the model in (4.1) can be categorized as (i) a statistical factor model (Ross, 1976) when both β and f are unspecified and inferred from the panel of asset returns, (ii) a characteristic-based factor model (Rosenberg, 1974) when β is fixed and determined by asset-specific characteristics while f is inferred from the data, or (iii) a macro-economic factor model (Chen et al., 1986) when f is observable and derived from macroeconomic or asset pricing theory while

[2]In Section 3 we analyze the impact of estimation errors in betas for the MFFM.

Table 4.1: Description of ETF contracts

ticker	description	sector / style classification	# trades per day
XLE.A	Energy Sector SPDR Fund	Energy	76,392
XLB.A	Materials Sector SPDR Fund	Materials	16,708
XLI.A	Industrial Sector SPDR Fund	Industrials	12,207
XLY.A	Consumer Discretionary Sector SPDR Fund	Consumer Discretionary	9,731
XLP.A	Consumer Staples Sector SPDR Fund	Consumer Staples	6,153
XLV.A	Health Care Sector SPDR Fund	Health Care	6,697
XLF.A	Financial Sector SPDR Fund	Financials	112,191
XLK.A	Technology Sector SPDR Fund	Information Technology	9,243
IYZ.N	iShares Telecommunications Sector Fund	Telecommunications	762
XLU.A	Utilities Sector SPDR Fund	Utilities	11,753
SPY.A	SPDR Trust Series 1	Large Cap	356,876
IWM.A	iShares Russell 2000 Index Fund	Small Cap	140,192
IVE.N	S&P 500 Value Index Fund	Value	3,030
IVW.N	S&P 500 Growth Index Fund	Growth	3,912
	Average across ETFs		54,703
	Average across S&P400 constituents		4,898
	Average across S&P500 constituents		19,395
	Average across S&P600 constituents		1,990

Note: This table lists the ETF contracts used in the empirical analysis, together with the average number of trades per day over the period January 2007 through April 2009. The "SMB" ("HML") factor is specified as IWM.A - SPY.A (IVE.N - IVW.N).

β is estimated from the data. See Grinold and Kahn (2000) or Connor et al. (2012) for further discussion.

The factor model we develop in this paper can be classified as a traditional macroeconomic model in the sense that the factors are observable and their loadings are estimated from the data. However, its construction is designed to make efficient use of high-frequency data while simultaneously avoiding the potentially severe biases induced by market micro-structure noise. Specifically, our "*mixed-frequency factor model*" involves the use of liquid assets as factors for precise estimation of the factor covariance matrix using high-frequency data, while factor loadings are estimated using lower-frequency returns of the, possibly illiquid, individual assets. The use of liquid factors in the MFFM is motivated by the empirical observation that a growing number of highly liquid exchange traded funds (ETFs) and futures contracts are now available that proxy commonly used country, industry, and style factors. With minimal spreads and accurate real-time pricing for many of these contracts, the effects of market microstructure noise are of little concern and the use of high-frequency data is justified. Particularly for a large and heterogeneous asset universe, however, many of the individual assets may be illiquid and contaminated by market microstructure effects at high sampling frequencies. To support this point Table 4.1 shows statistics on the ETFs we use in our empirical application. The average number of observations for these ETFs is over 54,000 per day. In contrast the average number of observations for a constituent of the S&P500 is just over 19,000 per day, and this drops to about only 2,000 per day for the constituents of the S&P600, i.e. the small cap index.

We now specialize the rather general result in Theorem 4.2.1. to the MFFM setting to gain further insights into its properties. We define F and $\mathcal{F}$ as the matrices of low- and high-frequency factor return observations with dimensions $(T \times K)$ and $(M \times K)$. Similarly, R_i and $\mathcal{R}_i$ denote the vectors of low- and high-frequency returns of asset i of length T and N_i, and τ_i the $(N_i \times 1)$ vector of time-stamps associated with $\mathcal{R}_i$.

Assumption N The factor returns $\mathcal{F}$ are jointly normal with zero mean, serially uncorrelated and observed without friction[3]. The (integrated) factor covariance matrix is estimated using the high-frequency factor returns as $\widehat{\Lambda} = \mathcal{F}'\mathcal{F}$.

Assumption O The asset return dynamics at low frequency are governed by a linear factor model as in (4.1) with i.i.d. normal residuals ε_i. The factor loadings are estimated by means of linear regression using the low-frequency returns $\widehat{\beta}_i = (F'F)^{-1}F'R_i$.

Corollary 4.2.2 *Let assumption **N**, **O**, and those in Theorem 4.2.1. hold. Then for $i \neq j$*

$$\mathsf{E}[\widehat{\gamma}_{ij}] = \gamma_{ij} \tag{4.6}$$

and

$$\mathsf{V}[\widehat{\gamma}_{ij}] = \frac{A}{T} + \frac{B}{M}, \tag{4.7}$$

where

$$\begin{aligned} A &= \sigma_j^2 \beta_i' \Lambda \beta_i + \sigma_i^2 \beta_j' \Lambda \beta_j + \sigma_i^2 \sigma_j^2 \frac{K}{T}, \\ B &= \sum_{m,n,p,q}^{K} \left(\beta_i(m)\beta_i(p) + \Sigma_{\beta,i}(m,p)\right)\left(\beta_j(n)\beta_j(q) + \Sigma_{\beta,j}(n,q)\right)\left(\Lambda_{pq}\Lambda_{nm} + \Lambda_{pm}\Lambda_{nq}\right) \end{aligned}$$

Proof See Appendix 4.A. ■

The above corollary provides insights into the properties of the MFFM covariance estimator. In particular, it is unbiased with a variance that can be attributed to the

[3] Given the highly liquid ETFs we propose as factors, see Table 4.1, it is justified to assume that factor returns are serially uncorrelated and observed without friction.

measurement error in factor loadings (i.e. A/T) and to the measurement error in the factor covariance matrix (i.e. B/M).[4]

To illustrate the efficiency of the MFFM in a bi-variate setting, we compare it to the (i) Hayashi and Yoshida (2005) estimator, (ii) realized covariance and (iii) realized covariance lead-lag estimator. For this purpose, we assume that intraday price observations for asset i (from which the returns $\mathcal{R}_i$ are computed) arrive according to a Poisson process with intensity $\lambda_i = \mathsf{E}[N_i]$. Further, we assume that prices are contaminated with i.i.d. microstructure noise with variance $\xi_i^2 = \pi_i \gamma_i^2 / \lambda_i$. We use closed-form expressions for the efficiency of the popular aforementioned estimators (see Griffin and Oomen (2011) for details) and compare these with the variance of the MFFM covariance estimator. To compute the variance of the MFFM covariance estimator, we need to make some assumptions about the underlying factor structure. Here, we use a setting with $K = 5$ factors, factor loadings $\beta_i = (0.5, -0.1, 0, 0.2, 0.6)'$, $\beta_j = (0.7, -0.2, -0.3, 0.4, 0.2)'$, and factor covariance matrix $\Lambda = I_K + \frac{1}{2}(1 - I_K)$. The specific or idiosyncratic risk component is $\sigma_h^2 = \beta_h' \Lambda \beta_h$ for $h \in \{i, j\}$ so that the R^2 of the factor regression is around 50% and the assets have a correlation of $\rho_{ij} \approx 40\%$ with:

$$V(r) = (\beta_i, \beta_j)' \Lambda (\beta_i, \beta_j) + \Sigma = \begin{pmatrix} 2.075 & 0.765 \\ 0.765 & 1.584 \end{pmatrix}$$

In Figures 4.1 and 4.2 the efficiency of the estimators is plotted against the number of returns an estimator has access to. Figure 4.1 displays the performance for asynchronously traded assets i and j that are observed without additive microstructure noise. Figure 4.2 shows the performance when the asynchronous returns are contaminated with microstructure noise.

From these graphs, we observe that for reasonable scenarios the MFFM comfortably outperforms the HY estimator unless a large number of intraday return observations on the individual assets is available. For instance, using 5-minute ($M = 78$) factor returns to estimate the 5×5 factor covariance matrix and 1 year ($T = 250$) of daily asset returns to estimate the 5×1 factor loading vector β, the MFFM delivers better estimates unless the HY estimator has access to more than 500 clean or 1250 noisy intraday (asynchronous) observations. The MFFM is also substantially more efficient than the realized covariance (lead–lag) estimator.

[4] Note that in some circumstances β is (assumed to be) known so that $V(\widehat{\gamma}_{ij}) = B/M$, see e.g. Grinold and Kahn (2000, Ch. 3).

Figure 4.1: Comparison of MFFM to Hayashi-Yoshida, RC and RC-LL in terms of ln MSE without microstructure noise

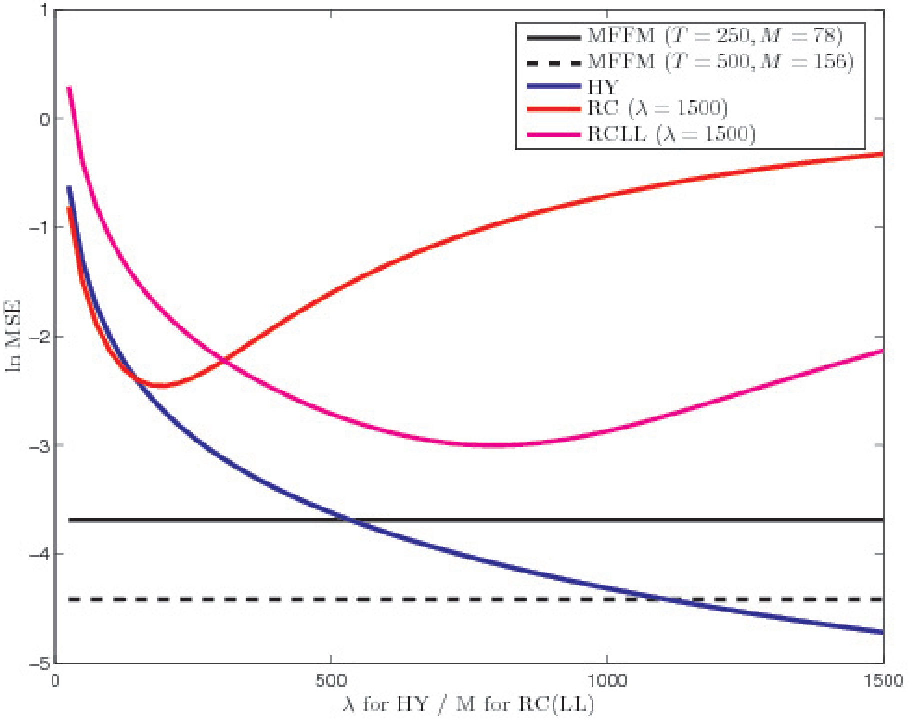
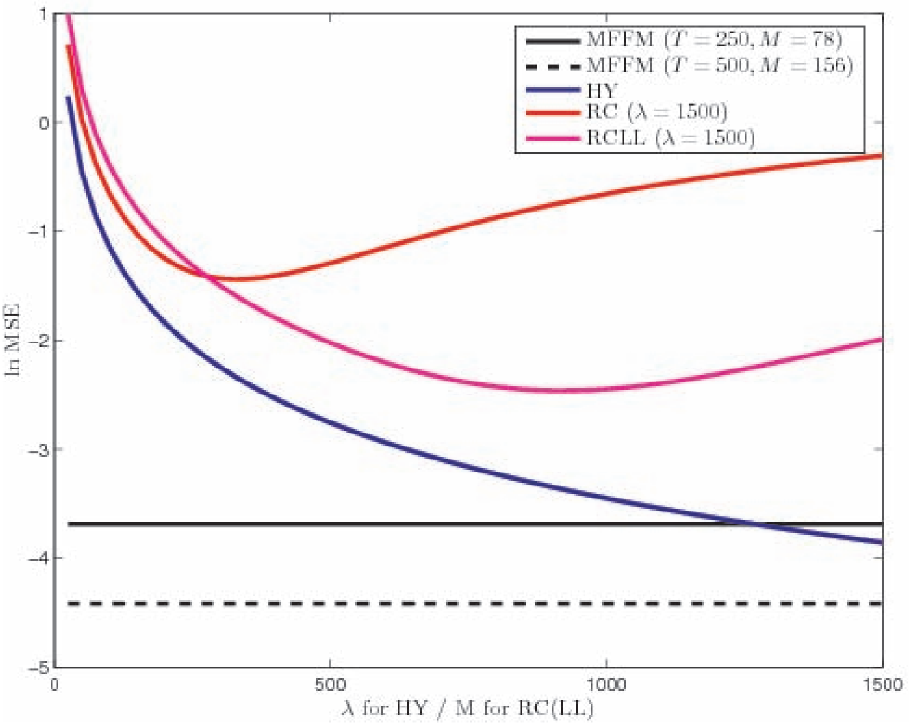


Figure 4.2: Comparison of MFFM to Hayashi-Yoshida, RC and RC-LL in terms of ln MSE with microstructure noise



4.3 Monte Carlo Simulation

The theoretical results presented in the previous section demonstrate the superior properties of MFFM compared to existing covariance estimators in a bi-variate setting. An important additional feature of the MFFM is that its factor structure ensures stable and positive definite covariance matrices in higher dimensional settings. In this section we provide further insights into this property of the MFFM by means of an extensive simulation study. In addition to increasing the dimension of the covariance matrix to realistic magnitudes of several hundreds of assets, we relax some of the assumptions made in the previous section to study the effects of estimation errors in the factor exposures for individual stocks.

4.3.1 Simulation design

We simulate returns for asset i at high frequency as

$$\mathcal{R}_{i,t_j} = \mathcal{F}_{t_j}\beta_i + \varepsilon_{i,t_j} + \eta_{i,t_j} - \eta_{i,t_{j-1}}$$

where $i = 1, 2, \ldots,$ (number of stocks), $j = 1, 2, \ldots, N_i$ (number of observations in a day), $0 \leq t_{j-1} < t_j \leq 1$, and $\mathcal{F}_{t_j}$ denotes the factor return between t_{j-1} and t_j. To ensure a realistic setup, we calibrate the data generating process (DGP) based on characteristics of the data used in the empirical application in Section 4. Specifically, the common factor $\mathcal{F}$ is a tri-variate Brownian motion with a covariance structure Λ as estimated for the daily Fama and French three-factor (market, size, and value) returns[5] over the period January 1998 through December 2007. The 3×1 vector of factor exposures β_i are obtained from regressing daily (corporate action adjusted) excess returns for each of the S&P500 constituents on the Fama and French three-factor returns, using the same sample period.

The idiosyncratic component $\varepsilon_{i,t_j} \sim$ i.i.d. $\mathcal{N}(0, \sigma_i^2(t_j - t_{j-1})/N_i)$ where σ_i^2 is the residual variance of the Fama and French regression for the i^{th} S&P500 constituent, the market microstructure noise component $\eta_{i,t_j} \sim$ i.i.d. $\mathcal{N}(0, \omega_i^2)$ where $\omega_i^2 = \frac{1}{4}(\beta_i'\Lambda\beta_i + \sigma_i^2)/N_i$,[6], and the observation times t_j are based on a Poisson process with intensity λ_i set to the average number of daily trades for the i^{th} S&P500 constituent, N_i.

This simulation setup ensures a realistic covariance structure of the 500-dimensional returns process at low frequency. At the same time, it incorporates non-synchronous

[5] Data available from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data library.html.

[6] As shown in Oomen (2009), this level of noise is representative for the S&P500 universe.

trading and market microstructure noise at high frequency. We simulate second-by-second factor prices for a 6,5 hour trading day (23,400 seconds) and residuals to generate stock returns according to the DGP. The Poisson process in combination with the market microstructure noise then provide the simulated stock price paths.

4.3.2 Covariance models

For the simulated asset returns we estimate the covariance matrix using either MFFM or the realized covariance matrix.

Realized covariance

The realized covariance is a popular and efficient estimator of the latent integrated covariance. RC converges in probability to the integrated covariance in the absence of noise, see Barndorff-Nielsen and Shephard (2004). The RC is estimated as the cross-product of intraday returns:

$$RC = \mathcal{R}'\mathcal{R}$$

where $\mathcal{R}$ is a $N \times S$ matrix of intraday returns. Here N is the number of non-overlapping intraday intervals where in each interval we take the last observed price. In case an interval has no price the last price of the previous interval is used, resulting in a zero return for that interval.

Mixed-frequency factor model

For the MFFM we need to estimate the factor loadings, the factor covariance matrix, and the residual variances. This will give us the MFFM-based covariance matrix as

$$\text{MFFM} = \tilde{\beta}'\widehat{\Lambda}\tilde{\beta} + \widehat{\Delta} \tag{4.8}$$

where $\widehat{\Lambda} = \mathcal{F}'\mathcal{F}$ is the estimated $K \times K$ realized factor covariance matrix, $\tilde{\beta}$ is the $K \times S$ matrix of factor loadings contaminated with i.i.d. measurement errors, and $\widehat{\Delta}$ is a $S \times S$ matrix with the estimated residual variances on the diagonal and zeroes elsewhere.

In empirical applications the factor covariances, residual variances and factor loadings can be estimated at different sampling frequencies. First, we propose to estimate the betas at the daily frequency. The main problem with estimating betas with intraday returns is that they can become severely biased towards zero due to the non-synchronicity of the relatively liquid factors and the considerably less liquid stocks. Also, Todorov and Bollerslev (2010) illustrate that jumps can cause single-day realized betas to exhibit erratic time-series behavior. We therefore propose a simple moving window history of 2.5 years of daily returns data that combined with OLS delivers betas that are smooth and by construction exhibit a much smaller variance than single-day realized betas while improving upon using monthly data.[7]

Second, the realized factor covariance matrix can be estimated at very high frequencies due to the high liquidity of ETF factor proxies. Third and finally, the residual variances can also be estimated using intraday data, but possibly at a lower frequency than the factor covariance matrix. This is to reduce the impact of the noise terms (η). We first compute the residuals, using $\varepsilon_{i,t_j} = \mathcal{R}_{i,t_j} - \mathcal{F}_{t_j}\tilde{\beta}_i$. Then we compute the variances of these residuals. While it is possible to use all intra-day returns for asset i for this purpose, due to market microstructure noise and the difference between the observation frequency for the factors and the stock prices these residual variances will be biased upwards. Below we examine to what extent lowering the sampling frequency to compute these residual variances reduces this bias.

4.3.3 Simulation results

As a measure of relative accuracy of the covariance estimates, we compute their distance to the true covariance matrix using the Frobenius norm. We do this separately for the diagonal and off-diagonal elements to disentangle the variance and covariance terms, i.e. we compute

$$\sum_{i=1}^{S} |\widehat{\Gamma}_{ii} - \Gamma_{ii}|^2 \qquad \text{and} \qquad 2\sum_{i=1}^{S}\sum_{j=i+1}^{S} |\widehat{\Gamma}_{ij} - \Gamma_{ij}|^2 \tag{4.9}$$

where $\Gamma = \beta'\Lambda\beta + \Sigma$, with Σ the diagonal matrix with the residual variances on the diagonal, and $\widehat{\Gamma}$ being either the MFFM or the RC covariance matrix estimate.

Non-synchronous prices, no noise

Figure 4.3 illustrates the performance of the RC and the MFFM when prices are non-synchronous but market microstructure noise is absent (i.e. $\omega_i = 0$). The covariance results illustrate that the MFFM has an excellent performance and is very robust across sampling frequencies. Furthermore, in contrast to RC, its performance is not affected by non-synchronicity.

[7] We have empirically experimented with the use of intraday data to estimate beta's. For sampling frequencies ranging from 15s to 65m we find that using intraday data to estimate beta's substantially increases the variance of the MFFM estimator. Aggregating realized betas to monthly or quarterly data and then applying EWMA smoothing helps to decrease the variance but the performance is inferior compared to using low-frequency beta's. Detailed results are available upon request.

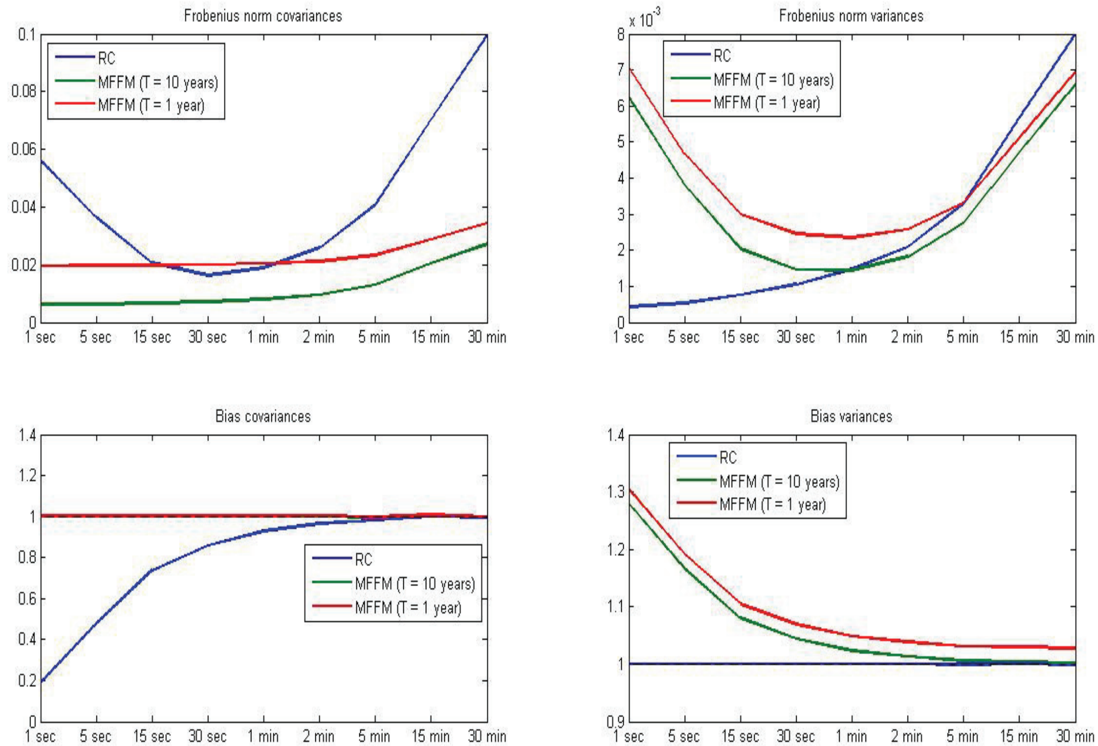


Figure 4.3: This Figure displays the Frobenius Norm (Equation (4.9)) of the (co)variance estimation errors for the MFFM (Equation (4.8)) and RC based on 10,000 repetitions in a setting with high liquidity for the individual stocks, the prices are non-synchronous and not contaminated by microstructure noise.

Non-synchronicity, however, does affect the MFFM variance estimates. This may seem counterintuitive at first as non-synchronicity usually affects the covariances and not so much the variances. The reason for the upward bias in the MFFM variances is caused by a mismatch between the very frequently observed factor returns and less frequently observed stock returns, which results in an additional quadratic bias term in the MFFM diagonal. The mismatch between liquid factors and less liquid stocks disappears when sampling at the 5-minute frequency or lower. Also note that with an increasing number of assets in a portfolio, the variance elements play a more limited role as the covariances become more dominant. For example, in the simulation with 500 stocks we only have 500 variances in contrast to 249,500 covariances. However, in some circumstances it may be interesting to introduce a third sampling frequency, that is, we can sample the residual returns at a lower sampling frequency than the sampling frequency for factor returns used to estimate the factor covariance matrix. We examine this possibility below.

Non-synchronous prices and market microstructure noise
Figure 4.4 illustrates the more practically relevant case where prices are non-synchronous but also contaminated by additive market microstructure noise.

Market microstructure noise does not deteriorate the performance of both covariance estimators as the noise is (assumed to be) cross-sectionally independent. However, the noise does affect the variances computed with the MFFM (through the residual variances) and RC. For both estimators the diagonal elements perform fairly similar at the 5min and lower sampling frequencies while the MFFM covariances are substantially more efficient.

Lower frequency for residual variances to reduce MFFM variance bias
Finally, we examine the effects of introducing a third sampling frequency, that is, sampling the residual returns at a lower sampling frequency than the sampling frequency for factor returns used to estimate the factor covariance matrix.

Note that for the MFFM we use the assumption that the common factors fully capture the correlation among asset returns thus the residual returns only enter the MFFM by adding the diagonal residual variances to the systematic variances. Hence introducing a third frequency still delivers a well conditioned positive semi-definite covariance matrix.

Figure 4.5 illustrates how reducing the sampling frequency of the residual variances relative to the frequency used for the factor covariances can improve the efficiency of the variance elements in the MFFM. If the sampling frequency for the factor returns is ultra-high (sampling more frequently than once a minute) we use

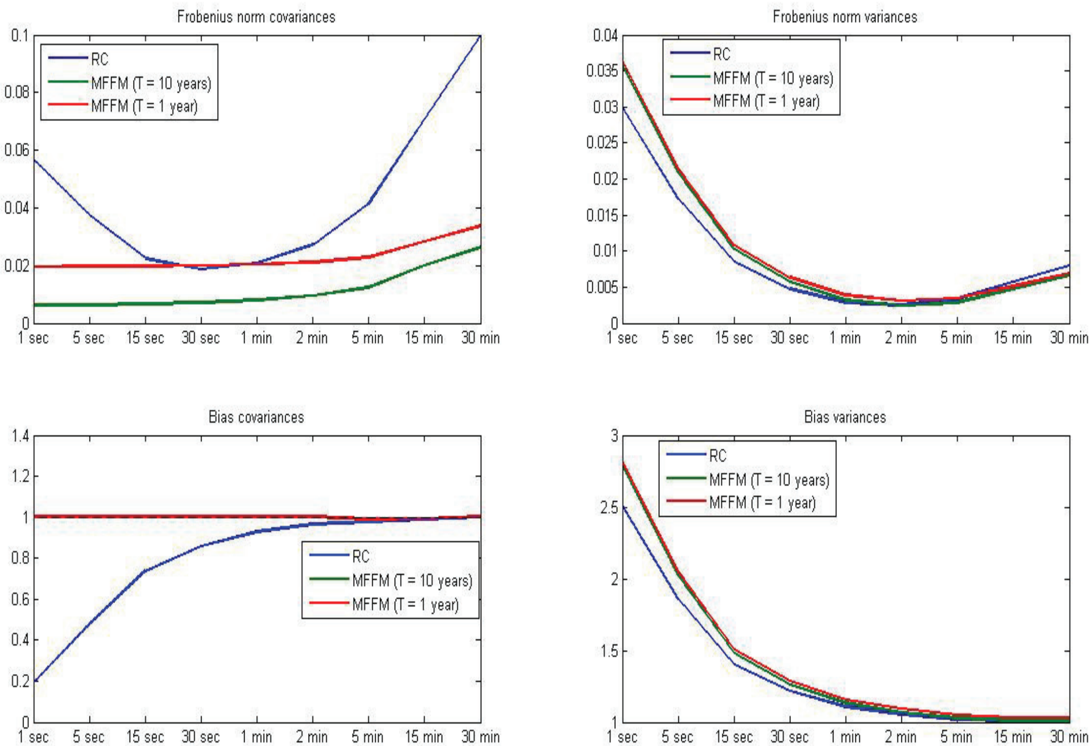


Figure 4.4: This Figure displays the Frobenius Norm (Equation (4.9)) of the variance and covariance estimation errors for the MFFM (Equation (4.8)) and RC based on 10,000 repetitions in a setting with high liquidity for the individual stocks, the prices are non-synchronous and contaminated by microstructure noise.

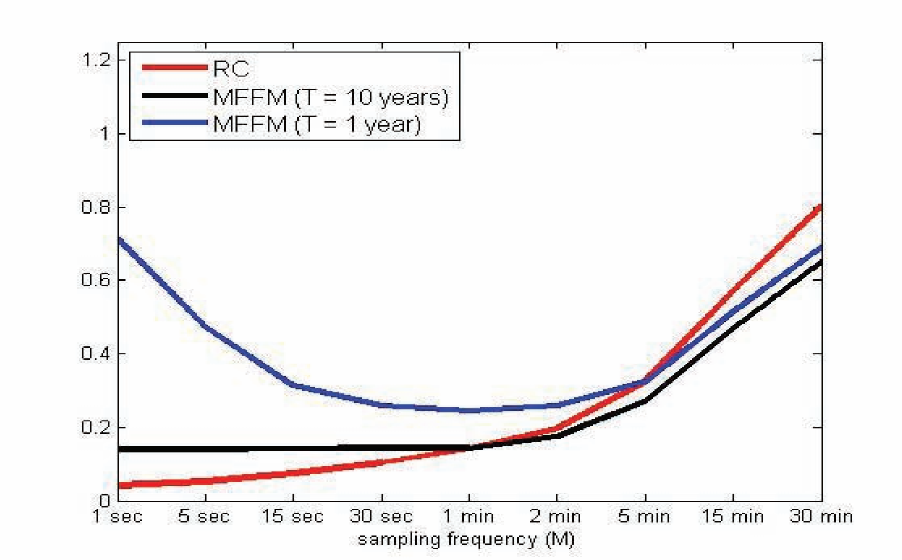


Figure 4.5: This Figure displays the Frobenius norm for the variance elements of the MFFM when a bias adjustment is used by introducing a lower 3rd sampling frequency for calculating idiosyncratic risk. The MFFM with small measurement errors in the betas (T = 10 years) is bias-adjusted while the case with larger measurement errors (T = 1 year) is not bias-adjusted. The sampling frequency used for residual risk is the 1m frequency if we use factor covariances sampled at higher frequencies. When the sampling frequency for the factor covariances is 1m or lower, then we use the same sampling frequency for residual risk.

the 1-minute frequency to sample the residual variances to restore the efficiency of the variance elements in the MFFM. At sampling frequencies lower than the 1-minute frequency we use the same frequency for the factor covariances and residual variances. Using a lower sampling frequency than the 1-minute frequency to calculate residual risk is of course also possible to eliminate the bias but would deteriorate the performance of the MFFM as it also increases the variance of the estimates. This is the well-known trade-off in the efficiency of high-frequency data estimates between bias and precision.

4.4 Empirical Applications

We apply the MFFM approach to three universes of stocks with different levels of market capitalization to assess its empirical performance compared to the realized covariance and the factor model based on daily data. Whereas in the simulation experiments reported in the previous section we evaluated the (relative) accuracy of *measurements* of daily covariances, here we focus on the performance in terms of out-of-sample *forecasts*. In empirical applications the "true" covariances are unobservable. For this reason we focus on forecasts instead of covariance measurements. We do this in two ways. First, we evaluate the forecasting performance of the MFFM and RC for the volatility of vast dimensional equally-weighted portfolios. Second, we compare the out-of-sample performance by constructing minimum tracking error portfolios.

4.4.1 Data

Our data sets comprises the constituents of the S&P500 (large caps), S&P400 (mid caps) and S&P600 (small cap) indexes. For each index we only use those stocks that were included in the index during the complete sample period, which runs from May 1, 2004 until April 30, 2009. This leaves 442 large-caps, 342 mid-caps and 491 small-caps. We collect high-frequency data from November 1, 2006 onwards. Specifically, we sample National Best Bid Best Offer (NBBO) mid-points, originating from NYSE and NASDAQ only, at the 15-seconds sampling frequency. The first 2.5 years of the sample period are used only to obtain estimates of the factor loadings in the MFFM, for which we require only daily (close-to-close) returns.

4.4.2 Covariance estimators

Volatilities and correlations of stock returns typically are time-varying. We incorporate this feature explicitly in the methodology that is used to obtain covariance forecasts, as described in detail below.

Realized Covariance

In the portfolio volatility forecasting exercise with S stocks we use the traditional RC estimator to obtain an estimate of the covariance matrix on day t, that is,

$$RC_t = \mathcal{R}_t'\mathcal{R}_t, \tag{4.10}$$

where $\mathcal{R}_t$ is the $N \times S$ matrix of (intraday) stock returns on day t.

In the minimum tracking error application we employ intraday *excess* stock returns net of the relevant benchmark, which for each of the three universes is taken to be the corresponding S&P index. The *active* realized covariance estimator is then computed as

$$RC_t^A = (\mathcal{R}_t - \mathcal{R}_{Mt}e)'(\mathcal{R}_t - \mathcal{R}_{Mt}e), \tag{4.11}$$

where $\mathcal{R}_{Mt}$ is a $N \times 1$ vector of intraday returns on the corresponding index, and e is an $S \times 1$ vector of ones. In both cases we include overnight returns by adding the outer product of the vector of *close-to-open* (active) returns.

Finally, we consider the RC_t and RC_t^A estimators for a range of intra-day sampling frequencies, equal to 15 seconds, 1, 5, 15, 30, 65 and 130 minutes. We also include the sample realized covariance based on daily *close-to-close* returns.

Mixed-frequency factor models

For the MFFM approach we employ a 12-factor model based on the Fama and French (1993) size and value factors and ten industry factors. The motivation to use 10 industry factors is that many stocks have activities in (and thus exposure to) multiple sectors, see Grinold and Kahn (2000), page 60. We allow for time-varying factor loadings, which are estimated using a moving window of 2.5 years (632 days) of daily *close-to-close* returns[8]. By means of the regression

$$R_{i,t-j} = F_{t-j}\beta_{i,t} + \varepsilon_{i,t-j}, \qquad \text{for } j = 0, 1, \ldots, L-1, \tag{4.12}$$

where R_{it} is a vector of daily returns on stock i, $F_t = [SMB_t \quad HML_t \quad I_1 \quad \ldots \quad I_{10}]$ is a matrix of factor returns on the size (Small-Minus-Big), value (High-Minus-Low) and industry factors, and L denotes the length of the moving window. The intraday residuals needed to compute idiosyncratic variances are obtained as

$$\hat{\varepsilon}_t = \mathcal{R}_t - \mathcal{F}_t\hat{\beta}_{t-1}, \tag{4.13}$$

Finally the MFFM covariance matrix estimate for day t is then computed as

$$\text{MFFM}_t = \hat{\beta}_{t-1}'\hat{\Lambda}_t\hat{\beta}_{t-1} + diag(\hat{\varepsilon}_t'\hat{\varepsilon}_t), \tag{4.14}$$

where $\hat{\Lambda}_t = \mathcal{F}_t'\mathcal{F}_t$ is the factor covariance matrix. The motivation to use 'lagged' factor loading estimates $\hat{\beta}_{t-1}$ (that is, based on the moving window that ends on

[8]In earlier studies on factor models the number of observations used for estimating betas is usually 3 to 5 years. Here we use 2.5 years as using a longer history would limit the number of constituents that survived our sample period, thereby reducing the dimension of the covariance matrix.

day $t-1$) rather than $\hat{\beta}_t$ stems from assumption *(iv)* in Theorem 4.2.1. stating that the measurement errors in the factor loadings and in the factor covariance matrix are orthogonal, i.e. $\beta^{\varepsilon} \perp \Lambda^{\varepsilon}$. By lagging the beta estimates in (4.13) and (4.14) we avoid the possibility that measurement errors in factor loadings are correlated with the measurement errors in the factor covariance matrix.[9] For the minimum tracking error application we follow the same approach, except that we use stock returns in excess of the returns on the relevant market index. Hence, we obtain estimates of the factor loadings from the regression

$$R_{i,t-j} - R_{M,t-j} = F_{t-j}\beta_t^A + \varepsilon_{i,t-j}^A, \qquad j = 0, 1, \ldots, L-1, \tag{4.15}$$

while we compute the *active* intraday residuals as

$$\hat{\varepsilon}_t^A = \mathcal{R}_t - \mathcal{R}_{M,t} - \mathcal{F}_t \hat{\beta}_{t-1}^A, \tag{4.16}$$

and the MFFM estimator for day t using (4.14). We include overnight returns in the factor covariance matrix $\hat{\Lambda}_t$ by adding the outer product of the vector of *close-to-open* factor returns, similar to including overnight stock returns in the realized covariance. For the idiosyncratic variances we also include the (active) residual overnight returns throughout the empirical analysis.

In the MFFM estimator in (4.14), we consider the same range of intra-day sampling frequencies for the factor covariance matrix and the idiosyncratic variances as used for the realized covariance estimator given in the previous subsection. Also, we include a conventional 'low-frequency' factor model where all parts of (4.14) are based on daily *close-to-close* returns.

4.4.3 Covariance matrix forecasts

We consider forecasts based on an exponentially weighted moving average (EWMA) scheme, motivated by the work of Foster and Nelson (1996) and Andreou and Ghysels (2002). In this framework, the covariance matrix forecast for day t, denoted $\Sigma_{t|t-1}$, is given by

$$\Sigma_{t|t-1} = \alpha \Sigma_{t-1|t-2} + (1-\alpha)\widehat{\Sigma}_{t-1}, \tag{4.17}$$

where the scalar α is a fixed decay parameter and $\widehat{\Sigma}_{t-1}$ is the covariance matrix estimate for day $t-1$ as given by either the RC estimator in (4.10) (or (4.11) in the minimum tracking error application) or the MFFM estimator in (4.14). We consider

[9] We have experimented with using $\hat{\beta}_t$ instead of $\hat{\beta}_{t-1}$, finding that this deteriorates the performance of the MFFM forecasts (although the differences are small).

several weighting schemes with $\alpha \in \{0.94, 0.75, 0.50, 0.25\}$. The value of 0.94 for α is the optimal decay parameter for daily data documented by RiskMetrics (see e.g. Zumbach (2006)). The use of smaller decay parameters allows us to examine the effects on the forecasting performance when putting more weight on more recent data. Smaller levels of α are also closer to our simulation study where in fact $\alpha = 0$. Further, using smaller values of α provides more insight in the quality of the covariance estimator itself rather than the 'smoothed' forecast. Less smoothing can be important also from an economic point of view, as it enables the forecasts to adjust more rapidly to important changes in variance and covariance dynamics, which for example occur at turning points between periods of high and low volatility.

We use the period from November 1, 2006 until December 31, 2006 as 'burn-in period' for the covariance dynamics in (4.17) and exclude these two months in the performance evaluations below. The out-of-sample period therefore runs from January 3, 2007 until April 30, 2009.

4.4.4 Equally-weighted portfolios

In our first forecasting exercise, we consider equally-weighted portfolios for the S&P500, S&P400 and S&P600 stock universes. As noted before, we only use the S constituents that were included in a single index during the complete sample period. For each universe the daily equally-weighted portfolio return is computed as $r_{p,t} = e'r_t$ where r_t is an $S \times 1$ vector of *close-to-close* returns on the individual stocks and e is the equal-weight vector with entries $1/S$. We obtain one-day ahead forecasts of the volatility of these equally-weighted portfolios as $\hat{\sigma}^2_{P,t|t-1} = e'\Sigma_{t|t-1}e$, using the MFFM- and RC-based covariance matrix forecasts from (4.17).

We evaluate the accuracy of the volatility forecasts in two ways. First, we run Mincer-Zarnowitz (MZ) regressions, in which the portfolio volatility proxy $\hat{\sigma}^2_{p,t|t}$ is regressed on a constant and one of the volatility forecasts, that is,

$$\hat{\sigma}^2_{p,t|t} = \gamma + \delta\hat{\sigma}^2_{p,t|t-1} + \varepsilon_t. \tag{4.18}$$

Here we use the squared daily return $r^2_{p,t}$ as the volatility proxy. Although this proxy is known to be noisy, at least it is unbiased. Obvious alternatives would be to use the RC or MFFM estimates of the covariance matrix for day t, but this might bias the

MZ regression towards one of the forecasts. Using the squared daily return avoids this issue.[10]

In addition we report results for forecast encompassing regressions where the squared daily return is regressed on the MFFM-based forecast $\hat{\sigma}^2_{p,MFFM,t|t-1}$ and a competing forecast $\hat{\sigma}^2_{p,X,t|t-1}$, that is,

$$r^2_{p,t} = \gamma + \delta_1 \hat{\sigma}^2_{p,MFFM,t|t-1} + \delta_2 \hat{\sigma}^2_{p,X,t|t-1} + \varepsilon_t. \tag{4.19}$$

These regressions can be used to obtain insights in how well the MFFM approach empirically competes with existing forecast methods. We consider two competing forecasts X, namely the RC at the same intraday sampling frequency as used for the MFFM and the daily factor model, denoted FM. Regression R^2's and coefficients are reported and statistically significant coefficients at the 5% level are displayed in bold fonts.

Figures 4.6–4.8 illustrates that the RC and MFFM provide very similar dynamics at the 5 min sampling frequency. In addition we observe that the estimates obtained with high-frequency data for the RC and MFFM are much more precise than their daily counterparts. The daily sample covariance and daily factor model estimates are "noisy". The equally-weighted portfolio volatility estimates are all plotted against the (scaled) daily absolute return.

We run Mincer-Zarnowitz and encompassing regression results with decay parameter $\alpha = \{0.94, 0.75, 0.50, 0.25, 0.00\}$. For space considerations we only report results for $\alpha = 0.94$ since the results for the other settings of α lead to similar conclusions. The only exception is that the performance of the daily factor model deteriorates rapidly for smaller α. This occurs because the daily factor model is based on only one observation per day and therefore require a longer history of covariance estimates to compete with the estimators based on higher sampling frequencies.

[10] Using the RC and the MFFM estimator based on a 5 min sampling frequency as the volatility proxy does not alter the main conclusions as reported here. The main difference is that we obtain higher regression R^2's that are about 10 to 15% higher than the R^2 for the daily squared return. In addition, we have considered the MZ regression using the absolute return as dependent variable (which then is regressed on a constant and the square root of $\hat{\sigma}^2_{p,t|t-1}$. This also results in higher R^2 values than those reported here (by about 5%), mostly because the absolute return is more robust to outliers. However, using this transformation of the variance does not lead to consistent forecast rankings when the forecast target is the conditional variance, see Patton (2011).

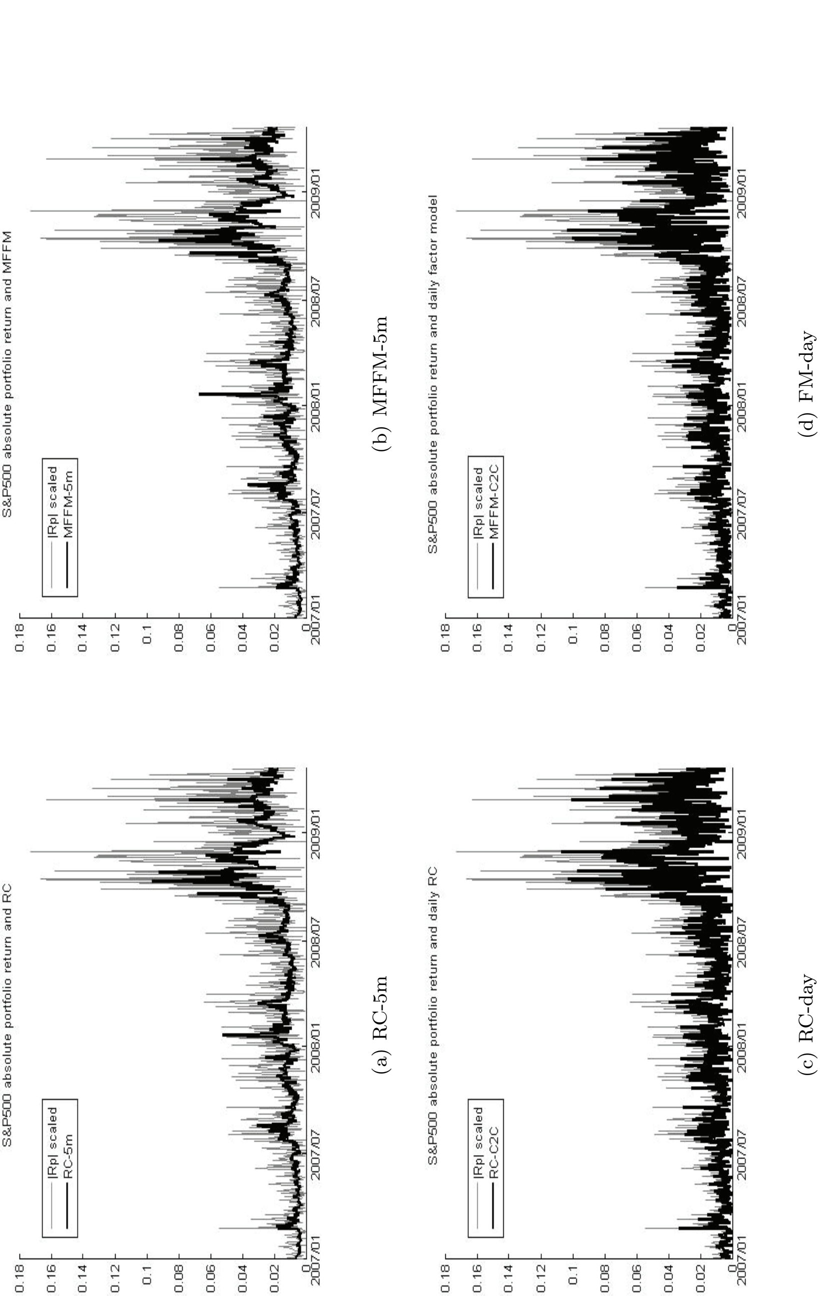
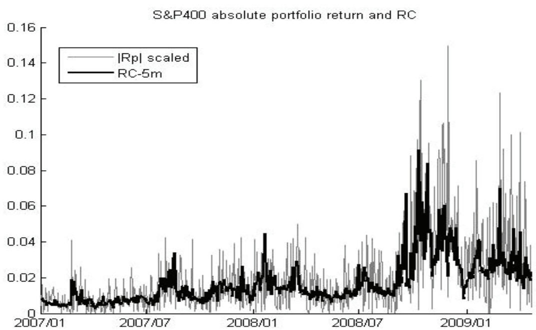
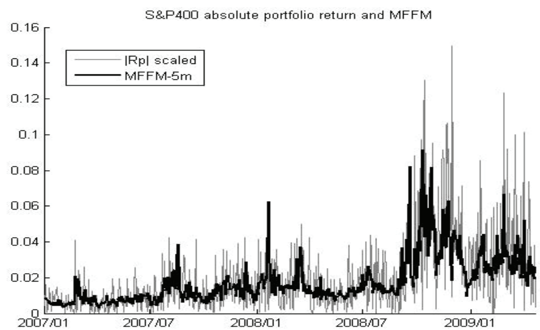


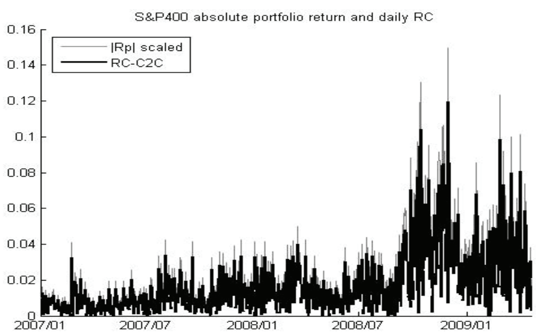
Figure 4.6: This Figure displays plots of the (scaled) absolute portfolio return of the equally-weighted S&P500 portfolio against the non-smoothed ($\alpha = 0$) volatility estimates obtained using the (a) realized covariance 5min (b) MFFM-5min (c) daily sample covariance and (d) daily factor model.



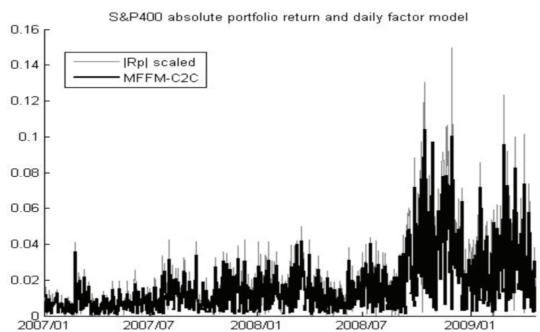
(a) RC-5m



(b) MFFM-5m

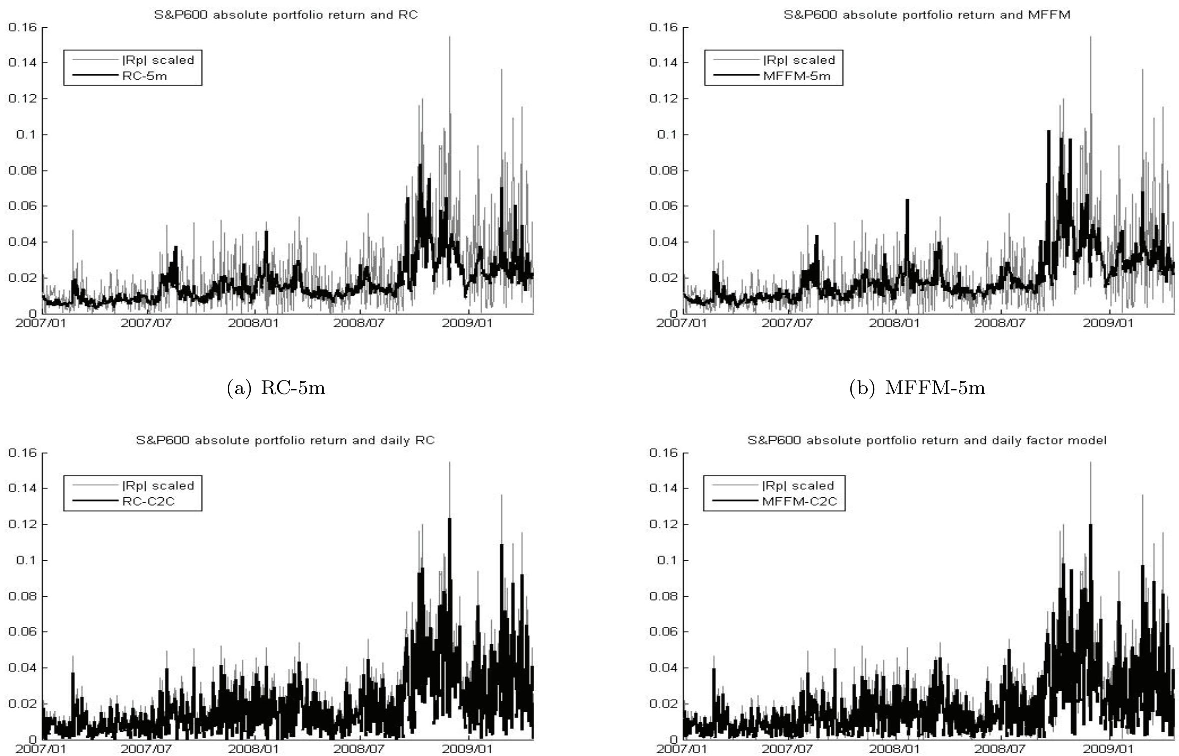


(c) RC-day



(d) FM-day

Figure 4.7: This Figure displays plots of the (scaled) absolute portfolio return of the equally-weighted S&P400 portfolio against the non-smoothed ($\alpha = 0$) volatility estimates obtained using the (a) realized covariance 5min (b) MFFM-5min (c) daily sample covariance and (d) daily factor model.



(a) RC-5m

(b) MFFM-5m

(c) RC-day

(d) FM-day

Figure 4.8: This Figure displays plots of the (scaled) absolute portfolio return of the equally-weighted S&P600 portfolio against the non-smoothed ($\alpha = 0$) volatility estimates obtained using the (a) realized covariance 5min (b) MFFM-5min (c) daily sample covariance and (d) daily factor model.

Table 4.2 summarizes the results for the S&P500 Mincer-Zarnowitz and encompassing regressions. Based on the Mincer-Zarnowitz regressions we find that the EWMA forecasts for the volatility of the equally-weighted portfolio have statistically significant coefficients. The constants, frequently interpreted as forecast bias, are statistically insignificant across all sampling frequencies. From the regression R^2's we learn that the results for RC and MFFM are very close, indicating that our factor structure indeed does a good job, and for both the RC and the MFFM we find that using high-frequency data improves the R^2 by about 3%.[11]

Table 4.2: S&P500 Portfolio volatility, Mincer-Zarnowitz and encompassing regressions

	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RC	**1.599**	**1.350**	**1.233**	**1.226**	**1.154**	**1.136**	**1.010**	**0.906**
R^2	0.241	0.242	0.244	0.246	0.244	0.245	0.229	0.205
Panel B: MFFM Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**1.490**	**1.366**	**1.299**	**1.311**	**1.253**	**1.233**	**1.120**	**1.036**
R^2	0.242	0.244	0.247	0.249	0.248	0.247	0.232	0.208
Panel C: MFFM + RC Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	1.236	4.353	4.179	3.745	3.591	2.091	1.841	2.230
RC	0.272	-2.968	-2.752	-2.296	-2.174	-0.795	-0.656	-1.056
R^2	0.242	0.246	0.250	0.252	0.250	0.248	0.232	0.209
Panel D: MFFM + FM Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
MFFM	**3.315**	**3.284**	**3.288**	**3.470**	**3.157**	**2.445**	**1.656**	
FM	-1.396	-1.609	-1.762	-1.904	-1.753	-1.141	-0.536	
R^2	0.257	0.264	0.271	0.277	0.272	0.261	0.234	

Note: This Table summarizes the results for Mincer-Zarnowitz and encompassing regressions using the daily squared portfolio return as unbiased proxy for the latent portfolio variance. The evaluation is based on 442 of the S&P500 constituents to forecast the variance of the equally-weighted portfolio one day ahead using EWMA covariance matrix forecasts with decay parameter $\alpha = 0.94$. Compared are the volatility forecasts generated with the MFFM, RC and the daily factor model. The out-of-sample period is Jan. 2007 – Apr. 2009. Coefficients that are statistically significant at the 5% level, based on Newey-West standard errors with 20 lags, are displayed in bold fonts.

[11] By lowering α the forecasting performance of the daily counter parts of the RC and MFFM deteriorates rapidly and regression coefficients become close to zero if we do not apply EWMA to generate forecasts due to the high variance of estimators based on daily data as displayed in Figure 4.6. The differences in R^2 between daily and high-frequency data when using non-smoothed estimates ($\alpha = 0$) are about 15%.

For the relatively liquid S&P500 encompassing regressions we find that the RC and MFFM forecasts do not encompass each other. The bias and loadings on the forecast have statistically insignificant Newey-West t-statistics at the 5% level. Using similar encompassing regressions, but now for the MFFM sampled at each intraday frequency against the daily factor model (FM), we find that the daily factor model forecasts are encompassed by the MFFM forecasts at each intraday sampling frequency. The improvement in regression the regression R^2 compared to regressing on MFFM only (see Panel B in Table 4.2) is also small.

Table 4.3: S&P400 Portfolio volatility, Mincer-Zarnowitz and encompassing regressions

	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RC	**2.105**	**1.594**	**1.369**	**1.365**	**1.257**	**1.220**	**1.064**	**0.915**
R^2	0.250	0.249	0.255	0.258	0.254	0.259	0.244	0.221
Panel B: MFFM Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**1.488**	**1.363**	**1.283**	**1.304**	**1.231**	**1.196**	**1.081**	**0.993**
R^2	0.259	0.259	0.262	0.266	0.266	0.266	0.252	0.230
Panel C: MFFM + RC Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**9.158**	**8.508**	**5.035**	**4.407**	**4.207**	2.469	2.362	3.429
RC	-11.051	**-8.535**	-4.070	-3.310	-3.119	-1.322	-1.287	-2.300
R^2	0.279	0.284	0.272	0.275	0.277	0.268	0.255	0.240
Panel D: MFFM + FM Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
MFFM	**2.355**	**2.333**	**2.324**	**2.527**	**2.457**	**1.923**	**1.355**	
FM	-0.630	-0.768	-0.881	-1.027	-1.087	-0.669	-0.271	
R^2	0.263	0.266	0.271	0.278	0.278	0.272	0.253	

Note: This Table summarizes the results for Mincer-Zarnowitz and encompassing regressions using the daily squared portfolio return as unbiased proxy for the latent portfolio variance. The evaluation is based on 342 of the S&P400 constituents to forecast the variance of the equally-weighted portfolio one day ahead using EWMA covariance matrix forecasts with decay parameter $\alpha = 0.94$. The out-of-sample period is Jan. 2007 – Apr. 2009. Coefficients that are statistically significant at the 5% level, based on Newey-West standard errors with 20 lags, are displayed in bold fonts.

For the S&P400 mid-cap universe Table 4.3 summarizes the forecast regression results. Using Mincer-Zarnowitz regressions we find that the forecasts based on RC and MFFM are statistically significant at each sampling frequency and the forecast bias is not significant. The regression R^2's for the MFFM regressions are higher than for the RC regressions. In the encompassing regression results for MFFM and RC (Panel C) we observe that, at high sampling frequencies, between 15 sec and 30 min, the MFFM forecasts encompass the RC forecasts if we use the squared daily return as proxy. Differences increase by moving from the relatively liquid S&P500 stocks to the less liquid S&P400 stocks where non-synchronicity plays a more important role. In line with the results for the S&P500 we find for the S&P400 universe that the daily factor model forecasts are encompassed by MFFM and this holds at every intraday sampling frequency, and adding the daily factor model forecasts to MFFM forecasts only improves the regression R^2 by about a half percent.

When we move to the relatively illiquid S&P600 constituents we observe in Table 4.4 that for the Mincer-Zarnowitz regressions the forecasts of RC and MFFM are significant at every frequency and the forecast bias is not. Similar to the S&P400 results we find that the regression R^2 for MFFM is higher than for RC. Using encompassing regressions we find that the MFFM forecasts are favored over the RC forecasts at very high frequencies between 15s and 1m. Consistent with the results for the S&P500 and S&P400 the MFFM forecasts obtained using intraday sampling encompass the factor model based on daily data.

4.4.5 Minimum tracking error portfolios

Given the one day ahead EWMA forecasts of the covariance matrices we construct minimum TE portfolios by calculating the standard fully-invested minimum variance portfolios (when using the *active* covariance matrix as we do here, then the minimum TE portfolio is the minimum variance portfolio):

$$w_t = \frac{\Sigma_{t|t-1}^{-1} e}{e' \Sigma_{t|t-1}^{-1} e} \tag{4.20}$$

where e is a $S \times 1$ vector of ones and Σ_t is the EWMA conditional covariance matrix forecast of RC or MFFM. The daily minimum TE portfolio returns are obtained by computing $R_{Pt} = w_t' r_t$ where r_t is the vector of daily stock returns. We calculate the *ex-post* tracking error using daily returns $\text{TE} = \text{Std}(R_P - R_M)$ and compare the results for the RC and the MFFM.

In this application we keep track of the daily turnover in the portfolio weights w_t which is directly associated with the transaction costs that an investor faces who wishes to re-balance his or her portfolio daily. We compute turnover by summing the absolute daily weight changes over the stock names,

$$TO_t = |w_t - w_{t-1}|'e. \tag{4.21}$$

We expect that a covariance estimator that is well-conditioned and numerically stable will result in smaller daily portfolio turnover. The daily turnover will also be related to the decay parameter α in (4.17) which is used to generate EWMA forecasts. A large decay parameter implies that more weight is assigned to historical estimates whereas a smaller decay parameter corresponds to assigning more weight to the most recent estimate(s). More weight on historical estimates will make an estimator more stable and cause less portfolio turnover but on the other hand, recent shifts in for example market volatility will be picked up at a slower pace.

Table 4.4: S&P600 Portfolio volatility, Mincer-Zarnowitz and encompassing regressions

	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RC	**2.878**	**2.080**	**1.656**	**1.590**	**1.406**	**1.306**	**1.125**	**0.916**
R^2	0.229	0.232	0.235	0.237	0.236	0.240	0.225	0.209
Panel B: MFFM Mincer-Zarnowitz								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**1.188**	**1.191**	**1.167**	**1.198**	**1.114**	**1.068**	**0.957**	**0.924**
R^2	0.241	0.241	0.241	0.244	0.243	0.241	0.228	0.216
Panel C: MFFM + RC Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MFFM	**2.226**	**3.397**	2.275	1.875	1.475	0.686	0.682	**2.254**
RC	-2.601	-3.942	-1.599	-0.918	-0.466	0.475	0.329	-1.345
R^2	0.244	0.247	0.243	0.245	0.243	0.242	0.228	0.220
Panel D: MFFM + FM Encompassing								
c	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
MFFM	**1.523**	**1.543**	**1.442**	**1.576**	**1.497**	**1.184**	0.783	
FM	-0.284	-0.298	-0.239	-0.322	-0.348	-0.111	0.180	
R^2	0.242	0.242	0.242	0.246	0.245	0.241	0.229	

Note: This Table summarizes the results for Mincer-Zarnowitz and encompassing regressions using the daily squared portfolio return as unbiased proxy for the latent portfolio variance. The evaluation is based on 491 of the S&P600 constituents to forecast the variance of the equally-weighted portfolio one day ahead using EWMA covariance matrix forecasts with decay parameter $\alpha = 0.94$. The out-of-sample period is Jan. 2007 – Apr. 2009. Coefficients that are statistically significant at the 5% level, based on Newey-West standard errors with 20 lags, are displayed in bold fonts.

4.4.6 Minimum tracking error results

Table 4.5 illustrates the performance in terms of annualized minimum tracking errors for the S&P500 large caps. Consistent with the simulation results we find that the MFFM covariance matrix estimator is remarkably robust across sampling frequencies indicating that, in contrast to RC, the factor covariance matrix can be estimated at very high frequencies as the level of market microstructure noise and non-synchronicity in the factors is relatively small compared to individual stocks. At almost each of the considered sampling frequencies and forecast weights α, the MFFM produces better results than RC. However, for the relatively liquid S&P500 universe the RC competes with the MFFM if we use a sampling frequency between 15s and 15m combined with $\alpha = 0.94$ for RC but deteriorates rapidly by putting more weight on the most recent estimates (lower α's). The difference between the MFFM and RC are small when we choose the best sampling frequency and forecasts weights for RC, but the differences are substantial on average across these settings. The covariance matrices considered here have a dimension of 442 and we find that at sampling frequencies of 30min and lower the RC is not well-conditioned and therefore not invertible, we indicate this with "NA". The naïvely diversified equally-weighted portfolio, advocated recently by DeMiguel et al. (2009), achieves a tracking error equal to 0.099 and is outperformed by the MFFM in each of the parameter settings and by most parameter settings for the RC given that these settings result in an invertible covariance matrix forecast.

Important differences in numerical stability of the covariance matrix forecasts are exemplified by the very large differences in portfolio turnover. At most of the sampling frequencies the difference is at least 8 times larger. This indicates, as noted by Fan et al. (2008), that using (sample) realized covariances for portfolio optimization can be "tricky" for vast dimensional portfolios. In contrast, due to its factor structure and the use of relatively liquid factors, the MFFM delivers exceptionally small levels of turnover associated with tracking errors that are at par with the best results for the realized covariance and outperform the realized covariance at all other settings without having to resort to putting a lot of weight on historical estimates. In fact the MFFM is found to be relatively insensitive to the choice of α and the sampling frequency. It is interesting to observe that on average the MFFM tracking errors are fairly constant across sampling frequencies and decay parameters. This is due to the factor structure which ensures stability of the covariance matrix. The level of turnover, however, does depend on the sampling frequency and decay parameter. The highest sampling frequencies produce very small levels of turnover because the (factor) covariance estimates are very precise. Applying a higher decay parameter, i.e. less weight on recent data, further smooths the covariance matrix and therefore reduces turnover. The lower sampling frequencies produce less precise (factor) covariance estimates and therefore higher levels of turnover. For lower sampling frequencies the covariance estimates are accurate on average due to the factor structure but they are less precise than when higher sampling frequencies are used.

Table 4.5: Annualized tracking errors S&P500 (large cap) universe

α	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC tracking error								
0.94	0.060	0.058	0.056	0.060	0.066	0.211	NA	NA
0.75	0.062	0.062	0.061	0.075	0.088	0.178	NA	NA
0.50	0.064	0.068	0.077	0.102	0.122	NA	NA	NA
0.25	0.068	0.077	0.099	0.127	NA	NA	NA	NA
Panel B: MFFM tracking error								
0.94	0.059	0.059	0.058	0.058	0.059	0.059	0.058	0.058
0.75	0.059	0.058	0.058	0.058	0.058	0.059	0.058	0.058
0.50	0.059	0.058	0.058	0.057	0.058	0.058	0.058	0.059
0.25	0.059	0.058	0.058	0.057	0.058	0.058	0.058	0.064
Panel C: RC turnover								
0.94	0.236	0.292	0.357	0.458	0.582	7.922	NA	NA
0.75	0.844	1.041	1.379	1.875	2.435	6.842	NA	NA
0.50	1.728	2.222	3.115	4.229	5.450	NA	NA	NA
0.25	2.861	3.960	5.748	7.668	NA	NA	NA	NA
Panel D: MFFM turnover								
0.94	0.028	0.032	0.037	0.042	0.046	0.051	0.057	0.079
0.75	0.089	0.101	0.120	0.141	0.158	0.182	0.209	0.297
0.50	0.162	0.185	0.225	0.269	0.307	0.361	0.420	0.603
0.25	0.239	0.275	0.339	0.414	0.478	0.573	0.675	0.986

Note: This table reports the *ex-post* annualized minimum tracking errors in percentages and the daily average portfolio turnover using 442 of the S&P500 constituents. The results are based on RC and MFFM to forecast the active covariance matrix one day ahead using EWMA forecasts over the sample period 3/1/2007 - 30/4/2009 with decay parameter α. For the MFFM we use a 12-factor model specification (size, value, and 10 industry factors). The "NA" entries indicate that the RC is not-invertible at certain combinations of sampling frequencies and weighting schemes.

For the S&P400, see Table 4.6, we decrease the naïve equally-weighted portfolio tracking error of 9.3% to 8.3% with RC and this result depends heavily on the forecast weighting scheme and sampling frequency. Using the MFFM further decreases the tracking error to 7.8% with results being robust. As expected, the tracking errors have increased for the S&P400 mid caps compared to the S&P500 large caps, see also Table 1 for the average number of trades per day in each S&P universe. Higher levels of non-synchronicity and microstructure noise in individual stocks explain this result. Similar to the portfolio turnover results for the S&P500 universe we find that the RC portfolios cause a daily turnover which is at least a factor 10 times larger than the turnover in the MFFM portfolios.

Table 4.6: Annualized tracking error S&P400 (mid cap) universe

α	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC tracking error								
0.94	0.089	0.085	0.083	0.087	0.091	0.275	NA	NA
0.75	0.095	0.095	0.096	0.114	0.122	0.305	NA	NA
0.50	0.102	0.104	0.118	0.156	0.174	NA	NA	NA
0.25	0.107	0.120	0.154	0.211	NA	NA	NA	NA
Panel B: MFFM tracking error								
0.94	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.078
0.75	0.078	0.078	0.078	0.078	0.078	0.077	0.078	0.079
0.50	0.078	0.078	0.078	0.078	0.078	0.078	0.078	0.081
0.25	0.078	0.078	0.078	0.078	0.079	0.078	0.079	0.086
Panel C: RC turnover								
0.94	0.251	0.341	0.411	0.484	0.597	3.397	NA	NA
0.75	0.915	1.183	1.520	1.939	2.427	6.485	NA	NA
0.50	1.869	2.429	3.403	4.364	5.357	NA	NA	NA
0.25	3.071	4.197	6.217	8.000	NA	NA	NA	NA
Panel D: MFFM turnover								
0.94	0.023	0.025	0.027	0.031	0.033	0.037	0.040	0.075
0.75	0.083	0.089	0.101	0.115	0.126	0.141	0.160	0.282
0.50	0.160	0.172	0.198	0.229	0.254	0.291	0.335	0.580
0.25	0.245	0.265	0.309	0.363	0.407	0.475	0.560	0.963

Note: This table reports the *ex-post* annualized minimum tracking errors in percentages and the daily average portfolio turnover using 342 of the S&P400 constituents. The tracking errors are based on RC and MFFM to forecast the active covariance matrix one day ahead using EWMA forecasts over the sample period 3/1/2007 - 30/4/2009 with decay parameter α. For the MFFM we use a 12-factor model specification (size, value, and 10 industry factors). The "NA" entries indicate that the RC is not-invertible at certain combinations of sampling frequencies and weighting schemes.

For the S&P600 small caps, where the level of non-synchronicity plays a more important role than for mid- and large-caps, we find larger tracking errors when using the RC because it is sensitive to market microstructure frictions and the increased portfolio dimension. The tracking errors for the RC are in fact larger, for every combination of sampling frequency and forecast weighting scheme, than that of the naïve $1/N$ portfolio which achieves a tracking error of 8.9%. The MFFM, however, achieves smaller tracking errors than it achieves for the S&P400 mid-caps, indicating that its performance is not compromised by the illiquidity of the S&P600 universe. The MFFM comfortably decreases the best RC tracking error, being 9.1%, to 7.0%

and the MFFM easily outperforms the naïve portfolio for all combinations of sampling frequencies and forecast weighting schemes. In line with the turnover results for the S&P500 and S&P400 we find that the RC portfolios have a turnover that is 12 times larger.

Table 4.7: Annualized tracking error S&P600 (small-cap) universe

α	15s	1m	5m	15m	30m	65m	130m	C2C
Panel A: RC tracking error								
0.94	0.098	0.091	0.091	0.095	0.105	NA	NA	NA
0.75	0.109	0.114	0.126	0.157	0.200	NA	NA	NA
0.50	0.121	0.135	0.183	0.249	NA	NA	NA	NA
0.25	0.136	0.162	NA	NA	NA	NA	NA	NA
Panel B: MFFM tracking error								
0.94	0.073	0.072	0.072	0.071	0.071	0.071	0.071	0.070
0.75	0.072	0.072	0.071	0.070	0.070	0.071	0.071	0.070
0.50	0.072	0.071	0.071	0.070	0.070	0.070	0.071	0.070
0.25	0.072	0.071	0.071	0.070	0.070	0.071	0.072	0.077
Panel C: RC turnover								
0.94	0.367	0.503	0.658	0.806	1.018	NA	NA	NA
0.75	1.393	1.885	2.649	3.389	4.219	NA	NA	NA
0.50	2.901	4.043	6.035	7.636	NA	NA	NA	NA
0.25	4.842	7.230	NA	NA	NA	NA	NA	NA
Panel D: turnover								
0.94	0.030	0.033	0.039	0.044	0.049	0.056	0.062	0.085
0.75	0.090	0.100	0.121	0.145	0.164	0.192	0.221	0.308
0.50	0.164	0.181	0.224	0.274	0.315	0.374	0.435	0.613
0.25	0.244	0.270	0.338	0.421	0.490	0.589	0.690	1.000

Note: This table reports the *ex-post* annualized minimum tracking errors in percentages and the daily average portfolio turnover using 491 of the S&P600 constituents. The tracking errors are based on RC and MFFM to forecast the active covariance matrix one day ahead using EWMA forecasts over the sample period 3/1/2007 - 30/4/2009 with decay parameter α. For the MFFM we use a 12-factor model specification (size, value, and 10 industry factors). The "NA" entries indicate that the RC is not-invertible at certain combinations of sampling frequencies and weighting schemes.

Note that outperforming the equally-weighted portfolio is not necessarily an easy task. DeMiguel et al. (2009) analyze various advanced methods consisting of Bayesian estimation, shrinkage, robust allocation etc. and find that none of the 14 models they implement can consistently outperform the $1/N$ portfolio. Hence, the fact that the MFFM consistently outperforms the $1/N$ and RC portfolios is encouraging support. Further, the results in Madhavan and Yang (2003) illustrate that using the sample (realized) covariance matrix for unrestricted optimization, results in a performance that is worse than the equally-weighted portfolio.

4.5 Conclusion

Recently there has been great interest in the use of high-frequency data to estimate variances and covariances. The advantage is that the use of high-frequency data results in more accurate covariance estimates, but on the other hand it also brings problems such as microstructure noise which reduces the efficiency of covariance estimators based on intraday data and non-synchronous trading leading to covariance estimates being biased towards zero. What so far has been lacking is to bring the merits of high-frequency data to factor models. With the introduction of exchange-traded funds important factors are now traded much more actively than individual stocks. For example the S&P500 ETFs (Spiders) have on average traded 18 times more frequently than the average individual stock in the S&P500. In this study we have proposed the Mixed Frequency Factor Model. In particular we can use ultra high-frequency data for ETFs to obtain a very accurate estimate of the factor covariance matrix, as prices are observed essentially free of noise. We use daily data to estimate the factor loadings conservatively to avoid problems inherent in the use high-frequency data for illiquid stocks and non-synchronicity biases between the returns on factors and stocks. Furthermore we take advantage of the facts that factor models can easily be applied to vast numbers of assets and that covariance matrices from factor models are less prone to error maximization in portfolio construction problems. Using Mincer-Zarnowitz and encompassing regressions we find that the MFFM portfolio volatility forecasts improve upon the daily factor and realized covariance forecasts when the forecasts are ranked on R^2 and as indicated by the positive weights on the MFFM versus negative weights on the RC and daily factor model. Adding a RC or daily factor model forecast to a MFFM forecast only improves the regression R^2 marginally. In a minimum tracking error application we reduce the tracking errors by using the MFFM rather than RC for computing the covariance matrix. The differences between RC and MFFM increase with the level of non-synchronicity between individual stocks, i.e. we find a larger difference when considering the S&P600 small caps than when we consider the S&P500 large caps. The RC outperforms the naïvely diversified equally-weighted $1/N$ portfolios when considering large- and mid-caps but fails by a substantial margin for the illiquid S&P600 small caps. The MFFM comfortably outperforms the $1/N$ portfolios regardless of the universe considered. For realized covariance the results in the tracking error applications depend severely on the sampling frequency and the weighting scheme applied to the past daily covariance matrices. In contrast, the performance of the MFFM is robust across sampling frequencies and weighting schemes and consistently outperforms RC and the naïve $1/N$ portfolios.

4.A Proofs

Proof of Theorem 4.2.1 Using the notation $\widehat{X} = X + X^{\varepsilon}$, we have for $i \neq j$:

$$\widehat{\gamma}_{ij} = \beta_i'\Lambda\beta_j + \beta_i'\Lambda\beta_j^{\varepsilon} + \beta_i'\Lambda^{\varepsilon}\beta_j + \beta_i'\Lambda^{\varepsilon}\beta_j^{\varepsilon} + \beta_i^{\varepsilon\prime}\Lambda\beta_j + \beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon} + \beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j + \beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j^{\varepsilon}.$$

Assumption (i) implies $\beta_i^{\varepsilon} \perp \beta_j^{\varepsilon}$ so that $E(\beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon}) = 0$. All other terms, except $\beta_i'\Lambda\beta_j$, are zero in expectation by assumptions (ii-iv). Hence, we have unbiasedness. To compute the variance of this estimator, note that all terms are mutually uncorrelated, so that the variance of the sum is the sum of the variances.

$$\begin{aligned}
V(\beta_i'\Lambda\beta_j^{\varepsilon}) &= \beta_i'\Lambda\Sigma_{\beta,j}\Lambda'\beta_i \\
V(\beta_i'\Lambda^{\varepsilon}\beta_j) &= E(\beta_i'\Lambda^{\varepsilon}\beta_j\beta_j'\Lambda^{\varepsilon\prime}\beta_i) = E(\mathrm{tr}(\beta_i\beta_i'\Lambda^{\varepsilon}\beta_j\beta_j'\Lambda^{\varepsilon\prime})) = g(\beta_i\beta_i', \beta_j\beta_j', \Phi) \\
V(\beta_i'\Lambda^{\varepsilon}\beta_j^{\varepsilon}) &= E(\beta_i'\Lambda^{\varepsilon}\beta_j^{\varepsilon}\beta_j^{\varepsilon\prime}\Lambda^{\varepsilon\prime}\beta_i) = E(\beta_i'\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime}\beta_i) = E(\mathrm{tr}(\beta_i\beta_i'\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime})) \\
&= g(\beta_i\beta_i', \Sigma_{\beta,j}, \Phi) \\
V(\beta_i^{\varepsilon\prime}\Lambda\beta_j) &= \beta_j'\Lambda'\Sigma_{\beta,i}\Lambda\beta_j \\
V(\beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon}) &= E(\beta_i^{\varepsilon\prime}\Lambda\beta_j^{\varepsilon}\beta_j^{\varepsilon\prime}\Lambda'\beta_i^{\varepsilon}) = E(\beta_i^{\varepsilon\prime}\Lambda\Sigma_{\beta,j}\Lambda'\beta_i^{\varepsilon}) = \mathrm{tr}(\Sigma_{\beta,i}\Lambda\Sigma_{\beta,j}\Lambda') \\
V(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j) &= E(\beta_j'\Lambda^{\varepsilon}\Sigma_{\beta,i}\Lambda^{\varepsilon\prime}\beta_j) = E(\mathrm{tr}(\beta_j\beta_j'\Lambda^{\varepsilon}\Sigma_{\beta,i}\Lambda^{\varepsilon\prime})) = g(\beta_j\beta_j', \Sigma_{\beta,i}, \Phi) \\
V(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j^{\varepsilon}) &= E(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\beta_j^{\varepsilon}\beta_j^{\varepsilon\prime}\Lambda^{\varepsilon\prime}\beta_i^{\varepsilon}) = E(\beta_i^{\varepsilon\prime}\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime}\beta_i^{\varepsilon}) = E(\mathrm{tr}(\Sigma_{\beta,i}\Lambda^{\varepsilon}\Sigma_{\beta,j}\Lambda^{\varepsilon\prime})) \\
&= g(\Sigma_{\beta,i}, \Sigma_{\beta,j}, \Phi)
\end{aligned}$$

All terms involving Λ^{ε} are of the form $E(\mathrm{tr}(AZBZ))$ where A, B, and Z are square symmetric matrices of equal dimension with A and B fixed and Z random with $E(\mathrm{vech}(Z)\mathrm{vech}(Z)') = \Phi$. Define $\overline{A} = AZ$ and $\overline{B} = BZ$ with

$$\overline{A}_{ij} = \sum_k A_{ik}Z_{kj} \qquad \text{and} \qquad \overline{B}_{ij} = \sum_m B_{im}Z_{mj}.$$

Then

$$\begin{aligned}
E(\mathrm{tr}(AZBZ)) &= \mathrm{tr}(E(\overline{A}\ \overline{B})) = \sum_{i,j} E(\overline{A}_{ij}\overline{B}_{ji}) = \sum_{i,j,k,m} A_{ik}B_{jm}E(Z_{kj}Z_{mi}) \\
&= \sum_{i,j,k,m} A_{ik}B_{jm}\Phi_{f(k,j),f(m,i)}.
\end{aligned}$$

■

Proof of Corollary 4.2.2 Given the assumptions, we have $\Sigma_{\beta,i} = \frac{1}{T}\sigma_i^2\Lambda^{-1}$ (asymptotically). Thus, $\beta_j'\Lambda\Sigma_{\beta,i}\Lambda'\beta_j = \frac{1}{T}\sigma_i^2\beta_j'\Lambda\beta_j$ and $\mathrm{tr}(\Sigma_{\beta,i}\Lambda\Sigma_{\beta,j}\Lambda') = \frac{1}{T^2}\sigma_i^2\sigma_j^2\mathrm{tr}(I_K) =$

$\frac{K}{T^2}\sigma_i^2\sigma_j^2$. Combining this, gives term A. For term B, note that

$$\begin{aligned} g(A,B,\Phi) &= \sum_{m,n,p,q}^{N} A_{mp}B_{nq}E((\widehat{\Lambda}_{pn}-\Lambda_{pn})(\widehat{\Lambda}_{qm}-\Lambda_{qm})) \\ &= \sum_{m,n,p,q}^{N} A_{mp}B_{nq}(E(\widehat{\Lambda}_{pn}\widehat{\Lambda}_{qm})-\Lambda_{pn}\Lambda_{qm}) \\ &= \frac{1}{M}\sum_{m,n,p,q}^{N} A_{mp}B_{nq}(\Lambda_{pq}\Lambda_{nm}+\Lambda_{pm}\Lambda_{nq}) \end{aligned}$$

using that for a multivariate normal random variable x with characteristic function $\ln\phi(\xi) = -\xi'\Sigma\xi/2$ we have

$$E(\widehat{\sigma}_{mn}\widehat{\sigma}_{pq}) = \sigma_{mn}\sigma_{pq} + \frac{\sigma_{mp}\sigma_{nq}+\sigma_{mq}\sigma_{np}}{M}$$

where

$$\widehat{\sigma}_{mn} \equiv \frac{1}{M}\sum_{i=1}^{M} x_i^{(m)}x_i^{(n)}$$

■

Nederlandse samenvatting (Summary in Dutch)

Dit proefschrift bestaat uit drie studies op het gebied van het meten en voorspellen van de volatiliteit van rendementen op financiële instrumenten met behulp van intradag data. Volatiliteit is een maatstaf voor de beweeglijkheid van financiële instrumenten zoals aandelen, staatsobligaties en goud. De volatiliteit op financiële markten is een indicator voor risico en daarmee een centraal onderwerp binnen het financieel-economische vakgebied. In ieder hoofdstuk worden nieuwe en originele schatters en methoden ontwikkeld voor het schatten van de (multivariate) volatiliteit voor financiële markt rendementen. Deze nieuwe methoden en technieken worden uitgebreid getest in gesimuleerde omgevingen die zo goed mogelijk de realiteit benaderen. Daarnaast wordt het gebruik van deze nieuwe methoden en technieken, en de daarmee samenhangende voordelen, geïllustreerd met behulp van empirische toepassingen.

Traditioneel wordt de volatiliteit geschat aan de hand van rendementen die berekend zijn met behulp van koersen die gemeten zijn op lage frequentie. Hierbij kan gedacht worden aan het gebruik van maandelijkse, wekelijkse of dagelijkse rendementen. Een belangrijk kenmerk van volatiliteit is dat het een dynamisch en dus tijdsvariërend proces is. Mandelbrot (1963) merkte al op dat periodes van hoge (lage) volatiliteit veelal gevolgd worden door periodes van hoge (lage) volatiliteit. Deze persistentie in het volatiliteitsproces verklaart deels de relatief goede voorspelbaarheid van de volatiliteit van aandelenrendementen. Deze voorspelbaarheid is met name aanzienlijk wanneer ze vergeleken wordt met de voorspelbaarheid van rendementen die, consistent met economische theorie, relatief laag of verwaarloosbaar is.

Merton (1980) merkte op dat de volatiliteit van rendementen over een vaste periode kan worden gekwantificeerd met hoge precisie mits gedurende die periode een voldoende groot aantal sub-periode rendementen beschikbaar is. Onder de aanname

van het afwezig zijn van marktfricties geldt statistisch gezien dat wanneer een groter aantal intra-periode observaties beschikbaar is, de schatting van de volatiliteit preciezer wordt. Dit is hoe tegenwoordig de dagelijkse volatiliteit met gebruik van intradag koersdata kan worden berekend. Door het sommeren van gekwadrateerde rendementen die intradag worden waargenomen kan de dagelijkse volatiliteit zeer precies worden geschat. Dit is het zogenaamde concept van 'gerealiseerde volatiliteit' en staat in sterk contrast tot populaire volatiliteit modellen als GARCH waarin de volatiliteit helaas niet observeerbaar is en dus parametrisch geschat moet worden met meetfouten.

Het concept van het meten van gerealiseerde volatiliteit op basis van intradag koersdata heeft sinds de studies van onder meer Andersen en Bollerslev (1998), waarin het idee is geformaliseerd, in hoog tempo aan populariteit gewonnen. Dit proefschrift omvat drie studies naar het gebruik van hoog frequente intradag koersdata voor het ontwikkelen van preciezere financiële risicomaatstaven en voorspellingen daarvan.

Een centraal en bijzonder relevant onderwerp in deze literatuur en in dit proefschrift is het ontwikkelen van maatstaven die robuust zijn voor realistische marktfricties. Bij dit soort fricties kan gedacht worden aan het feit dat transactieprijzen op en neer springen tussen bied- en laat-prijzen, zelfs wanneer de onderliggende 'efficiënte' prijs constant is. In een multivariate context is het feit dat transacties in verschillende financiële instrumenten asynchroon plaatsvinden een complicerende factor. Deze frictie leidt er in de regel toe dat wanneer men covarianties of correlaties schat met behulp van op hoge frequentie waargenomen koersdata, de samenhang tussen financiële instrumenten verkeerd zal worden ingeschat omdat deze maatstaven dan bloot staan aan een afwijking die de meting richting nul drukt. Dit leidt in de regel tot een potentieel gevaarlijke onderschatting van risicomaatstaven voor vermogensportefeuilles.

In hoofdstuk 2 introduceren wij een verbeterde methodologie om schattingen en voorspellingen van volatiliteit, gemeten met behulp van de Realized Range, te corrigeren voor de effecten van marktfricties. Wanneer de Realized Range gebruik maakt van intradag hoog en laag prijzen in bijvoorbeeld intervallen van vijf minuten, dan zal het feit dat de prijzen heen en weer bewegen tussen bied- en laat prijzen zorgen voor een overschatting van de volatiliteit. De verbetering die wij voorstellen is gebaseerd op het relaxeren van een belangrijke impliciete aanname in de studie van Christensen et al. (2009). Aan de hand van gesimuleerde en empirische voorspellingen van volatiliteit illustreren wij de voordelen die samenhangen met het gebruik van de door ons voorgestelde methodiek, die onder meer gebaseerd is op simulatie en sortering, om te corrigeren voor het bestaan van marktfricties.

In hoofdstuk 3 breiden wij de zogenaamde univariate Realized Range methodiek uit naar een multivariate context. De Realized Range is gebaseerd op het gebruik van hoog- en laag-prijzen die worden waargenomen gedurende intradag intervallen. De dagelijkse Range als risicomaatstaf is in multivariate context uitgebreid door Brandt en Diebold (2006). Daarnaast introduceren Martens en Van Dijk (2007) en Christensen en Podolskij (2007) het gebruik van intradag hoog-laag prijsverschillen voor om in univariate context volatiliteit preciezer te kunnen schatten. Het combineren van beide ideeën resulteert in de nieuwe covariantieschatter, de Realized Co-Range. Deze schatter wordt uitgebreid bestudeerd in een gesimuleerde omgeving met inachtneming van de meest relevante marktfricties. Naast het simulatie gedeelte bevat deze studie ook een empirische toepassing voor het dynamisch alloceren van vermogen over aandelen-, staatsobligatie- en goud-futures contracten.

In hoofdstuk 4 introduceren wij een nieuwe covariantie-matrix schatter die gebruik maakt van hoge-frequentie koersdata. De meeste studies in de context van multivariate schattingen van volatiliteit op basis van intradag koersdata, zijn niet verder gekomen dan methoden te ontwikkelen die gebruikt kunnen worden voor portefeuilles bestaande uit een zeer beperkt aantal instrumenten zoals veelal slechts twee of hooguit dertig aandelen. Dit is te wijten aan de theoretische eigenschappen van deze schatters die de maximale dimensie van een covariantie- of correlatie-matrix van rendementen drastisch beperken.

Door het gebruik van intradag data te combineren met het gebruik van factormodellen ontwikkelen wij een zogenaamd Mixed-Frequency Factor Model. Dit model combineert het gebruik van ultra-hoge frequentie data voor het zeer precies schatten van de covariantie-matrix van factorrendementen, met het gebruik van dagelijkse aandelen- en factorrendementen om de factorcoëfficiënten van aandelen te schatten. Voor de factoren wordt gebruikgemaakt van relatief zeer liquide Exchange Traded Funds (ETFs), waarvoor de impact van marktfricties in essentie verwaarloosbaar is. Voor het schatten van de factorcoëfficiënten van individuele aandelen wordt gekozen voor het gebruik van rendementen op lagere frequentie om zo de ongewenste effecten van voor aandelen bestaande marktfricties en asynchroniteit te omzeilen. Dit resulteert in een statistisch gezien zeer efficiënte schatter voor de covariantie-matrix van aandelenrendementen. Een belangrijk praktisch probleem dat vooralsnog onopgelost was binnen de literatuur over het schatten van hoogdimensionale covariantie matrices met behulp van intradag data wordt in de voorgestelde methodiek succesvol omzeild door het gebruik van een factor model. De voordelen van deze nieuwe schatter worden onderbouwd in theorie en in een uitgebreide simulatie-analyse waarin marktfricties in acht worden genomen bestudeerd. Dit resulteert in bemoedigende resultaten, te meer de schatter ook empirisch succesvol is geïmplementeerd voor portefeuilles bestaande uit honderden aandelen.

Bibliography

Aït-Sahalia, Y. and Mancini, L.: 2008, Out of sample forecasts of quadratic variation, *Journal of Econometrics* **147**, 17–33.

Aït-Sahalia, Y., Mykland, P. and Zhang, L.: 2005, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* **18**, 351–416.

Andersen, T. and Bollerslev, T.: 1998, Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review* **39**(4), 885–905.

Andersen, T., Bollerslev, T., Christoffersen, P. and Diebold, F.: 2006a, Volatility and correlation forecasting, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Elsevier, North-Holland, pp. 777–878.

Andersen, T., Bollerslev, T., Diebold, F. and Labys, P.: 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T., Bollerslev, T., Diebold, F. and Labys, P.: 2003, Modeling and forecasting realized volatility, *Econometrica* **71**, 529–626.

Andersen, T., Bollerslev, T., Diebold, F. and Vega, C.: 2007, Real-time price discovery in stock, bond and foreign exchange markets, *Journal of International Economics* **73**, 251–277.

Andersen, T., Bollerslev, T. and Meddahi, N.: 2011, Realized volatility forecasting and market microstructure noise, *Journal of Econometrics* **160**(1), 220–234.

Andreou, E. and Ghysels, E.: 2002, Rolling-sample volatility estimators: Some new theoretical, simulation and empirical results, *Journal of Business and Economic Statistics* **20**, 363–376.

Awartani, B., Corradi, V. and Distaso, W.: 2009, Testing market microstructure effects with an application to the Dow Jones Industrial Average, *Journal of Business and Economic Statistics* **27**, 251–265.

Bandi, F. M. and Russell, J. R.: 2006, Separating microstructure noise from volatility, *Journal of Financial Economics* **79**, 655–692.

Bandi, F. M. and Russell, J. R.: 2008a, Microstructure noise, realized variance, and optimal sampling, *Review of Economic Studies* **75**, 339–369.

Bandi, F. M., Russell, J. R. and Zhu, Y.: 2008, Using high-frequency data in dynamic portfolio choice, *Econometric Reviews* **27**, 163–198.

Bandi, F. and Russell, J.: 2005, Realized covariation, realized beta, and microstructure noise. University of Chicago, Working Paper.

Bandi, F. and Russell, J.: 2008b, Microstructure noise, realized variance and optimal sampling, *Review of Economic Studies* **75**(2), 339–369.

Bannouh, K., Martens, M., Oomen, R. and Van Dijk, D.: 2012, Realized mixed-frequency factor models for vast-dimensional covariance estimation. Available at http://repub.eur.nl/res/pub/37470/.

Bannouh, K., Martens, M. and Van Dijk, D.: 2012, Measuring and forecasting volatility with the realized range in the presence of noise and non-trading. Working Paper.

Bannouh, K., Van Dijk, D. and Martens, M.: 2009, Range-based covariance estimation using high-frequency data: The realized co-range, *Journal of Financial Econometrics* **7**(4), 341–372.

Bardorff-Nielsen, O. and Shephard, N.: 2001, Non-gaussian ornstein-uhlenbeck based models and some of their applications in financial economics, *Journal of the Royal Statistical Society B* **63**, 167–241.

Barndorff-Nielsen, O. E. and Shephard, N.: 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society Series B* **64**, 253–280.

Barndorff-Nielsen, O. E. and Shephard, N.: 2004, Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics, *Econometrica* **72**, 885–925.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N.: 2009, Realised kernels in practice: trades and quotes, *Econometrics Journal* **12**, 1–32.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N.: 2011, Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading, *Journal of Econometrics* **162**, 149–169.

Beckers, S.: 1983, Variances of security price returns based on high, low, and closing prices, *Journal of Business* **56**, 97–112.

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

Bollerslev, T., Law, T. H. and Tauchen, G.: 2008, Risk, jumps, and diversification, *Journal of Econometrics* **144**, 234–256.

Bollerslev, T., Tauchen, G. and Zhou, H.: 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* **22**(11), 4463–4492.

Bollerslev, T. and Todorov, V.: 2011, Tails, fears and risk premia, *Journal of Finance* **66**(6), 2165–2211.

Brandt, M. and Diebold, F.: 2006, A no-arbitrage approach to range-based estimation of return covariances and correlations, *Journal of Business* **79**(1), 61–74.

Brownlees, C. and Gallo, G.: 2006, Financial econometric analysis at ultra-high frequency: data handling concerns, *Computational Statistics & Data Analysis* **51**(4), 2232–2245.

Brunetti, C. and Lildholdt, P. M.: 2007, Range-based (co)variance estimation: The co-range. Johns Hopkins University, Working Paper.

Busch, T., Christensen, B. and Nielsen, M.: 2010, The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets, *Journal of Econometrics* **160**, 48–57.

Chamberlain, G. and Rothschild, M.: 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* **51**(5), 1281–1304.

Chan, L., Karceski, J. and Lakonishok, J.: 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* **12**, 937–974.

Chen, N., Roll, R. and Ross, S.: 1986, Economic forces and the stock market, *Journal of Business* **59**, 383–403.

Christensen, K. and Podolskij, M.: 2006, Range-based estimation of quadratic variation. Working Paper 37/06, Technical University Dortmund.

Christensen, K. and Podolskij, M.: 2007, Realized range-based estimation of integrated variance, *Journal of Econometrics* **141**, 323–349.

Christensen, K., Podolskij, M. and Vetter, M.: 2009, Bias-correcting the realized range-based variance in the presence of market microstructure noise, *Finance and Stochastics* **13**, 239–268.

Comte, F. and Renault, E.: 1998, Long-memory in continuous-time stochastic volatility models, *Mathematical Finance* **8**, 291–323.

Connor, G., Goldberg, L. and Korajczyk, R.: 2012, *Portfolio Risk Management*, Princeton University Press.

Connor, G. and Korajczyk, R.: 1994, A test for the number of factors in an approximate factor model, *Journal of Finance* **48**(4), 1263–1291.

Davies, R. and Harte, D.: 1987, Tests for hurst effect, *Biometrika* **74**, 95–102.

De Pooter, M., Martens, M. and Van Dijk, D.: 2008, Predicting the daily covariance matrix for s&p 100 stocks using intraday data - but which frequency to use?, *Econometric Reviews* **27**, 199–229.

De Rossi, G., Zhang, H., Jessop, D. and Jones, C.: 2012, High frequency data for low frequency managers: hedging market exposure. White paper, UBS Investment Research.

DeMiguel, V., Garlappi, L. and Uppal, R.: 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* **22**, 1915–1953.

Diebold, F. and Strasser, G.: 2012, On the correlation structure of microstructure noise: a financial economic approach. NBER Working Paper 16469.

Dobrev, D.: 2007, Capturing volatility from large price moves: Generalized range theory and applications. Working Paper, Northwestern University.

Duyvesteyn, J., Martens, M. and Safavi, S.: 2011, Forecasting bond returns using jumps in intraday prices, *Journal of Fixed Income* **20**(4), 80–90.

Engle, R.: 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* **50**(4), 1987–1007.

Engle, R.: 2000, The econometrics of ultra high frequency data, *Econometrica* **68**, 1–22.

Epps, T. W.: 1979, Comovements in stock prices in the very short run, *Journal of the American Statistical Association* **74**, 291–298.

Fama, E. and French, K.: 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**, 3–56.

Fan, J., Fan, Y. and Lv, J.: 2008, High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* **147**, 186–197.

Fleming, J., Kirby, C. and Ostdiek, B.: 2001, The economic value of volatility timing, *Journal of Finance* **56**, 329–352.

Fleming, J., Kirby, C. and Ostdiek, B.: 2003, The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics* **67**(3), 473–509.

Foster, D. and Nelson, D.: 1996, Continuous record asymptotics for rolling sample variance estimators, *Econometrica* **64**(3), 139–174.

French, K., Schwert, G. and Stambaugh, R.: 1987, Expected stock return and volatility, *Journal of Financial Economics* **19**, 3–29.

Garman, M. and Klass, M.: 1980, On the estimation of security price volatilities from historical data, *Journal of Business* **53**, 67–78.

Ghysels, E. and Sinko, A.: 2011, Volatility forecasting and microstructure noise, *Journal of Econometrics* **160**(1), 257–271.

Griffin, J. E. and Oomen, R. C. A.: 2011, Covariance measurement in the presence of non-synchronous trading and market microstructure noise, *Journal of Econometrics* **160**, 58–68.

Griffin, J. and Oomen, R.: 2008, Sampling returns for realized variance calculations: tick time or transaction time?, *Econometric Reviews* **27**, 230–253.

Grinold, R. and Kahn, R.: 2000, Active portfolio management, a quantitative approach for producing superior returns and controlling risk, *2nd Edition, McGraw-Hill* .

Halbleib, R. and Voev, V.: 2011, Forecasting covariance matrices: A mixed frequency approach. University of Brussels, Working Paper.

Hansen, P. and Lunde, A.: 2006, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* **24**, 127–218.

Hansen, P., Lunde, A. and Voev, V.: 2010, Realized beta garch: A multivariate garch model with realized measures of volatility and covolatility. University of Aarhus, Working Paper.

Hautsch, N., Kyj, L. and Malec, P.: 2011, The merit of high-frequency data in portfolio allocation. Humboldt-University Berlin, Working Paper.

Hautsch, N., Kyj, L. and Oomen, R.: 2012, A blocking and regularization approach to high dimensional realized covariance estimation, *Journal of Applied Econometrics* **27**, 625–645.

Hayashi, T. and Yoshida, N.: 2005, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* **11**, 359–379.

Heston, S.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* **6**(2), 327–343.

Hsieh, D.: 1991, Chaos and nonlinear dynamics: applications to financial markets, *Journal of Finance* **46**, 1839–1877.

Ingersoll, J.: 1984, Some results in the theory of arbitrage pricing, *Journal of Finance* **39**(4), 1021–1039.

Jagannathan, R. and Ma, T.: 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* **58**(4), 1651–1684.

Kunitomo, N.: 1992, Improving the parkinson method of estimating security price volatilites, *Journal of Business* **65**, 295–302.

Kyj, L., Ostdiek, B. and Ensor, K.: 2009, Covariance estimation in dynamic portfolio optimization: A realized single factor model. Working Paper.

Liu, L., Patton, A. and Sheppard, K.: 2012, Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. Duke University, Working Paper.

Madhavan, A. and Yang, J.: 2003, Practical risk management, *Journal of Portfolio Management* **30**, 73–85.

Mandelbrot, B.: 1963, The variation of certain speculative prices, *Journal of Business* **36**(4), 394–419.

Marsh, T. and Rosenfeld, E.: 1986, Non-trading, market making, and estimates of stock price volatility, *Journal of Financial Economics* **15**, 395–472.

Martens, M.: 2006, Estimating ubiased and precise realized covariances. Erasmus University Rotterdam, Working Paper.

Martens, M. and Van Dijk, D.: 2007, Measuring volatility with the realized range, *Journal of Econometrics* **138**(1).

McAleer, M. and Medeiros, M.: 2008, Realized volatility: a review, *Econometric Reviews* **27**, 10–45.

Merton, R.: 1980, On estimating the expected return on the market: an exploratory investigation, *Journal of Financial Economics* **8**(4), 323–361.

Noureldin, D., Shephard, N. and Sheppard, K.: 2012, Multivariate high-frequency-based volatility (heavy) models, *Journal of Applied Econometrics* . Forthcoming.

Oomen, R. C. A.: 2005, Properties of bias-corrected realized variance under alternative sampling schemes, *Journal of Financial Econometrics* **3**, 555–577.

Oomen, R. C. A.: 2009. A universal scaling law of noise in financial markets.

Parkinson, M.: 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* **53**, 61–65.

Patton, A.: 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* **160**(1), 246–256.

Rogers, L. and Satchell, S.: 1991, Estimating variance from high, low and closing prices, *Annals of Applied Probability* **1**(4), 504–512.

Roll, R.: 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* **39**(4), 1127–1139.

Rosenberg, B.: 1974, Extra-market components of covariance in security returns, *Journal of Financial and Quantitative Analysis* **9**, 263–274.

Ross, S.: 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* **13**, 341–360.

Schwert, G.: 1990, Stock market volatility, *Financial Analysts Journal* **46**, 23–34.

Schwert, G. and Seguin, P.: 1990, Heteroscedasticity in stock returns, *Journal of Finance* **45**, 1129–1155.

Sheppard, K.: 2006, Realized covariance and scrambling. Working Paper.

Taylor, S. and Xu, X.: 1997, The incremental volatility information in one million foreign exchange rate quotations, *Journal of Empirical Finance* **4**, 355–380.

Todorov, V. and Bollerslev, T.: 2010, Jumps and betas: A new framework for disentangling and estimating systematic risks, *Journal of Econometrics* **157**(2), 220–235.

Voev, V. and Lunde, A.: 2007, Integrated covariance estimation using high-frequency data in the presence of noise, *Journal of Financial Econometrics* **5**(1), 68–104.

Wright, J. and Zhou, H.: 2009, Bond risk premia and realized jump risk, *Journal of Banking & Finance* **33**(12), 23332345.

Yang, Z. and Zhang, Q.: 2000, Drift-independent volatility estimation based on high, low, open, and close prices, *Journal of Business* **73**, 477–491.

Zhang, L.: 2011, Estimating covariation: Epps effect and microstructure noise, *Journal of Econometrics* **160**(1), 33–47.

Zhang, L., Mykland, P. and Aït-Sahalia, Y.: 2005, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100**, 1394–1411.

Zhou, B.: 1996, High-frequency data and volatility in foreign exchange rates, *Journal of Business and Economic Statistics* **14**(1), 45–52.

Zumbach, G.: 2006, The riskmetrics 2006 methodology. Available at SSRN: http://ssrn.com/abstract=1420185.

Biography

Karim Bannouh was born in De Bilt on 2 October, 1980. Karim studied financial econometrics at Erasmus University Rotterdam. In September 2006 he started as a PhD-candidate at the Econometric institute (EI) and Erasmus Research Institute of Management (ERIM) at Erasmus University Rotterdam. During his PhD-track his work was presented at several international conferences in Aarhus, Bergamo, Geneva, London, Oxford and Stanford. The article version of Chapter 3 is published in the *Journal of Financial Econometrics*. After his PhD-track, he worked at the firm-wide ING Group Model Validation department with a focus on financial market risk and trading risk models across asset classes. In May 2011 he joined Saemor Capital, a market neutral hedge fund that uses quantitative strategies, as a portfolio manager with risk management, asset allocation and market timing as his specific fields of interest.

ERIM Ph.D. Series Research in Management

Erasmus Research Institute of Management (ERIM)

ERIM Electronic Series Portal: `http://hdl.handle.net/1765/1`

Acciaro, M., *Bundling Strategies in Global Supply Chains*,
Promoter(s): Prof.dr. H.E. Haralambides,
EPS-2010-197-LIS, http://hdl.handle.net/1765/19742

Agatz, N.A.H., *Demand Management in E-Fulfillment*,
Promoter(s): Prof.dr.ir. J.A.E.E. van Nunen, EPS-2009-163-LIS,
http://hdl.handle.net/1765/15425

Alexiev, A., *Exploratory Innovation: The Role of Organizational and Top Management Team Social Capital*,
Promoter(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda,
EPS-2010-208-STR, http://hdl.handle.net/1765/20632

Asperen, E. van, *Essays on Port, Container, and Bulk Chemical Logistics Optimization.*
Promoter(s): Prof.dr.ir. R. Dekker,
EPS-2009-181-LIS, http://hdl.handle.net/1765/17626

Assem, M.J. van den, *Deal or No Deal? Decision Making under Risk in a Large-Stake TV Game Show and Related Experiments*,
Promoter(s): Prof.dr. J. Spronk,
EPS-2008-138-F&A, http://hdl.handle.net/1765/13566

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*,
Promoter(s): Prof.dr.ir. B.G.C. Dellaert,

EPS-2011-241-MKT, http://hdl.handle.net/1765/23670

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*,
Promoter(s): Prof.dr. B. Krug,
EPS-2012-262-ORG, http://hdl.handle.net/1765/32345

Bezemer, P.J., *Diffusion of Corporate Governance Beliefs: Board Independence and the Emergence of a Shareholder Value Orientation in the Netherlands*. Promoter(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2009-192-STR, http://hdl.handle.net/1765/18458

Binken, J.L.G., *System Markets: Indirect Network Effects in Action, or Inaction*,
Promoter(s): Prof.dr. S. Stremersch,
EPS-2010-213-MKT, http://hdl.handle.net/1765/21186

Blitz, D.C., *Benchmarking Benchmarks*,
Promoter(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor,
EPS-2011-225-F&A, http://hdl.handle.net/1765/22624

Boer-Sorbn, K., *Agent-Based Simulation of Financial Markets: A modular, Continuous-Time Approach*,
Promoter(s): Prof.dr. A. de Bruin,
EPS-2008-119-LIS, http://hdl.handle.net/1765/10870

Boon, C.T., *HRM and Fit: Survival of the Fittest!?*,
Promoter(s): Prof.dr. J. Paauwe & Prof.dr. D.N. den Hartog,
EPS-2008-129-ORG, http://hdl.handle.net/1765/12606

Borst, W.A.M., *Understanding Crowdsourcing: Effects of Motivation and Rewards on Participation and Performance in Voluntary Online Activities*,
Promoter(s): Prof.dr.ir. J.C.M. van den Ende & Prof.dr.ir. H.W.G.M. van Heck,
EPS-2010-221-LIS, http://hdl.handle.net/1765/21914

Braun, E., *City Marketing: Towards an Integrated Approach*,
Promoter(s): Prof.dr. L. van den Berg,
EPS-2008-142-MKT, http://hdl.handle.net/1765/13694

Brumme, W.-H., *Manufacturing Capability Switching in the High-Tech Electronics Technology Life Cycle*,
Promoter(s): Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. L.N. Van Wassenhove,
EPS-2008-126-LIS, http://hdl.handle.net/1765/12103

Budiono, D.P., *The Analysis of Mutual Fund Performance: Evidence from U.S. Equity Mutual Funds*,
Promoter(s): Prof.dr. M.J.C.M. Verbeek,
EPS-2010-185-F&A, http://hdl.handle.net/1765/18126

Burger, M.J., *Structure and Cooptition in Urban Networks*,
Promoter(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur,
EPS-2011-243-ORG, http://hdl.handle.net/1765/26178

Burgers, J.H., *Managing Corporate Venturing: Multilevel Studies on Project Autonomy, Integration, Knowledge Relatedness, and Phases in the New Business Development Process*,
Promoter(s): Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda,
EPS-2008-136-STR, http://hdl.handle.net/1765/13484

Camacho, N.M., *Health and Marketing; Essays on Physician and Patient Decision-making*,
Promoter(s): Prof.dr. S. Stremersch,
EPS-2011-237-MKT, http://hdl.handle.net/1765/23604

Carvalho de Mesquita Ferreira, L., *Attention Mosaics: Studies of Organizational Attention*,
Promoter(s): Prof.dr. P.M.A.R. Heugens & Prof.dr. J. van Oosterhout,
EPS-2010-205-ORG, http://hdl.handle.net/1765/19882

Chen, C.-M., *Evaluation and Design of Supply Chain Operations Using DEA*,
Promoter(s): Prof.dr. J.A.E.E. van Nunen,
EPS-2009-172-LIS, http://hdl.handle.net/1765/16181

Chen, H., *Individual Mobile Communication Services and Tariffs*,
Promoter(s): Prof.dr. L.F.J.M. Pau,
EPS-2008-123-LIS, http://hdl.handle.net/1765/11141

Defilippi Angeldonis, E.F., *Access Regulation for Naturally Monopolistic Port Terminals: Lessons from Regulated Network Industries*,
Promoter(s): Prof.dr. H.E. Haralambides,
EPS-2010-204-LIS, http://hdl.handle.net/1765/19881

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*,
Promoter(s): Prof.dr.ir. J.C.M. van den Ende,
EPS-2012-255-ORG, http://hdl.handle.net/1765/ 31174

Derwall, J.M.M., *The Economic Virtues of SRI and CSR*,
Promoter(s): Prof.dr. C.G. Koedijk,
EPS-2007-101-F&A, http://hdl.handle.net/1765/8986

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*,
Promoter(s): Prof.dr. D. De Cremer & Prof.dr. E. van Dijk,
EPS-2011-232-ORG, http://hdl.handle.net/1765/23268

Diepen, M. van, *Dynamics and Competition in Charitable Giving*,
Promoter(s): Prof.dr. Ph.H.B.F. Franses,
EPS-2009-159-MKT, http://hdl.handle.net/1765/14526

Dietvorst, R.C., *Neural Mechanisms Underlying Social Intelligence and Their Relationship with the Performance of Sales Managers*,
Promoter(s): Prof.dr. W.J.M.I. Verbeke,
EPS-2010-215-MKT, http://hdl.handle.net/1765/21188

Dietz, H.M.S., *Managing (Sales)People towards Perfomance: HR Strategy, Leadership & Teamwork*,
Promoter(s): Prof.dr. G.W.J. Hendrikse,
EPS-2009-168-ORG, http://hdl.handle.net/1765/16081

Doorn, S. van, *Managing Entrepreneurial Orientation*,
Promoter(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda,
EPS-2012-258-STR, http://hdl.handle.net/1765/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory*,
Promoter(s): Prof.dr. W.F.C. Verschoor,
EPS-2012-257-F&A, http://hdl.handle.net/1765/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*,
Promoter(s): Prof.dr. A. de Jong,
EPS-2011-240-F&A, http://hdl.handle.net/1765/26041

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*,
Promoter(s): Prof.dr.ir. R. Dekker,
EPS-2011-247-LIS, http://hdl.handle.net/1765/26509

Eijk, A.R. van der, *Behind Networks: Knowledge Transfer, Favor Exchange and Performance*,
Promoter(s): Prof.dr. S.L. van de Velde & Prof.dr.drs. W.A. Dolfsma,
EPS-2009-161-LIS, http://hdl.handle.net/1765/14613

Elstak, M.N., *Flipping the Identity Coin: The Comparative Effect of Perceived, Projected and Desired Organizational Identity on Organizational Identification and Desired Behavior*,
Promoter(s): Prof.dr. C.B.M. van Riel,
EPS-2008-117-ORG, http://hdl.handle.net/1765/10723

Erken, H.P.G., *Productivity, R&D and Entrepreneurship*,
Promoter(s): Prof.dr. A.R. Thurik,
EPS-2008-147-ORG, http://hdl.handle.net/1765/14004

Essen, M. van, *An Institution-Based View of Ownership*,
Promoter(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens,
EPS-2011-226-ORG, http://hdl.handle.net/1765/22643

Feng, L., *Motivation, Coordination and Cognition in Cooperatives*,
Promoter(s): Prof.dr. G.W.J. Hendrikse,
EPS-2010-220-ORG, http://hdl.handle.net/1765/21680

Gertsen, H.F.M., *Riding a Tiger without Being Eaten: How Companies and Analysts Tame Financial Restatements and Influence Corporate Reputation*,
Promoter(s): Prof.dr. C.B.M. van Riel,
EPS-2009-171-ORG, http://hdl.handle.net/1765/16098

Gijsbers, G.W., *Agricultural Innovation in Asia: Drivers, Paradigms and Performance*,
Promoter(s): Prof.dr. R.J.M. van Tulder,
EPS-2009-156-ORG, http://hdl.handle.net/1765/14524

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*,
Promoter(s): Prof.dr.ir. B.G.C. Dellaert,
EPS-2012-256-MKT, http://hdl.handle.net/1765/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*,
Promoter(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek,
EPS-2012-264-F&A, http://hdl.handle.net/1765/37170

Gong, Y., *Stochastic Modelling and Analysis of Warehouse Operations*,
Promoter(s): Prof.dr. M.B.M. de Koster & Prof.dr. S.L. van de Velde,
EPS-2009-180-LIS, http://hdl.handle.net/1765/16724

Greeven, M.J., *Innovation in an Uncertain Institutional Environment: Private Software Entrepreneurs in Hangzhou, China*,
Promoter(s): Prof.dr. B. Krug,
EPS-2009-164-ORG, http://hdl.handle.net/1765/15426

Guenster, N.K., *Investment Strategies Based on Social Responsibility and Bubbles*, Promoter(s): Prof.dr. C.G. Koedijk, EPS-2008-175-F&A, http://hdl.handle.net/1765/16209

Hakimi, N.A, *Leader Empowering Behaviour: The Leader's Perspective: Understanding the Motivation behind Leader Empowering Behaviour*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2010-184-ORG, http://hdl.handle.net/1765/17701

Halderen, M.D. van, *Organizational Identity Expressiveness and Perception Management: Principles for Expressing the Organizational Identity in Order to Manage the Perceptions and Behavioral Reactions of External Stakeholders*,
Promoter(s): Prof.dr. S.B.M. van Riel,
EPS-2008-122-ORG, http://hdl.handle.net/1765/10872

Hensmans, M., *A Republican Settlement Theory of the Firm: Applied to Retail Banks in England and the Netherlands (1830-2007)*,
Promoter(s): Prof.dr. A. Jolink & Prof.dr. S.J. Magala,
EPS-2010-193-ORG, http://hdl.handle.net/1765/19494

Hernandez Mireles, C., *Marketing Modeling for New Products*,
Promoter(s): Prof.dr. P.H. Franses,

EPS-2010-202-MKT, http://hdl.handle.net/1765/19878

Hessels, S.J.A., *International Entrepreneurship: Value Creation Across National Borders*,
Promoter(s): Prof.dr. A.R. Thurik,
EPS-2008-144-ORG, http://hdl.handle.net/1765/13942

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*,
Promoter(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda,
EPS-2012-259-STR, http://hdl.handle.net/1765/32167

Hoedemaekers, C.M.W., *Performance, Pinned down: A Lacanian Analysis of Subjectivity at Work*,
Promoter(s): Prof.dr. S. Magala & Prof.dr. D.H. den Hartog,
EPS-2008-121-ORG, http://hdl.handle.net/1765/10871

Hoever, I.J., *Diversity and Creativity: In Search of Synergy*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2012-267-ORG, http://hdl.handle.net/1765/1
Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*,
Promoter(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik,
EPS-2011-246-STR, http://hdl.handle.net/1765/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*,
Promoter(s): Prof.dr. D. De Cremer & Dr. M. van Dijke,
EPS-2011-244-ORG, http://hdl.handle.net/1765/26228

Huang, X., *An Analysis of Occupational Pension Provision: From Evaluation to Redesign*,
Promoter(s): Prof.dr. M.J.C.M. Verbeek & Prof.dr. R.J. Mahieu,
EPS-2010-196-F&A, http://hdl.handle.net/1765/19674

Hytönen, K.A. *Context Effects in Valuation, Judgment and Choice*,
Promoter(s): Prof.dr.ir. A. Smidts,
EPS-2011-252-MKT, http://hdl.handle.net/1765/30668

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*,
Promoter(s): Prof.dr. L.G. Kroon,
EPS-2011-222-LIS, http://hdl.handle.net/1765/22156

Jaspers, F.P.H., *Organizing Systemic Innovation*,
Promoter(s): Prof.dr.ir. J.C.M. van den Ende,
EPS-2009-160-ORG, http://hdl.handle.net/1765/14974

Jennen, M.G.J., *Empirical Essays on Office Market Dynamics*,
Promoter(s): Prof.dr. C.G. Koedijk & Prof.dr. D. Brounen,
EPS-2008-140-F&A, http://hdl.handle.net/1765/13692

Jiang, T., *Capital Structure Determinants and Governance Structure Variety in Franchising*,
Promoter(s): Prof.dr. G. Hendrikse & Prof.dr. A. de Jong,
EPS-2009-158-F&A, http://hdl.handle.net/1765/14975

Jiao, T., *Essays in Financial Accounting*,
Promoter(s): Prof.dr. G.M.H. Mertens,
EPS-2009-176-F&A, http://hdl.handle.net/1765/16097

Kaa, G. van, *Standard Battles for Complex Systems: Empirical Research on the Home Network*,
Promoter(s): Prof.dr.ir. J. van den Ende & Prof.dr.ir. H.W.G.M. van Heck,
EPS-2009-166-ORG, http://hdl.handle.net/1765/16011

Kagie, M., *Advances in Online Shopping Interfaces: Product Catalog Maps and Recommender Systems*,
Promoter(s): Prof.dr. P.J.F. Groenen,
EPS-2010-195-MKT, http://hdl.handle.net/1765/19532

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*,
Promoter(s): Prof.dr. S. Stremersch,
EPS-2011-239-MKT, http://hdl.handle.net/1765/23610

Karreman, B., *Financial Services and Emerging Markets*,
Promoter(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings,
EPS-2011-223-ORG, http://hdl.handle.net/1765/22280

Klein, M.H., *Poverty Alleviation through Sustainable Strategic Business Models: Essays on Poverty Alleviation as a Business Strategy*,
Promoter(s): Prof.dr. H.R. Commandeur,
EPS-2008-135-STR, http://hdl.handle.net/1765/13482

Krauth, E.I., *Real-Time Planning Support: A Task-Technology Fit Perspective*,
Promoter(s): Prof.dr. S.L. van de Velde & Prof.dr. J. van Hillegersberg,
EPS-2008-155-LIS, http://hdl.handle.net/1765/14521

Kwee, Z., *Investigating Three Key Principles of Sustained Strategic Renewal: A Longitudinal Study of Long-Lived Firms*,
Promoter(s): Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda,
EPS-2009-174-STR, http://hdl.handle.net/1765/16207

Lam, K.Y., *Reliability and Rankings*,
Promoter(s): Prof.dr. P.H.B.F. Franses,
EPS-2011-230-MKT, http://hdl.handle.net/1765/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*,
Promoter(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens,
EPS-2012-253-ORG, http://hdl.handle.net/1765/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*,
Promoter(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer,
EPS-2011-236-MKT, http://hdl.handle.net/1765/23504

Larco Martinelli, J.A., *Incorporating Worker-Specific Factors in Operations Management Models*,
Promoter(s): Prof.dr.ir. J. Dul & Prof.dr. M.B.M. de Koster,
EPS-2010-217-LIS, http://hdl.handle.net/1765/21527

Li, T., *Informedness and Customer-Centric Revenue Management*,
Promoter(s): Prof.dr. P.H.M. Vervest & Prof.dr.ir. H.W.G.M. van Heck,
EPS-2009-146-LIS, http://hdl.handle.net/1765/14525

Liere, D.W. van, *Network Horizon and the Dynamics of Network Positions: A Multi-Method Multi-Level Longitudinal Study of Interfirm Networks*,
Promoter(s): Prof.dr. P.H.M. Vervest,
EPS-2007-105-LIS, http://hdl.handle.net/1765/10181

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*,
Promoter(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak,
EPS-2011-229-F&A, http://hdl.handle.net/1765/22814

Maas, A.A., van der, *Strategy Implementation in a Small Island Context: An Integrative Framework*,
Promoter(s): Prof.dr. H.G. van Dissel,
EPS-2008-127-LIS, http://hdl.handle.net/1765/12278

Maas, K.E.G., *Corporate Social Performance: From Output Measurement to Impact Measurement*,
Promoter(s): Prof.dr. H.R. Commandeur,
EPS-2009-182-STR, http://hdl.handle.net/1765/17627

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*,
Promoter(s): Prof.dr. D.J.C. van Dijk,
EPS-2011-227-F&A, http://hdl.handle.net/1765/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*,
Promoter(s): Prof.dr. Ph.H.B.F. Franses,
EPS-2012-266-MKT, http://hdl.handle.net/1765/34930

Meuer, J., *Configurations of Inter-Firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*,
Promoter(s): Prof.dr. B. Krug,
EPS-2011-228-ORG, http://hdl.handle.net/1765/22745

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*,
Promoter(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda,

EPS-2012-260-S&E, http://hdl.handle.net/1765/32343

Moitra, D., *Globalization of R&D: Leveraging Offshoring for Innovative Capability and Organizational Flexibility*,
Promoter(s): Prof.dr. K. Kumar,
EPS-2008-150-LIS, http://hdl.handle.net/1765/14081

Moonen, J.M., *Multi-Agent Systems for Transportation Planning and Coordination*,
Promoter(s): Prof.dr. J. van Hillegersberg & Prof.dr. S.L. van de Velde,
EPS-2009-177-LIS, http://hdl.handle.net/1765/16208

Nalbantov G.I., *Essays on Some Recent Penalization Methods with Applications in Finance and Marketing*,
Promoter(s): Prof. dr P.J.F. Groenen,
EPS-2008-132-F&A, http://hdl.handle.net/1765/13319

Nederveen Pieterse, A., *Goal Orientation in Teams: The Role of Diversity*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2009-162-ORG, http://hdl.handle.net/1765/15240

Nguyen, T.T., *Capital Structure, Strategic Competition, and Governance*,
Promoter(s): Prof.dr. A. de Jong,
EPS-2008-148-F&A, http://hdl.handle.net/1765/14005

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in Short-term Planning and in Disruption Management*,
Promoter(s): Prof.dr. L.G. Kroon,
EPS-2011-224-LIS, http://hdl.handle.net/1765/22444

Niesten, E.M.M.I., *Regulation, Governance and Adaptation: Governance Transformations in the Dutch and French Liberalizing Electricity Industries*,
Promoter(s): Prof.dr. A. Jolink & Prof.dr. J.P.M. Groenewegen,
EPS-2009-170-ORG, http://hdl.handle.net/1765/16096

Nieuwenboer, N.A. den, *Seeing the Shadow of the Self*,
Promoter(s): Prof.dr. S.P. Kaptein,
EPS-2008-151-ORG, http://hdl.handle.net/1765/14223

Nijdam, M.H., Leader Firms: *The Value of Companies for the Competitiveness of the Rotterdam Seaport Cluster*,
Promoter(s): Prof.dr. R.J.M. van Tulder,
EPS-2010-216-ORG, http://hdl.handle.net/1765/21405

Noordegraaf-Eelens, L.H.J., *Contested Communication: A Critical Analysis of Central Bank Speech*,
Promoter(s): Prof.dr. Ph.H.B.F. Franses,
EPS-2010-209-MKT, http://hdl.handle.net/1765/21061

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*,
Promoter(s): Prof.dr. G. van der Pijl & Prof.dr. H. Commandeur & Prof.dr. M. Keil,
EPS-2012-263-S&E, http://hdl.handle.net/1765/34928

Nuijten, I., *Servant Leadership: Paradox or Diamond in the Rough? A Multidimensional Measure and Empirical Evidence*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2009-183-ORG, http://hdl.handle.net/1765/21405

Oosterhout, M., van, *Business Agility and Information Technology in Service Organizations*,
Promoter(s): Prof.dr.ir. H.W.G.M. van Heck,
EPS-2010-198-LIS, http://hdl.handle.net/1765/19805

Oostrum, J.M., van, *Applying Mathematical Models to Surgical Patient Planning*,
Promoter(s): Prof.dr. A.P.M. Wagelmans,
EPS-2009-179-LIS, http://hdl.handle.net/1765/16728

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on Bureaucracy and Formal Rules*,
Promoter(s): Prof.dr. P.P.M.A.R. Heugens,
EPS-2011-231-ORG, http://hdl.handle.net/1765/23250

Otgaar, A.H.J., *Industrial Tourism: Where the Public Meets the Private*,
Promoter(s): Prof.dr. L. van den Berg,
EPS-2010-219-ORG, http://hdl.handle.net/1765/21585

Ozdemir, M.N., *Project-level Governance, Monetary Incentives and Performance in Strategic R&D Alliances*,
Promoter(s): Prof.dr.ir. J.C.M. van den Ende,
EPS-2011-235-LIS, http://hdl.handle.net/1765/23550

Peers, Y., *Econometric Advances in Diffusion Models*,
Promoter(s): Prof.dr. Ph.H.B.F. Franses,
EPS-2011-251-MKT, http://hdl.handle.net/1765/30586

Pince, C., *Advances in Inventory Management: Dynamic Models*,
Promoter(s): Prof.dr.ir. R. Dekker,
EPS-2010-199-LIS, http://hdl.handle.net/1765/19867

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*,
Promoter(s): Prof.dr. M.J.C.M. Verbeek,
EPS-2012-254-F&A, http://hdl.handle.net/1765/30848

Potthoff, D., *Railway Crew Rescheduling: Novel Approaches and Extensions*,
Promoter(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. L.G. Kroon,
EPS-2010-210-LIS, http://hdl.handle.net/1765/21084

Poruthiyil, P.V., *Steering Through: How Organizations Negotiate Permanent Uncertainty and Unresolvable Choices*,
Promoter(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S. Magala,
EPS-2011-245-ORG, http://hdl.handle.net/1765/26392

Pourakbar, M. *End-of-Life Inventory Decisions of Service Parts*,
Promoter(s): Prof.dr.ir. R. Dekker,
EPS-2011-249-LIS, http://hdl.handle.net/1765/30584

Prins, R., *Modeling Consumer Adoption and Usage of Value-Added Mobile Services*,
Promoter(s): Prof.dr. Ph.H.B.F. Franses & Prof.dr. P.C. Verhoef,
EPS-2008-128-MKT, http://hdl.handle.net/1765/12461

Quak, H.J., *Sustainability of Urban Freight Transport: Retail Distribution and Local Regulation in Cities*,
Promoter(s): Prof.dr. M.B.M. de Koster,
EPS-2008-124-LIS, http://hdl.handle.net/1765/11990

Quariguasi Frota Neto, J., *Eco-efficient Supply Chains for Electrical and Electronic Products*,
Promoter(s): Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. H.W.G.M. van Heck,
EPS-2008-152-LIS, http://hdl.handle.net/1765/14785

Radkevitch, U.L, *Online Reverse Auction for Procurement of Services*,
Promoter(s): Prof.dr.ir. H.W.G.M. van Heck,
EPS-2008-137-LIS, http://hdl.handle.net/1765/13497

Rijsenbilt, J.A., *CEO Narcissism; Measurement and Impact*,
Promoter(s): Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur,
EPS-2011-238-STR, http://hdl.handle.net/1765/23554

Roelofsen, E.M., *The Role of Analyst Conference Calls in Capital Markets*,
Promoter(s): Prof.dr. G.M.H. Mertens & Prof.dr. L.G. van der Tas RA,
EPS-2010-190-F&A, http://hdl.handle.net/1765/18013

Rook, L., *Imitation in Creative Task Performance*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2008-125-ORG, http://hdl.handle.net/1765/11555

Rosmalen, J. van, *Segmentation and Dimension Reduction: Exploratory and Model-Based Approaches*,
Promoter(s): Prof.dr. P.J.F. Groenen,
EPS-2009-165-MKT, http://hdl.handle.net/1765/15536

Roza, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of Innovation, Absorptive Capacity and Firm Size*,
Promoter(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch,
EPS-2011-214-STR, http://hdl.handle.net/1765/22155

Rus, D., *The Dark Side of Leadership: Exploring the Psychology of Leader Self-serving Behavior*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2009-178-ORG, http://hdl.handle.net/1765/16726

Samii, R., *Leveraging Logistics Partnerships: Lessons from Humanitarian Organizations*,
Promoter(s): Prof.dr.ir. J.A.E.E. van Nunen & Prof.dr.ir. L.N. Van Wassenhove,
EPS-2008-153-LIS, http://hdl.handle.net/1765/14519

Schaik, D. van, *M&A in Japan: An Analysis of Merger Waves and Hostile Takeovers*,
Promoter(s): Prof.dr. J. Spronk & Prof.dr. J.P.M. Groenewegen,
EPS-2008-141-F&A, http://hdl.handle.net/1765/13693

Schauten, M.B.J., *Valuation, Capital Structure Decisions and the Cost of Capital*,
Promoter(s): Prof.dr. J. Spronk & Prof.dr. D. van Dijk,
EPS-2008-134-F&A, http://hdl.handle.net/1765/13480

Schellekens, G.A.C., *Language Abstraction in Word of Mouth*,
Promoter(s): Prof.dr.ir. A. Smidts,
EPS-2010-218-MKT, ISBN: 978-90-5892-252-6, http://hdl.handle.net/1765/21580

Sotgiu, F., *Not All Promotions are Made Equal: From the Effects of a Price War to Cross-chain Cannibalization*,
Promoter(s): Prof.dr. M.G. Dekimpe & Prof.dr.ir. B. Wierenga,
EPS-2010-203-MKT, http://hdl.handle.net/1765/19714

Srour, F.J., Dissecting Drayage: *An Examination of Structure, Information, and Control in Drayage Operations*,
Promoter(s): Prof.dr. S.L. van de Velde,
EPS-2010-186-LIS, http://hdl.handle.net/1765/18231

Stam, D.A., *Managing Dreams and Ambitions: A Psychological Analysis of Vision Communication*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2008-149-ORG, http://hdl.handle.net/1765/14080

Stienstra, M., *Strategic Renewal in Regulatory Environments: How Inter- and Intra-organisational Institutional Forces Influence European Energy Incumbent Firms*,
Promoter(s): Prof.dr.ir. F.A.J. Van den Bosch & Prof.dr. H.W. Volberda,
EPS-2008-145-STR, http://hdl.handle.net/1765/13943

Sweldens, S.T.L.R., *Evaluative Conditioning 2.0: Direct versus Associative Transfer of Affect to Brands*,
Promoter(s): Prof.dr. S.M.J. van Osselaer,
EPS-2009-167-MKT, http://hdl.handle.net/1765/16012

Szkudlarek, B.A., *Spinning the Web of Reentry: [Re]connecting reentry training theory and practice*,
Promoter(s): Prof.dr. S.J. Magala,

EPS-2008-143-ORG, http://hdl.handle.net/1765/13695

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*,
Promoter(s): Prof.dr. P.G.J. Roosenboom,
EPS-2012-265-F&A, http://hdl.handle.net/1765/37265

Tempelaar, M.P., *Organizing for Ambidexterity: Studies on the Pursuit of Exploration and Exploitation through Differentiation, Integration, Contextual and Individual Attributes*,
Promoter(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda,
EPS-2010-191-STR, http://hdl.handle.net/1765/18457

Tiwari, V., *Transition Process and Performance in IT Outsourcing: Evidence from a Field Study and Laboratory Experiments*,
Promoter(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. P.H.M. Vervest,
EPS-2010-201-LIS, http://hdl.handle.net/1765/19868

Tröster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*,
Promoter(s): Prof.dr. D.L. van Knippenberg,
EPS-2011-233-ORG, http://hdl.handle.net/1765/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision Making*,
Promoter(s): Prof.dr.ir. B.G.C. Dellaert,
EPS-2012-268-MKT, http://hdl.handle.net/1765/1

Tuk, M.A., *Is Friendship Silent When Money Talks? How Consumers Respond to Word-of-Mouth Marketing*,
Promoter(s): Prof.dr.ir. A. Smidts & Prof.dr. D.H.J. Wigboldus,
EPS-2008-130-MKT, http://hdl.handle.net/1765/12702

Tzioti, S., *Let Me Give You a Piece of Advice: Empirical Papers about Advice Taking in Marketing*,
Promoter(s): Prof.dr. S.M.J. van Osselaer & Prof.dr.ir. B. Wierenga,
EPS-2010-211-MKT, hdl.handle.net/1765/21149

Vaccaro, I.G., *Management Innovation: Studies on the Role of Internal Change Agents*,
Promoter(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda,
EPS-2010-212-STR, hdl.handle.net/1765/21150

Verheijen, H.J.J., *Vendor-Buyer Coordination in Supply Chains*,
Promoter(s): Prof.dr.ir. J.A.E.E. van Nunen,
EPS-2010-194-LIS, http://hdl.handle.net/1765/19594

Verwijmeren, P., *Empirical Essays on Debt, Equity, and Convertible Securities*,
Promoter(s): Prof.dr. A. de Jong & Prof.dr. M.J.C.M. Verbeek,
EPS-2009-154-F&A, http://hdl.handle.net/1765/14312

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*,
Promoter(s): Prof.dr. Ph.H.B.F. Franses,

EPS-2011-250-MKT, http://hdl.handle.net/1765/30585

Waard, E.J. de, *Engaging Environmental Turbulence: Organizational Determinants for Repetitive Quick and Adequate Responses*,
Promoter(s): Prof.dr. H.W. Volberda & Prof.dr. J. Soeters,
EPS-2010-189-STR, http://hdl.handle.net/1765/18012

Wall, R.S., *Netscape: Cities and Global Corporate Networks*, Promoter(s): Prof.dr. G.A. van der Knaap, EPS-2009-169-ORG, http://hdl.handle.net/1765/16013
Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*,
Promoter(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak,
EPS-2011-248-LIS, http://hdl.handle.net/1765/26564

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*,
Promoter(s): Prof.dr. M.J.C.M. Verbeek,
EPS-2011-242-F&A, http://hdl.handle.net/1765/26066

Weerdt, N.P. van der, *Organizational Flexibility for Hypercompetitive Markets: Empirical Evidence of the Composition and Context Specificity of Dynamic Capabilities and Organization Design Parameters*,
Promoter(s): Prof.dr. H.W. Volberda,
EPS-2009-173-STR, http://hdl.handle.net/1765/16182

Wubben, M.J.J., *Social Functions of Emotions in Social Dilemmas*,
Promoter(s): Prof.dr. D. De Cremer & Prof.dr. E. van Dijk,
EPS-2009-187-ORG, http://hdl.handle.net/1765/18228

Xu, Y., *Empirical Essays on the Stock Returns, Risk Management, and Liquidity Creation of Banks*,
Promoter(s): Prof.dr. M.J.C.M. Verbeek,
EPS-2010-188-F&A, http://hdl.handle.net/1765/18125

Yang, J., *Towards the Restructuring and Co-ordination Mechanisms for the Architecture of Chinese Transport Logistics*,
Promoter(s): Prof.dr. H.E. Harlambides,
EPS-2009-157-LIS, http://hdl.handle.net/1765/14527

Yu, M., *Enhancing Warehouse Perfromance by Efficient Order Picking*,
Promoter(s): Prof.dr. M.B.M. de Koster,
EPS-2008-139-LIS, http://hdl.handle.net/1765/13691

Zhang, D., *Essays in Executive Compensation*,
Promoter(s): Prof.dr. I. Dittmann,
EPS-2012-261-F&A, http://hdl.handle.net/1765/32344

Zhang, X., *Scheduling with Time Lags*,
Promoter(s): Prof.dr. S.L. van de Velde,
EPS-2010-206-LIS, http://hdl.handle.net/1765/19928

Zhou, H., Knowledge, *Entrepreneurship and Performance: Evidence from Country-level and Firm-level Studies*,
Promoter(s): Prof.dr. A.R. Thurik & Prof.dr. L.M. Uhlaner,
EPS-2010-207-ORG, http://hdl.handle.net/1765/20634

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*,
Promoter(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen,
EPS-2011-234-ORG, http://hdl.handle.net/1765/23422

Zwart, G.J. de, *Empirical Studies on Financial Markets: Private Equity, Corporate Bonds and Emerging Markets*,
Promoter(s): Prof.dr. M.J.C.M. Verbeek & Prof.dr. D.J.C. van Dijk,
EPS-2008-131-F&A, http://hdl.handle.net/1765/12703

MEASURING AND FORECASTING FINANCIAL MARKET VOLATILITY USING HIGH-FREQUENCY DATA

My research on measuring, modeling and forecasting financial market volatility as a doctoral student is presented in this dissertation in the form of three chapters. Chapter 2 of this dissertation introduces a novel heuristic bias-correction that aims at improving realized range-based volatility estimates. The third chapter introduces an innovative approach for estimating covariances using high-low price ranges sampled at intraday frequencies. The fourth chapter introduces a new covariance matrix estimator that is based on the idea of combining the merits of factor models and high-frequency data.

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

Erasmus Research Institute of Management - ERiM

ERIM PhD Series
Research in Management

Erasmus Research Institute of Management - ERiM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam (EUR)
P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl

Design & layout: B&T Ontwerp en advies (www.b-en-t.nl) Print: Haveka (www.haveka.nl)