Three Essays on Estimation and Dynamic Modelling of Multivariate Market Risks using High Frequency Financial Data

Dissertation

zur Erlangung des Grades
Doktor der Wirtschaftswissenschaften (Dr. rer. pol.)
am FachbereichWirtschaftswissenschaften
der Universität Konstanz

vorgelegt von:
Valeri Voev
Wollmatingerstr. 110
78467 Konstanz

Tag der mündlichen Prüfung: 11. Februar 2008

1. Referent: Prof. Dr. Winfried Pohlmeier
2. Referent: Prof. Dr. Asger Lunde

Preamble

Coming to Konstanz in October 2001 was a turning point in my career, not only because the university proved to be one of the leading universities in Germany, but mainly because of the quantitative orientation and methods used to teach Economics. The Master's Programme "International Economic Relations" gave me a very solid background to pursue a doctoral degree during the four years which I spent at the Chair of Econometrics of Prof. Winfried Pohlmeier, who gave me the opportunity to indulge in my research interests and for which I am deeply indebted to him. I am particularly grateful for his professional advice and for achieving to create and maintain a working atmosphere at his Chair, in which I could fully develop my potential and which offered me great opportunities for research cooperation. I would not have achieved this level in my professional career, without his support and guidance.

I am also very thankful to Prof. Pohlmeier for taking the initiative and responsibility to coordinate the Research Training Network "Microstructure of Financial Markets in Europe" funded by the European Commission, which gave me the opportunity to make two 6-month research visits at the Aarhus School of Business, where I was warmly accepted by Prof. Asger Lunde, who provided me with more than what was necessary to make my stay there very pleasant and productive. The time spent in Aarhus had a great influence on my professional development and it opened new horizons for my research.

Undoubtedly, I owe the writing of this thesis to my supervisors and colleagues in Konstanz and Aarhus. I would not have been able to make it so far, though, without the unconditional support of my parents and my wife Hao Kejia.

I would like to thank my colleagues Roxana Chiriac and Ingmar Nolte with whom it is a pleasure and always an exciting experience to work with. To my colleagues Sandra Nolte, Derya Uysal, Laura Wichert, Anton Flossmann and Rémi Piatek at the Chair of Econometrics and Peter Nyberg and Anders Wilhelmsson at the Aarhus School of Business I am grateful for many inspirational conversations.

Contents

Introduction **7**

Zusammenfassung **11**

1 Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise **15**

1.1 Introduction 15
1.2 Theoretical Framework 17
1.3 Covariance Estimation with High Frequency Data: General Discussion 19
1.4 Extensions of the *CC* Estimator: Bias Correction and Subsampling . 24
1.5 Empirical Results and Monte Carlo Study 34
1.5.1 Empirical Results 34
1.5.2 Simulation Design 37
1.5.3 Simulation Results 38
1.6 Conclusion 42
Bibliography 44
Appendix: Proofs 46

2 Estimating High-Frequency Based (Co-) Variances: A Unified Approach **55**

2.1 Introduction 55
2.2 Theoretical Setup 56
2.3 Estimation Procedures 59
2.3.1 Variance Estimation 60
2.3.2 Covariance Estimation 62
2.4 Monte Carlo Study 66
2.4.1 Simulation Setup 67
2.4.2 Estimators 69

2.4.3 Simulation Results 71
2.5 Conclusion 77
Bibliography 78
Appendix 80

3 Dynamic Modelling of Large Dimensional Covariance Matrices 82
3.1 Introduction 82
3.2 Forecasting models 84
3.2.1 A sample covariance forecast 85
3.2.2 A shrinkage sample covariance forecast 85
3.2.3 A RiskMetrics™ forecast 87
3.2.4 A simple realized covariance forecast 88
3.2.5 A shrinkage realized covariance forecast 88
3.2.6 Dynamic realized covariance forecasts 92
3.3 Data 93
3.4 Results 94
3.5 Conclusion 100
Bibliography 103

Complete Bibliography 105

List of Tables

1.1 T-statistics for the significance of the cross covariance function $\gamma(l)$ of the noise process of several stock pairs from the DJIA index in 2004. 36
1.2 Bias, standard deviation and RMSE (in percent) of covariance estimators. Case of stochastic correlation and correlated noise. 40

2.1 Description of the Monte Carlo Simulation Scenarios. 69
2.2 Mean, median, maximum and minimum of the RMSE rankings and of the relative RMSE across simulation scenarios using the Bayesian Information Criterion. 74
2.3 Mean, median, maximum and minimum of the RMSE rankings and of the relative RMSE across simulation scenarios using the modified Bayesian Information Criterion. 75

3.1 Results from the Diebold-Mariano tests. 99
3.2 Root mean squared prediction errors and corresponding ranks of the forecasting models based on the Frobenius norm. 100

Introduction

This dissertation consists of three stand-alone research papers, all of which treat the topic of estimation and dynamic modelling of multivariate volatility by employing the information contained in high-frequency data, which became available in the last 10 – 15 years. The main focus of all three studies is the multivariate application, in which one is interested in estimating and modelling the covariance matrix of more than two financial assets. Main motivation is that in practice, an economic agent is rarely exposed to a single source of risk, and it is exactly the correlations between risks, which make risk management so important. If risks were not correlated, the concepts of hedging, portfolio diversification and risk management would not have come into existence. The availability of high-frequency data opened new frontiers in the field of risk management not only to financial econometricians and mathematicians, but also to practitioners, who are now able to measure and manage risk much more accurately than only several years ago. It is exactly this relevance and novelty of the field that makes it currently a very active area of research.

The three chapters of this thesis can broadly be separated into two categories – estimation (Chapter 1 and 2) and dynamic modelling (Chapter 3), and are intentionally arranged in a particular sequence in the thesis. The first paper is mainly concerned with how to obtain a precise estimate of the covariance between two assets in the presence of a host of market microstructure frictions. An extension to this problem, where both the estimation of variances and covariances is addressed in a theoretically unified framework, is presented in the second paper, which also develops new estimation techniques improving substantially the efficiency of existing univariate and multivariate estimators. In the third chapter I abstract from the issue of market microstructure, starting from the point where a series of covariance matrices is available, for which a suitable time-series model is to be developed with the aim of making risk forecasts. Thus, the exposition in the thesis evolves logically from the problem of estimating a single covariance, through the estimation of a possibly high-dimensional covariance matrix, to the issue of dynamic modelling and forecast-

ing of the multivariate risks. This general introduction aims to summarize the main findings resulting from the separate studies.

Chapter 1 is a reprint of the article "Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise", published jointly with Asger Lunde in the *Journal of Financial Econometrics*, Vol. 5 (1), Winter 2007, pp. 68 – 104. The paper analyzes the impact of market microstructure noise and non-synchronicity on realized covariance type estimators. The main focus is on the estimator recently proposed by Hayashi & Yoshida (2005), who develop a technique to resolve the problem of the empirically observed biases in the covariance estimates constructed on the basis of data with increasing frequency. Epps (1979) documented this phenomenon and attributed it to the non-synchronous trading times across assets observed at higher frequencies. While the estimator proposed by Hayashi & Yoshida (2005) is unbiased and consistent in the presence of non-synchronicity, we show that market microstructure frictions can affect its properties significantly, leading to biasedness and inconsistency. The main contribution of the paper is to propose a new estimation approach, which restores the initial desirable statistical properties of the Hayashi & Yoshida (2005) estimator under very general assumptions on the noise process, in particular relaxing the hypothesis of an i.i.d. noise, usually encountered in the literature. Furthermore, we demonstrate that this generality is essential, by showing empirically that at very high frequencies returns of financial assets tend to exhibit significant lagged cross-correlations. As traditional tests for (cross-) correlations based on synchronous observations of a bivariate process are not applicable in our setup, we develop a method for statistical inference based on the directly available unsynchronized observations. We derive analytically the variance of our covariance estimator, and show in a simulation experiment that it is able to outperform various alternatives proposed in the literature, also in the cases where we have misspecification.

The second chapter is a joint work with Ingmar Nolte, in which we propose a unified theoretical framework which allows us to develop novel approaches to jointly estimate variances and covariances of financial assets. The necessity to bring both the univariate and multivariate aspects together is motivated by the existing divergence in the literature, in which one strand is devoted to the analysis of the univariate problem, while another one focuses on the estimation of a single covariance as is the case for the model in Chapter 1. As a consequence, there is a lack of a unified treatment of both issues which is indispensable for a real-world practical application.

This discrepancy is further exacerbated by the often differing sets of assumptions being made in the two literature strands. The reason for this is the difference in the impact of the market microstructure frictions on the properties of univariate and multivariate estimators. A simple example in this respect is the existence of non-synchronicity, which is not an issue to be concerned with for the variance estimation, but is extremely important for the covariance measurement.

Given the importance of both volatility and correlation for risk management, this paper establishes a theoretical setup in which the whole correlation matrix is estimated under consideration of the particular differences mentioned above. We develop a class of innovative estimation techniques, which are particularly easy to implement in practice as they involve running simple OLS regressions. Within the proposed model class, we put forward a data-driven procedure to select the best-performing model among the alternatives in the class for the particular data at hand. In order to compare the performance of our estimators to other univariate and multivariate approaches we carry out an extensive Monte Carlo experiment in which we simulate a wide range of possible price and noise data generating processes. The results show that our approaches are clearly outperforming all existing methods across the range of scenarios. For the "average" trading scenario, which describes the data generating process of observed market data quite well, the efficiency gains resulting from our approach are in the range of 35% to 50% compared to the next-best alternative outside our class of models. Apart from allowing for a very efficient estimation of the covariance matrix of interest, the proposed approach delivers estimates of the variance and autocovariance function of the noise process, which shed light on the degree of market efficiency.

The last chapter in the thesis is a reprint of the article "Dynamic Modelling of Large Dimensional Covariance Matrices", published in the volume "Recent Developments in High Frequency Financial Econometrics" of the series "Studies in Empirical Economics", published by Springer. The paper proposes a modelling framework for the dynamics of high-dimensional covariance matrices. Main challenge to the traditional multivariate volatility models, such as multivariate Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and multivariate Stochastic Volatility (SV) models is the so called "curse of dimensionality", which refers to the exponential increase of the model parameters with respect to the dimension of the model. An important condition, which has to be guaranteed by any multivariate volatility model is that the resulting model-implied and forecasted matrix should be positive (semi-)

definite. The model, which I develop in Chapter 3 of this dissertation is designed in a way to automatically fulfill this positivity condition, and is particularly suitable for applications with a large number of assets. In order to apply the model, one first needs to construct a series of covariance matrices, which are then subsequently modelled within a time series framework.

The main feature of the model is that it uses a transformation of the series of covariance matrices which decomposes them into so-called Cholesky factors. The advantage of this decomposition is that the covariance matrix forecasts resulting from the time series model for the Cholesky factors are by construction positive definite, without the necessity of imposing restrictions on the model parameters. Thus, standard Autoregressive Moving Average (ARMA) models can be applied to capture the dynamics of the Cholesky factors, which in turn are re-transformed to produce the matrix forecast. In an empirical application the model performance is compared against alternative approaches, feasible in large dimensional systems, by means of Diebold-Mariano tests, which are used to determine whether a given forecasting model is statistically significantly better than a competing model. The test results confirm the superiority of the methodology proposed in the paper against alternatives such as the RiskMetrics model, often applied among practitioners.

Zusammenfassung

Diese Dissertation besteht aus drei eigenständigen Forschungspapieren, die sich alle mit der Schätzung und der dynamischen Modellierung von multivariaten Volatilitätsmatrizen anhand von hochfrequenten Finanzmarktzeitreihen beschäftigen. Motiviert wird die Arbeit vor allem durch die Tatsache, dass ökonomische Entscheidungsträger mehrere, verschiedene Risiken in ihren Entscheidungen berücksichtigen müssen, und es insbesondere die Korrelationen solcher Risiken sind, die ein ausgefeiltes Risikomanagement heutzutage, insbesondere in der Finanzwirtschaft, unabdingbar machen. Wenn Risiken untereinander nicht korreliert wären, hätten die Konzepte des Hedgings, der Portfoliodiversifikation und des Risikomanagements nie entwickelt werden können. Die Verfügbarkeit ultra-hochfrequenter Daten schuf schlagartig viele neue Möglichkeiten im Gebiet des Risikomanagements für Finanzökonometriker und -mathematiker ebenso wie für Anwender in der Praxis, die nun in der Lage sind, mit Hilfe dieser Daten Risiken viel genauer zu messen und zu kontrollieren, als dies noch vor wenigen Jahren möglich war. Es sind die Relevanz und Neuartigkeit dieses Gebiets, die es zu einem regen Forschungsbereich machen.
Die drei Kapitel dieser Arbeit können grob in zwei Kategorien eingeteilt werden – Schätzung (Kapitel 1 und 2) und dynamische Modellierung (Kapitel 3), und sind bewusst in dieser Dissertation in einer bestimmten Reihenfolge zusammengestellt. Die erste Arbeit beschreibt ein Vorgehen, mit dem eine präzise Schätzung der Kovarianz zweier Finanzwerte, unter Berücksichtigung einer Vielfalt von Marktmikrostruktur Eigenschaften, erhalten werden kann. Eine Erweiterung dieses Problems, in der sowohl die Schätzung der Varianzen als auch die der Kovarianzen in einem einheitlichen theoretischen Rahmen behandelt wird, findet sich in der zweiten Arbeit. Darüber hinaus, werden in dieser Arbeit neue Schätztechniken entwickelt, die eine deutlich höhere Effizienz im Vergleich zu existierenden uni- und multivariaten Schätzern aufweisen. Im dritten Kapitel abstrahiere ich von dem Einfluss der Marktmikrostruktur und gehe davon aus, dass der Ökonometriker bereits über eine Zeitreihe von Kovarianzmatrizen verfügt, für die ein angemessenes Zeitreihenmod-

ell entwickelt werden soll, mit dem Ziel Risikovorhersagen zu machen. Durch diese Anordnung der Kapitel wird der Leser beginnend mit der Schätzung zunächst einer einzigen Kovarianz gefolgt von möglicherweise höher-dimensionalen Kovarianzmatrizen, zu dem Problem der dynamischen Modellierung und Vorhersagbarkeit von multivariaten Risiken geführt. Diese allgemeine Einleitung versucht die Hauptresultate der einzelnen Studien zusammenzufassen.

Das erste Kapitel ist bereits unter dem Titel "Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise" gemeinsam mit Asger Lunde im *Journal of Financial Econometrics*, Vol. 5 (1), Winter 2007, pp. 68 - 104 veröffentlicht worden. Diese Arbeit analysiert den Einfluß von Marktmikrostruktur Noise- und Asynchronitätseffekten auf realisierte Kovarianzschätzer. Der Fokus liegt hierbei auf dem kürzlich von Hayashi & Yoshida (2005) vorgeschlagenen Schätzer, der in der Lage ist, das Problem der empirisch beobachteten Verzerrungen zu lösen, die entstehen, wenn Kovarianzschätzungen basierend auf hochfrequenter Daten berechnet werden. Epps (1979) dokumentiert dieses Phänomen und führt es auf die bei höherer Frequenz erkennbare Nichtsynchronität der Zeitpunkte zurück, zu denen verschiedene Aktien gehandelt werden. Obwohl der Schätzer von Hayashi & Yoshida (2005), der auf nichtsynchronen Beobachtungen basiert, unter idealen Bedingungen unverzerrt und konsistent ist, zeigen wir, dass Marktmikrostruktureffekte zu Verzerrung und Inkonsistenz dieses Schätzers führen. Die wesentliche Errungenschaft der Arbeit, ist die Entwicklung einer neuen Schätzmethode, welche in der Lage ist, die ursprünglichen, wünschenswerten Eigenschaften des Hayashi-Yoshida Schätzers unter sehr allgemeinen Annahmen an den Noise-prozess wiederherzustellen. Insbesondere wird die starke Annahme eines unabhängigen identisch verteilten Noise abgeschwächt. Des Weiteren zeigen wir, dass diese Verallgemeinerung in Realität notwendig ist, da bei sehr hohen Frequenzen, Renditen von Finanzwerten signifikante verzögerte Kreuzkorrelationen aufweisen. Da traditionelle Tests für Kreuzkorrelationen basierend auf synchronen Beobachtungen eines bivariaten Prozesses in unserem Modell nicht anwendbar sind, entwickeln wir eine Methode, die es uns ermöglicht statistische Inferenz anhand der vorhandenen nichtsynchronen Beobachtungen zu betreiben. Wir leiten die Varianz unseres Kovarianzschätzers unter bestimmten Annahmen analytisch her und zeigen in einer Simulation, dass dieser Schätzer bessere Ergebnisse liefert, als andere Methoden, die in der Literatur vorgeschlagen werden, selbst in dem Fall, in dem Fehlspezifikationen vorliegen.

Das darauf folgende Kapitel ist eine Arbeit, die in Zusammenarbeit mit Ingmar Nolte

entstanden ist. Hier entwickeln wir neue Ansätze um die Varianzen und Kovarianzen von Finanzwerten in einem einheitlichen theoretischen Rahmen zu schätzen. Das Interesse sowohl uni- als auch multivariate Aspekte gemeinsam zu betrachten wird hauptsächlich dadurch motiviert, dass in der bestehenden Literatur kaum Beiträge über die gemeinsame Schätzung von Varianzen und Kovarianzen existieren, welche aber von entscheidender Bedeutung für praktische Problemstellungen ist. Desweiteren, werden in Papieren über Volatilitätsschätzung im Vergleich zu Artikeln über Kovarianzschätzung, häufig unterschiedliche theoretischen Annahmen getroffen. Der Grund hierfür liegt in den verschiedenen Einflüssen, die Marktmikrostruktureffekte auf die Eigenschaften der uni- und multivariaten Schätzer haben. Ein einfaches Beispiel dafür ist das nichtsynchrone Handel, welches für die Varianzschätzung keine besondere Rolle spielt, für die Kovarianzschätzung jedoch extrem wichtig ist.
In Anbetracht der Wichtigkeit von Volatilitäten und Korrelationen für das Risikomanagement, schafft diese Arbeit einen theoretischen Rahmen, der es erlaubt die gesamte Korrelationsmatrix unter Berücksichtigung der oben genannten Unterschiede zu schätzen. Wir entwickeln eine Klasse von innovativen, einfach anwendbaren Schätztechniken, die auf herkömmlichen KQ Schätzungen basieren. Innerhalb der betrachteten Modellklasse führen wir eine datengetriebene Prozedur ein, um das am besten für die bestimmte Anwendung geeignete Modell der Klasse auszuwählen. Um die Leistung unserer Schätzer mit alternativen Schätzungsverfahren zu vergleichen führen wir eine umfangreiche Monte Carlo Simulation durch, die eine breite Vielfalt von möglichen Handelsszenarien sowie auch Preis- und Noiseprozesse abdeckt. Die Ergebnisse zeigen, dass unser Ansatz allen betrachteten Alternativen überlegen ist. Für das "durchschnittliche" Handelsszenario, welches den datengenerierenden Prozess von tatsächlich beobachteten Marktdaten gut beschreibt, liegen die Effizienzgewinne in der Größenordnung von 30% - 50%, verglichen mit der nächstbesten Alternative außerhalb unserer Modellklasse. Der vorgeschlagene Ansatz liefert neben einer sehr effizienten Schätzung der Kovarianzmatrix auch Schätzungen für die Varianz und Autokovarianzfunktion des Noiseprozesses, die ihrerseits aussagekräftig für die Effizienz des betrachteten Marktes sind.
Das letzte Kapitel der Arbeit ist ein Abdruck des Artikels "Dynamic Modelling of Large Dimensional Covariance Matrices", das im Band "Recent Developments in High Frequency Financial Econometrics" der Serie "Studies in Empirical Economics" erschienen ist. In dieser Arbeit wird ein Modell für die Dynamik höherdimensionaler Kovarianzmatrizen entwickelt. Die größte Schwierigkeit für die traditionellen multi-

variaten Volatilitätsmodelle wie MGARCH (Multivariate Generalized Autoregressive Conditional Heteroscedasticity) und multivariate Stochastic Volatility (SV) Modelle besteht in dem "Fluch der Dimensionalität", der sich auf den exponentiellen Zuwachs an Modellparametern in Relation zur Modelldimension bezieht. Eine wichtige Bedingung, die von allen multivariaten Volatilitätsmodellen erfüllt werden soll, ist, dass die vorhergesagten Matrizen positiv (semi-)definit sein müssen.

Das im dritten Kapitel dieser Dissertation entwickelte Modell ist so konstruiert, dass es diese Positivitätsbedingung automatisch erfüllt und sich hervorragend für Anwendungen mit einer großen Anzahl von Anlagen eignet. In dieser Arbeit ist davon ausgegangen, dass dem Ökonometriker eine Zeitreihe von Kovarianzmatrizen zu Verfügung steht, die anschließend dynamisch modelliert wird. Das Besondere an dem Modell ist, dass es eine Transformation der Folge der Kovarianzmatrizen nutzt, welche diese in sogenannte Cholesky Faktoren zerlegt. Der Vorteil dieser Zerlegung ist, dass die Kovarianzmatrixvorhersage, die aus dem Zeitreihenmodell für die Cholesky Faktoren entsteht, per Konstruktion positiv definit ist, ohne Restriktionen für die Modellparameter zu spezifizieren. Es können daher herkömmliche ARMA (Autoregressive Moving Average) Modelle benutzt werden um die Dynamik der Cholesky Faktoren zu modellieren. Die daraus resultierende Cholesky-Faktorvorhersagen werden ihrerseits wieder zu einer Kovarianzmatrixvorhersage zurücktransformiert. In einer empirischen Anwendung wird das Modell anhand von Diebold-Mariano Tests, mit Alternativen verglichen, die für höherdimensionale Anwendungen geeignet sind. Die Testergebnisse bestätigen die Überlegenheit des vorgeschlagenen Modells gegen Alternativen wie z.B. dem in der Praxis oft eingesetzten RiskMetrics Modell.

Chapter 1

Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise

1.1 Introduction

It is now widely accepted, that market microstructure noise causes observed market prices to deviate from some efficient price which has the martingale property. This led to the idea of viewing observed prices as noisy measures of the latent true price process. A classical example of how market frictions distort efficient prices is Roll's (1984) model. It shows that the presence of bid-ask spread leads to a negative first-order correlation in observed price changes.

In this paper, we are interested in estimating the integrated covariance of the latent price process. A recently proposed alternative to the realized covariance is the *cumulative covariance* (CC) estimator developed in Hayashi & Kusuoka (2004) and Hayashi & Yoshida (2004, 2005). We examine the properties of this estimator under a very flexible noise specification and find that it is biased and inconsistent. Consequently, in a first step we propose a bias correction in the spirit of the realized kernels of Barndorff-Nielsen, Hansen, Lunde & Shephard (2006). The resulting estimator hence accounts for the two sources of bias in high-frequency covariance measurement – non-synchronicity and noise. In order to choose the length of the kernel we develop an approach to estimate the cross-correlation function of the noise in *calendar time* using the raw tick data without interpolation. The choice of calendar time here is important as there is no straightforward way of defining tick-time dependence in the

multivariate case. In a second step, once the bias is corrected for, we show how to improve the efficiency of the estimator by means of subsampling with an optimal number of grids.

Analyzing stock data for 2004 we find that observed prices do not seem to conform to the martingale plus i.i.d. noise assumptions since there are significant lead-lag patterns across stocks in the DJIA index, which confirms the practical relevance of the general assumptions we make. We use covariance signature plots to verify that bias correction works in practice, while the efficiency gains achieved by subsampling are documented with the help of a simulation experiment.

We benefited from a number of papers that derive an expression for the bias and a bias correction of realized variance in the univariate case, such as Bandi & Russell (2005*a*), Oomen (2005), Zhang, Mykland & Aït-Sahalia (2005), Barndorff-Nielsen et al. (2006) and Hansen & Lunde (2006), among others. In the extension to the multivariate framework, the additional complication of non-synchronous trading in different assets arises. The non-synchronicity leads to the so called Epps effect due to Epps (1979), which manifests itself as a bias towards zero as the sampling frequency increases. An investigation of the determinants of the Epps effect can be found in Renò (2001). While the realized covariance suffers from the Epps effect and is therefore biased, the CC estimator is shown to be unbiased and consistent under the assumption that the observations are uncontaminated by noise. Important empirical studies on the properties of different realized covariance-based estimators are Martens (2004) and de Pooter, Martens & van Dijk (2006).

While non-synchronicity is of extreme importance in covariance estimation, we focus our discussion on noise for two reasons. The first reason is that in the absence of noise, the CC estimator seems to be optimal with non-synchronously observed diffusion processes. It uses all available data and is unbiased and consistent. The second reason is that Zhang (2006*b*) has studied extensively the last-tick interpolation based realized covariance under non-synchronicity and noise. Important issues for covariance estimation based on synchronized observations, such as how often to sample and what kind of synchronization bias corrections to employ, have been addressed by Bandi & Russell (2005*b*) and Zhang (2006*b*). In a similar setup, Sheppard (2005) has also analyzed the effect of scrambled prices on the bias of realized variance and covariance. It seems that a treatment of a non-synchronicity adjusted estimator (the CC estimator) when observed prices are noisy is still lacking in the literature. An exception is the concurrent and independent research of Griffin & Oomen (2006) who

examine the mean and variance of the CC estimator under i.i.d. (across time and assets) noise and exogenous Poisson arrival times with constant intensities. Hence, it is of interest to relate their results to ours whenever possible.
The remainder of the paper is structured as follows: in Section1.2 we present the theoretical assumptions on the price and noise processes. Section1.3 describes several of the recently introduced high-frequency covariance estimators and draws some conclusions regarding their performance in the presence of non-synchronicity and noise. The core of the paper, the theoretical development of the proposed extensions to the CC estimator, is contained in Section1.4. Section1.5 includes an empirical application of the estimators and a short simulation experiment, and Section1.6 concludes. The proofs are collected in the Appendix.

1.2 Theoretical Framework

We consider a K-dimensional vector of efficient prices, given by $\mathbf{p}^*(t) = \mathbf{a}^*(t)+\mathbf{m}^*(t)$, where $\mathbf{a}^*(t)$ is a drift term with continuous finite-variation paths and $\mathbf{m}^*(t)$ is a local martingale. The quadratic covariation matrix-valued process is defined as

$$\mathbf{C}(t) = \operatorname*{plim}_{M\to\infty} \sum_{j=0}^{M-1} \{\mathbf{p}^*(t_{j+1}) - \mathbf{p}^*(t_j)\}\{\mathbf{p}^*(t_{j+1}) - \mathbf{p}^*(t_j)\}',$$

for any sequence of partitions $t_0 = 0 < t_1 < \ldots < t_M = t$ with $\sup_j\{t_{j+1}-t_j\} \to 0$ as $M \to \infty$. Under the assumption that the drift process is continuous, the quadratic covariation of the log-price process equals the quadratic covariation of the martingale component. This result holds irrespective of the presence of jumps in the local martingale component (see Barndorff-Nielsen & Shephard (2004), henceforth BNS). In practice, we have for each time period denoted by h (usually a day), M intra-period observations. These could be irregularly spaced, as in the case of transactions data, or equidistant (e.g., an observation every 5 minutes). If the observations are regularly spaced, with $\delta = h/M$ being the time between observations, the j-th intra-period return for the i-th period is defined as:

$$\mathbf{r}^*_{j,i} = \mathbf{p}^*\left((i-1)h + j\delta\right) - \mathbf{p}^*\left((i-1)h + (j-1)\delta\right), \qquad j = 1\ldots M.$$

The realized covariance matrix for period i is given by:

$$\mathbf{RC}^{(M)} = \sum_{j=1}^{M} \mathbf{r}^*_{j,i} \mathbf{r}^{*\prime}_{j,i}. \tag{1.1}$$

BNS (2004) derive the asymptotic distribution of the realized covariance under the assumption that the true price process belongs to the class of continuous semi-martingales with stochastic volatility and show that the presence of drift does not affect the asymptotic results. In our study, we derive finite sample properties and a drift would only unnecessarily complicate the derivations.[1] Therefore, we make the following assumption:

Assumption 1.1. *The efficient price process has no drift, such that* $\mathbf{p}^*(t) = \mathbf{m}^*(t)$. $\mathbf{m}^*(t)$ *is a multivariate stochastic volatility process satisfying*

$$\mathbf{m}^*(t) = \int_0^t \mathbf{\Theta}(u) d\mathbf{W}(u)$$

where $\mathbf{\Theta}$ *is the spot covolatility process and* $\mathbf{W}$ *is a vector standard Brownian motion of dimension* q*. All the elements of* $\mathbf{\Theta}(t)\mathbf{\Theta}(t)'$ *satisfy the Lipschitz condition.*

In this setting we have the following important relationships. The spot covariance is defined as $\mathbf{\Sigma}(t) = \mathbf{\Theta}(t)\mathbf{\Theta}(t)'$ and its increment over a subinterval j is given by $\mathbf{\Sigma}_j = \int_{t_{j-1}}^{t_j} \mathbf{\Sigma}(u)du$. For stochastic volatility martingales, the quadratic covariation equals the integrated covariance given by $\mathbf{IC}(t) = \int_0^t \mathbf{\Sigma}(u)du$, and it follows that realized covariance consistently estimates increments of integrated covariance, $\mathbf{IC}_i = \mathbf{IC}(hi) - \mathbf{IC}(h(i-1))$. This result is of particular importance since BNS(2002) have indicated that in the univariate case this increment is the variance of the i-th period log-return conditional on the path of the volatility process. In the multivariate extension, BNS(2004) show that realized covariance is asymptotically normal with a $K \times K$ matrix of means $\mathbf{IC}_i$. The asymptotic covariance of $\sqrt{\delta^{-1}}(\mathbf{RC}^{(M)} - \mathbf{IC}_i)$ is a $K^2 \times K^2$ matrix $\mathbf{\Psi}_i$, whose generic element, corresponding to the covariance between the (A, B) and (A', B') element of $\mathbf{RC}^{(M)}$, is given by

$$\psi_{i,(AB,A'B')} = \int_{h(i-1)}^{hi} \left\{\sigma_{AA'}(u)\sigma_{BB'}(u) + \sigma_{AB'}(u)\sigma_{BA'}(u)\right\} du, \tag{1.2}$$

[1] For example, with only two intraday returns, a drift of 20% per year would lead to a bias in the order of 10^{-6}, which diminishes quickly as the number of observations increases.

where $\sigma_{AB}(t)$ is the (A, B) element of the $\mathbf{\Sigma}(t)$ process. This matrix is unknown but can be consistently estimated (see BNS (2004)).

The theory presented above suggests that if we could directly observe $\mathbf{p}^*$, we would use all available observations to compute $\mathbf{RC}^{(M)}$. Market microstructure effects, however, distort the price process. The bid-ask bounce, non-synchronous trading and price discreetness are perhaps the most important reasons the observed price process does not conform to the martingale assumption. For this reason, in empirical work one must differentiate between the "true" price process which is assumed to be a martingale and the observed process, which is a noisy signal of the former: $\mathbf{p}(t) = \mathbf{p}^*(t) + \mathbf{u}(t)$, where $\mathbf{u}(t)$ is a vector error term capturing all market microstructure effects.

In the following we focus on estimating the covariance over a single period (t ranging from 0 to 1, with h representing a trading day), so henceforth we drop the index i. In the presence of noise, the observed return is given by $\mathbf{r}_j = \mathbf{r}_j^* + \mathbf{e}_j$, where $\mathbf{e}_j = \mathbf{u}_j - \mathbf{u}_{j-1}$ and $\mathbf{u}_j = \mathbf{u}(j\delta)$. Initially, we make only a stationarity assumption about the noise process:

Assumption 1.2. *The noise process* $\mathbf{u}$ *is covariance stationary with autocovariance function given by* $\mathbf{\Gamma}(l) = \mathrm{E}[\mathbf{u}(t)\mathbf{u}'(t+l)]$.

In addition to allowing for serial correlation in the noise process, this characterization does not exclude dependence between the noise and the efficient price. Note that we have defined the dynamics of the noise process in calendar time. Usually dependence in tick time is more intuitive and easier to work with; in the multivariate case, however, there might be a considerable difference with respect to the assets' trading (quoting) activity. Therefore, defining dependence in tick time in this context is troublesome.

1.3 Covariance Estimation with High Frequency Data: General Discussion

We are now in a position to derive a general expression for the bias of the realized covariance $\mathbf{RC}^{(M)}$. Obviously, when noise is present, the estimator (1.1) is only feasible with observed returns in the place of efficient returns. In order to isolate the bias caused by noise from the bias due to the Epps effect, we initially assume that observations are *synchronous*. When characterizing estimators using non-synchronous observations, we will relax this assumption.

Theorem 1.1. *Given that price observations are synchronous and Assumptions 1.1 and 1.2 hold, the bias of the realized covariance is given by*

$$\mathrm{E}\left[\mathbf{RC}^{(M)} - \mathbf{IC}\right] = \boldsymbol{\Upsilon}_M + \boldsymbol{\Upsilon}_M' + M\left(2\boldsymbol{\Gamma}(0) - \boldsymbol{\Gamma}(\delta) - \boldsymbol{\Gamma}'(\delta)\right),$$

where $\boldsymbol{\Upsilon}_M \equiv \mathrm{E}\left[\sum_{j=1}^{M} \mathbf{r}_j^* \mathbf{e}_j'\right]$.

Proof. See the Appendix.

This theorem is a straightforward extension of the corresponding unvariate results in Hansen & Lunde (2006). It is interesting to note the difference from the univariate problem of estimating integrated variance. If $\boldsymbol{\Upsilon}_M = 0$, Hansen & Lunde (2006) show that the bias of realized variance is positive. In contrast, we show in the following example that this is not necessarily the case for realized covariances. We consider two assets A and B and introduce the notation $\gamma(l)$ for the (A, B) element of $\boldsymbol{\Gamma}(l)$.[2]

Example 1.1. *Assume that there are two assets A and B, the contemporaneous noise correlation is zero ($\gamma(0) = 0$), and let the first asset "lead" the second, such that $\gamma(\delta) > 0$, while $\gamma(-\delta) = 0$. The noise is independent of the price process, such that $\boldsymbol{\Upsilon}_M = 0$. The sign of the bias is given by*

$$\begin{aligned} &sign\left(2\boldsymbol{\Gamma}(0) - \boldsymbol{\Gamma}(\delta) - \boldsymbol{\Gamma}'(\delta)\right) = sign\left(\boldsymbol{\Gamma}(0) - \boldsymbol{\Gamma}(\delta) + \boldsymbol{\Gamma}(0) - \boldsymbol{\Gamma}(-\delta)\right) \\ &= \left[\begin{pmatrix} + & 0 \\ 0 & + \end{pmatrix} - \begin{pmatrix} \times & + \\ 0 & \times \end{pmatrix}\right] + \left[\begin{pmatrix} + & 0 \\ 0 & + \end{pmatrix} - \begin{pmatrix} \times & 0 \\ + & \times \end{pmatrix}\right] = \begin{pmatrix} + & - \\ 0 & + \end{pmatrix} + \begin{pmatrix} + & 0 \\ - & + \end{pmatrix} = \begin{pmatrix} + & - \\ - & + \end{pmatrix}, \end{aligned}$$

where the symbol $\times$ signifies that the element could be either positive or negative, but the resulting sum is unambiguously positive by the Cauchy-Schwartz inequality.

The example shows that the realized variance is biased upwards, while the covariance could be biased downwards, possibly exacerbating the negative bias due to non-synchronous trading (Epps effect). Even if $\boldsymbol{\Upsilon}_M \neq 0$, we might find it reasonable that if the noise is mainly due to the trading process of its own asset, as with the non-synchronous revision of quotes in Hansen & Lunde (2006), then the off-diagonal elements of $\boldsymbol{\Upsilon}_M$ will be close to zero and the above result will still hold off the diagonal. For a very thorough treatment of the last-tick realized covariance we refer

[2] Owing to the properties of the multivariate autocovariance function, we have that the (B, A) element of $\boldsymbol{\Gamma}(l)$ is equal to $\gamma(-l)$. In general, where it is not further specified, we use the subscript (A, B) to denote the (A, B) element of variance-covariance matrices, e.g. $IC_{(A,B)}$. The notation $\gamma(l)$ is used as a simplification, since it is extensively used in the proofs.

the reader to Zhang (2006*b*), who derives its mean squared error (MSE) and optimal sampling frequency with non-synchronicity and i.i.d. noise. Bandi & Russell (2005*b*) also study the issue of optimal sampling frequency, based on the idea that realized covariance based on high-frequency data essentially estimates the moments of the noise, rather than the integrated covariance. The estimated noise moments and integrated quarticity can then be used to compute the MSE of the estimator. The optimal frequency is chosen as the value of M, which minimizes the MSE. We will denote this estimator, applied with a second order lead and lag bias correction (as in Bandi & Russell (2005*b*)), by $B\&R$. As any estimator based on interpolated prices, this one also suffers from the Epps effect when trading is non-synchronous.

Generally, the realized covariance can be modified to include L leads and U lags to cancel the autocorrelations induced by noise and/or non-synchronicity, resulting in the formulation

$$\mathbf{RCLL}^{(M)} = \sum_{j=1}^{M} \sum_{l=-L}^{U} \mathbf{r}_{j+l} \mathbf{r}_j'. \tag{1.3}$$

In the bivariate case this estimator has been studied extensively by Griffin & Oomen (2006). They derive its bias and variance under non-synchronicity and examine its MSE under various choices of L and U. In our simulation experiment, for the sake of completeness, we include this estimator with $L = U = 1$, which we denote by $\mathbf{RC}_{AC_1}^{(M)}$.

One of the major drawbacks of realized covariance is that it is based on (last-tick) interpolation, which makes it susceptible to the Epps effect. An interesting approach to partially reduce the effect of interpolation is the so-called "Replace all" estimator, which is a realized covariance based on non-synchronous data. This estimator is used in Martens (2004) and is based on the synchronization technique of Harris, McInish, Shoesmith & Wood (1995) as follows: a first price tuple is obtained as soon as all assets have traded; then, the next one is recorded as soon as all of them have traded again (the most slowly trading asset determines when this happens), setting the prices of the "quicker" assets to their most recent values. In this case, the recorded price of the asset which traded last will indeed be the price at that point of time, while the prices for the other assets will be determined using last-tick interpolation from their latest transactions (or quotes). We call the time elapsed from this latest transaction to the recording of the observation the *interpolation span*. Returns, computed with these matched price tuples, are then used to compute the realized covariance. Hence, we denote the estimator by RC^{ra}. The bias of this estimator can

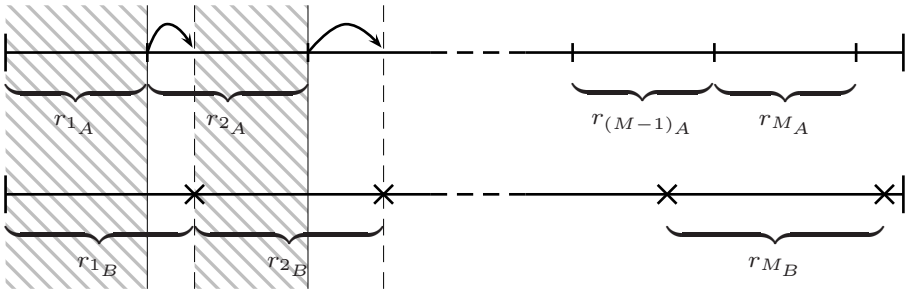

Figure 1.1: An example of the "Replace all" matching scheme and realized covariance based thereof. The expected value of RC^{ra} is equal to the sum of the increments of integrated covariance over the shaded regions. The arrows represent last-tick interpolation.

be obtained as a corollary of Theorem 1 in Zhang (2006*b*). To illustrate the nature of this bias, we have depicted a possible sequence of observations in Figure 1.1. Under this scenario, we have that $RC^{ra} = r_{1_A} r_{1_B} + r_{2_A} r_{2_B} + \ldots + r_{M_A} r_{M_B}$. The expected value of this sum equals the sum of the increments of integrated covariation over the shaded regions only, since cross products of non-overlapping returns are zero in expectation. Thus, the bias of RC^{ra} is equal to the integral of $\sigma_{AB}(t)$ over the non-shaded regions.[3]

An estimation procedure which solves the non-synchronicity problem is the cumulative covariance (CC) estimator proposed by Hayashi & Yoshida (2005), who show that it is unbiased and consistent in the absence of noise. Some additional notation is needed in order to present this estimator. First, denote by M_A the number of trades (quotes) for asset A and by M_B the number of trades (quotes) for asset B. Let

$$\Pi_A = \{t_{j_A} : j_A = 1, 2, \ldots, M_A\} \quad \text{and} \quad \Pi_B = \{t_{j_B} : j_B = 1, 2, \ldots, M_B\}$$

denote the sets of transaction (quoting) times of asset A and B, respectively. The following results are derived under the assumption that both assets trade at $t_0 = 0$ and $t_{M_A} = t_{M_B} = 1$, where $t = 1$ is the end of the trading day. The tick returns of A and B are given by $r_{j_A} = p^A_{j_A} - p^A_{j_A - 1} = r^*_{j_A} + e_{j_A}$ and $r_{j_B} = p^B_{j_B} - p^B_{j_B - 1} = r^*_{j_B} + e_{j_B}$, respectively, where p^A and p^B are the A and B elements of the price vector $\mathbf{p}$. The cumulative covariance estimator is given by

$$CC_{(A,B)} = \sum_{j_A, j_B} r_{j_A} r_{j_B} \mathbb{1}_{\{(t_{j_A - 1}, t_{j_A}] \cap (t_{j_B - 1}, t_{j_B}] \neq \emptyset\}}.$$

[3] See Equation 7 in Zhang (2006*b*).

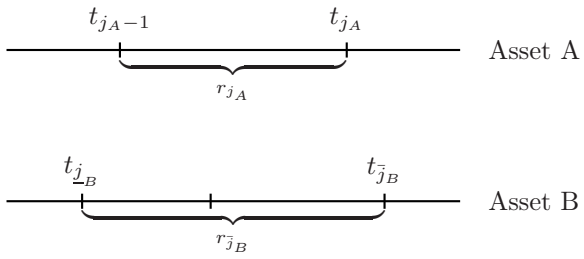

Figure 1.2: An example for a pair $r_{j_A}, r_{\overline{j}_B}$, where $r_{\overline{j}_B}$ includes two tick returns of asset B.

The defining feature of this estimator is that it adds products of returns to the sum as long as the corresponding intervals overlap. Thus, a given tick return of asset A (which we will call the *base asset*), r_{j_A}, is multiplied by (possibly) several tick returns of asset B, spanning a period starting before or at t_{j_A-1} and reaching beyond or up to t_{j_A}. It is because of this particular feature, and due to the martingale properties of the true price process, that this estimator does not suffer from the Epps effect. It is important to realize that from a practitioner's point of view, this estimator is easy to implement and does not rely on choices of synchronization methods and sampling schemes. These features make it attractive to use in practice.
In order to derive a more workable expression for the estimator, define

$$\underline{j}_B(j_A) : t_{\underline{j}_B} = \max\{t \in \Pi_B : t \leq t_{j_A-1}\} \quad \text{and} \quad \overline{j}_B(j_A) : t_{\overline{j}_B} = \min\{t \in \Pi_B : t \geq t_{j_A}\};$$

that is, the most recent transaction of asset B before t_{j_A-1}, and the first one following t_{j_A}, respectively. Then, (suppressing the dependence on j_A) define $r_{\overline{j}_B} = p^B_{\overline{j}_B} - p^B_{\underline{j}_B}$, which could be a simple tick return if $(\underline{j}_B, \overline{j}_B)$ form a pair (t_{j_B-1}, t_{j_B}) or a sum of several tick returns (e.g., see Figure 1.2 where $r_{\overline{j}_B}$ consists of 2 tick returns of asset B). As usual, we have that $r_{\overline{j}_B} = r^*_{\overline{j}_B} + e_{\overline{j}_B}$, the sum of the efficient return and the noise return over the interval $(\underline{j}_B, \overline{j}_B]$. Then the estimator can be written as

$$CC_{(A,B)} = \sum_{j_A=1}^{M_A} r_{j_A} r_{\overline{j}_B}.$$

We note here that a change of the base asset does not change the estimator. In practice, it is easier to set the less frequently traded asset as the base asset, which also determines the order of the discretization error in the variance of the estimator.[4]

[4]If, for example, one of the assets trades very frequently, but the other one trades once at the beginning of the trading day and once more at the end, then the CC estimator is simply the cross-product of daily returns.

Under our general noise specification, we can derive the following theorem:

Theorem 1.2. *Under Assumptions 1.1 and 1.2, the bias of the cumulative covariance estimator is given by*

$$\mathrm{E}\left[CC_{(A,B)} - IC_{(A,B)}\right] = b_{M_A} + c_{M_A} + d_{M_A}, \tag{1.4}$$

where

$$b_{M_A} = \mathrm{E}\left[\sum_{j_A=1}^{M_A} e_{j_A} r^*_{\overline{j}_B}\right], \quad c_{M_A} = \mathrm{E}\left[\sum_{j_A=1}^{M_A} r^*_{j_A} e_{\overline{j}_B}\right], \quad d_{M_A} = \mathrm{E}\left[\sum_{j_A=1}^{M_A} e_{j_A} e_{\overline{j}_B}\right].$$

Proof. See the Appendix.

The three components of the bias arise due to dependence of the noise process A and the price process B, dependence of the noise process B and the price process A, and (serial) cross-correlation between the noise processes, respectively. Intuitively, the terms b_{M_A} and c_{M_A} will be of minor practical importance, since dependence between the price of one asset and the noise of another is improbable. The term d_{M_A}, though, will not be innocuous, as will be confirmed in our empirical analysis.
In the absence of noise, the estimator is unbiased since all terms vanish. Interestingly, if the arrival processes of both assets are orderly, it remains unbiased in the presence of i.i.d. noise, even if there is contemporaneous correlation across the noise elements(i.e., $\mathbf{\Gamma}(0)$ is not diagonal). This follows since in this case the probability of A and B trading at the same time is zero. In fact, if the noise is exogenous to the price process, then it suffices that $\gamma(l) = 0$ for $l \neq 0$, which is a milder condition than independence. The rejection of this condition by the data, however, necessitates the use of bias-correction in practice.

1.4 Extensions of the CC Estimator: Bias Correction and Subsampling

In order to derive the variance of the CC estimator, which plays an important role in our further developments, we first focus on the i.i.d. noise assumption.

Assumption 1.3.

(*i*) $\mathbf{p}^* \perp\!\!\!\perp \mathbf{u}$; $\mathbf{u}(s) \perp\!\!\!\perp \mathbf{u}(t), s \neq t$; *and* $\mathrm{E}[\mathbf{u}(t)] = 0$ *for all* t;

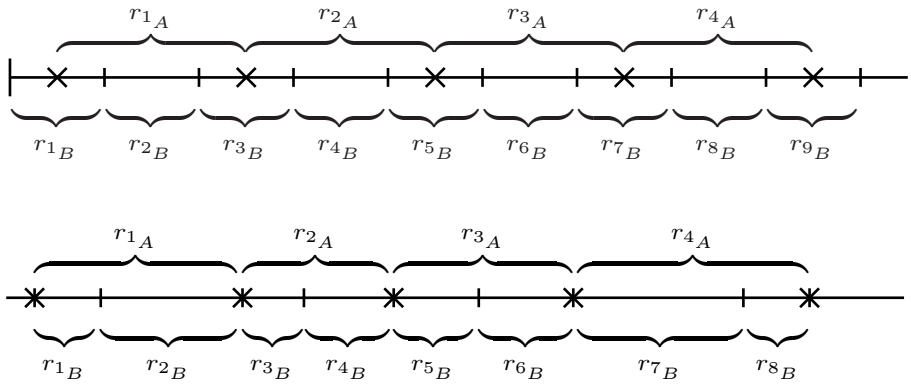

Figure 1.3: The top panel presents a regular non-synchronous trading scenario, and the lower panel gives an irregular synchronous trading scenario. Trading times of assets A and B are represented by a cross ($\times$) and a tick ($|$), respectively.

(*ii*) $\mathrm{E}\left[\mathbf{u}(t)\mathbf{u}'(t)\right] = \mathbf{\Gamma}(0) = \Omega$, *a matrix with finite elements for all t;*

(*iii*) $\mathrm{E}\left[\mathbf{u}(t)\mathbf{u}'(t) \otimes \mathbf{u}(t)\mathbf{u}'(t)\right] = \mu_4$, *a* $K^2 \times K^2$ *matrix with finite elements for all t.*

This noise specification allows only for contemporaneous correlation in the noise and forbids any dependence between the noise and the true price process. If Ω is diagonal, we can view this as the simple bid-ask bounce model of Roll (1984). This assumption has been heavily used in the extant literature on high-frequency (co-) volatility estimation. In the multivariate setting, Zhang (2006*b*) derives the MSE and optimal sampling frequency of the realized covariance, while in an independent and concurrent study Griffin & Oomen (2006) examine also its lead- and lag- adjustments and the CC estimator under Poisson arrival times. For our purposes, we only consider the CC estimator and compare our findings to the results of Griffin & Oomen (2006). In the following two lemmas, we examine two extreme trading scenarios in which asset A is trading twice slower than asset B and:

i.) both assets trade in regular intervals and no trade of asset A occurs simultaneously with a trade of asset B, which we will call *regular non-synchronous trading* (see the upper panel of Figure 1.3);

ii.) asset B's trading times are random and asset A trades every second time asset B does, which we will call *irregular synchronous trading* (see the lower panel of Figure 1.3).

Figure 1.3 illustrates a sequence of several tick returns of asset A and B in both trading scenarios. In both cases A is the slower asset and is the base asset used to

compute the CC estimator. These two extreme scenarios allow us to examine how the degree of non-synchronicity affects the variance of the CC estimator.

Lemma 1.1. *Given Assumptions 1.1 and 1.3, and regular non-synchronous trading, the variance of the CC estimator is*

$$\begin{aligned}\mathrm{V}\left[CC_{(A,B)}\right] = &\sum_{j_A=1}^{M_A}\left(\sigma_{j_A,(AA)}\sigma_{\overline{j}_B(j_A),(BB)} + \sigma^2_{j_A,(AB)}\right) + 2\omega_{AA}\sum_{j_B=1}^{M_B}\sigma_{j_B,(BB)} \\ &+ 2\omega_{BB}\sum_{j_A=1}^{M_A}\sigma_{j_A,(AA)} + 4M_A\omega_{AA}\omega_{BB} + 2\sum_{j_A=1}^{M_A-1}\sigma_{t_{j_A}:t_{\overline{j}_B(j_A)},(AB)}\sigma_{t_{\underline{j}_B(j_A+1)}:t_{j_A},(AB)}.\end{aligned}$$

The expressions $\sigma_{t_{j_A}:t_{\overline{j}_B(j_A)},(AB)}$ *and* $\sigma_{t_{\underline{j}_B(j_A+1)}:t_{j_A},(AB)}$ *denote the integrated covariance over* $(t_{j_A}, t_{\overline{j}_B(j_A)}]$ *and* $(t_{\underline{j}_B(j_A+1)}, t_{j_A}]$*, respectively.*

Proof. See the Appendix.

The odd-looking last term in the expression above appears because whenever there is non-synchronicity, the summands in the CC will be first-order autocorrelated. This arises since the two neighboring returns of the base asset, r_{j_A} and r_{j_A+1}, are both multiplied with the return of asset B from $t_{\underline{j}_B(j_A+1)}$ to $t_{\overline{j}_B(j_A)}$, resulting in accumulation of additional discretization error compared to the case with synchronicity presented below. To simplify the expression, we introduce the parameter $\kappa = \mu_{4(AA,BB)}/\mu_{4\text{norm}}$ as the cross fourth moment, relative to the normal distribution.[5]

Lemma 1.2. *Given Assumptions 1.1 and 1.3, and irregular synchronous trading, the variance of the CC estimator is*

$$\begin{aligned}\mathrm{V}\left[CC_{(A,B)}\right] = &\sum_{j_A=1}^{M_A}(\sigma_{j_A,(AA)}\sigma_{j_A,(BB)} + \sigma^2_{j_A,(AB)}) + 2\omega_{AA}\sum_{j_B=1}^{M_B}\sigma_{j_B,(BB)} \\ &+ 2\omega_{BB}\sum_{j_A=1}^{M_A}\sigma_{j_A,(AA)} + 2(2\kappa+1)M_A\omega_{AA}\omega_{BB} + 2(4\kappa-1)M_A\omega^2_{AB} \\ &- 2\kappa\omega_{AA}\omega_{BB} - 2(2\kappa-1)\omega^2_{AB} + 2\omega_{AB}\sum_{j_A=1}^{M_A}\sigma_{j_A,(AB)}.\end{aligned}$$

Proof. See the Appendix.

[5]From Anderson (2003) we use that if $X = (X_1, X_2) \sim N\left(\mathbf{0}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right)$, then $\mu_{4\text{norm}} \equiv \mathrm{E}\left[X_1^2X_2^2\right] = \sigma_{11}\sigma_{22} + 2\sigma_{12}^2$ (note the similarity to $\psi_{i,(AA,BB)}$ in (1.2)).

It is clear that in both cases the estimator is inconsistent, and the minimum of the variance will be attained for a finite sampling frequency, which balances between the discretization error and the impact of the noise. Comparing our variance expressions to the variance derived in Griffin & Oomen (2006) under their condition $\omega_{AB} = 0$, we recognize the same terms, with the difference that in our derivation the number of observations is explicit ($M_B = 2M_A$), while in theirs, it depends on the intensities of the corresponding processes.

Examining the two variance expressions, it is revealed that in the non-synchronous case the variance due to discretization is relatively larger (due to the last term in the variance expression in Lemma 1.1), while the variance due to noise is relatively smaller. This is intuitive, since on the one hand, the more non-synchronicity there is, the more products of non-overlapping returns are contributing to the variance. On the other hand, synchronicity brings about accumulation of noise (that is, of course, if there is some contemporaneous dependence across assets). Therefore, synchronicity influences the ratio of noise-to-discretization induced variance. In addition, under our assumptions, more synchronicity induces more bias. Both effects lead to higher optimal sampling frequency when less synchronicity is present.

In what follows we will present some new results on the CC estimator in the presence of correlated noise. Our focus on this particular estimator and noise specification is motivated firstly by our empirical section, where it is shown that the noise is not i.i.d. Secondly, the CC estimator successfully solves at least the non-synchronicity issue and it seems to be a promising endeavor to examine it whenever noise is present.

Regarding the dependence of the noise, we need one additional assumption, which is similar to the one in Bandi & Russell (2005*b*) in that it allows for temporal dependence in the noise components. Here, we make a further generalization by allowing the noise to be contemporaneously as well as serially correlated with the price process.

Assumption 1.4. *The noise process has finite dependence in the sense that* $\mathbf{\Gamma}(l) = \mathbf{0}$ *for all* $l > \theta_0$ *for some finite* $\theta_0 \geq 0$ *and* $\mathrm{E}\left[\mathbf{u}(t)|\,\mathbf{p}^*(l)\right] = 0$ *for* $|t - l| > \theta_0$.

An important feature of this noise specification is that the dependence vanishes after a finite displacement θ_0. Since in market microstructure theory, the noise is viewed as having a transient influence as opposed to the persistent effect of fundamental information, we find this assumption theoretically sound. In Section1.3, we illustrated that the cumulative covariance estimator is a sum of products of the type $r_{j_A} r_{\bar{j}_B}$, where $r_{\bar{j}_B}$ is such that it spans an interval which contains the interval $(t_{j_A-1}, t_{j_A}]$.

Then it follows that if θ_0 is small, while the differences $|t_{j_A-1} - t_{\underline{j}_B}|$ and $|t_{\overline{j}_B} - t_{j_A}|$ are large, the estimator will still remain unbiased because the dependence of the noise is short-lived relative to the trading intensity of the assets under consideration. This scenario, however, is not likely when the assets trade actively.

Having established the bias of the CC estimator in Theorem 1.2, we now turn to the problem of bias correction. Assume we choose b^+ and b^- as the number of ticks such that for each j_A, $|t_{j_A-1} - t_{\underline{j}_B-b^-}| > \theta_0$ and $|t_{j_A} - t_{\overline{j}_B+b^+}| > \theta_0$. We allow for different values of the correction parameters b^- (lags) and b^+ (leads), as we can have asymmetric dependence patterns. The choice of θ_0 in practice will be made clear, when we discuss a feasible way to estimate $\gamma(l)$ and to choose b^+ and b^- later in this section.[6] Based on the univariate kernels of Hansen & Lunde (2006), we suggest the following bias-corrected cumulative covariance estimator

$$CC^{bc}_{(A,B)} = \sum_{j_A=1}^{M_A} r_{j_A} \left(p^B_{\overline{j}_B+b^+} - p^B_{\underline{j}_B-b^-} \right),$$

for which we provide the following theorem.

Theorem 1.3. *Given Assumptions 1.1, 1.2 and 1.4, and if b^+ and b^- are chosen as described above, we have that*

$$\mathrm{E}\left[CC^{bc}_{(A,B)} - IC_{(A,B)}\right] = 0.$$

Proof. See the Appendix.

The kernel structure of the estimator can be observed in the following decomposition

$$\begin{aligned} CC^{bc}_{(A,B)} &= \sum_{j_A=1}^{M_A} \left(r_{j_A} \left(\sum_{h=0}^{b^--1} r_{\underline{j}_B-h} + r_{\overline{j}_B} + \sum_{h=1}^{b^+} r_{\overline{j}_B+h} \right) \right) \\ &= \underbrace{\sum_{j_A=1}^{M_A} r_{j_A} r_{\overline{j}_B}}_{CC_{(A,B)}} + \sum_{j_A=1}^{M_A} \sum_{h=0}^{b^--1} r_{j_A} r_{\underline{j}_B-h} + \sum_{j_A=1}^{M_A} \sum_{h=1}^{b^+} r_{j_A} r_{\overline{j}_B+h}, \end{aligned}$$

which also shows that one can think of this estimator as a non-synchronous version of the general lead-lag corrected realized covariance with b^- lags and b^+ leads. We do not provide a statement about the consistency of this estimator, but since the kernel

[6] In the empirical section we find that the lead and lag dependence can be quite asymmetric and hence the cutoff point θ_0 will differ for the lead and lag correction.

bandwidth is always greater than θ_0 (does not converge to zero), we hypothesize that it is not consistent (see the discussion in Section 4 in Hansen & Lunde (2006)). A similar bias correction is also proposed independently in the final section of Griffin & Oomen (2006). They leave it open for exploration, recognizing the difficulty of bias correction in tick time leading to asymmetry in the bias-corrected estimator with respect to the ordering of assets. This difficulty arises if we attempt to define dependence in tick time which, as mentioned above, is problematic. As a solution, we suggest to choose b^- and b^+ in ticks, but in such a way that the kernel window spans approximately the same *period of calendar time* for each j_A.

In order to determine the values of b^- and b^+ in the bias correction, we need an estimate of the cross-correlation function $\gamma(l)$. Choosing a correct width for the kernel is important, as a kernel which is too narrow will not eliminate the bias, while a kernel which is too wide will increase the variance of the estimator. The key idea here is to use products of non-overlapping returns, for which the expectation of the cross-products of efficient returns is zero due to the martingale assumption. We use the following example to demonstrate this idea for the case $l \leq 0$. For positive values of l the same type of argument is valid if we change the ordering of the assets.

Example 1.2. *Consider the pair of non-overlapping returns $r_{j_A} r_{\underline{j}_B}$ depicted in Figure 1.4 and denote*

$$\Delta_{\underline{j}_B} = t_{\underline{j}_B} - t_{\underline{j}_B - 1}, \qquad \Delta_{j_A \underline{j}_B} = t_{j_A - 1} - t_{\underline{j}_B}, \qquad \Delta_{j_A} = t_{j_A} - t_{j_A - 1}.$$

Since $\mathrm{E}\left[r^*_{j_A} r^*_{\underline{j}_B}\right] = 0$ *due to the martingale assumption, we have that*

$$\begin{aligned}\mathrm{E}\left[r_{j_A} r_{\underline{j}_B}\right] = &\underbrace{\mathrm{E}\left[\left(u^A_{j_A} - u^A_{j_A-1}\right)\left(p^{*B}_{\underline{j}_B} - p^{*B}_{\underline{j}_B-1}\right)\right]}_{(1)} + \underbrace{\mathrm{E}\left[\left(p^{*A}_{j_A} - p^{*A}_{j_A-1}\right)\left(u^B_{\underline{j}_B} - u^B_{\underline{j}_B-1}\right)\right]}_{(2)} \\ &+ \underbrace{\mathrm{E}\left[\left(u^A_{j_A} - u^A_{j_A-1}\right)\left(u^B_{\underline{j}_B} - u^B_{\underline{j}_B-1}\right)\right]}_{(3)}.\end{aligned}$$

Terms (1) and (2) are asymptotically zero when $\Delta_{j_A} \to 0$ and $\Delta_{\underline{j}_B} \to 0$. Furthermore, as mentioned earlier, the existence of dependence between the price and the noise of different assets is not very realistic. Therefore, we set expressions (1) and

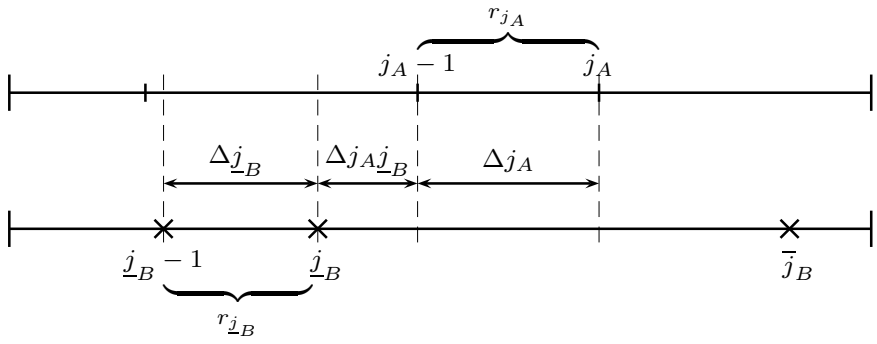

Figure 1.4: Estimation of the cross-covariance function $\gamma(l)$: illustration of the notation.

(2) to zero and working out expression (3) we are left with

$$\mathrm{E}\left[r_{j_A} r_{\underline{j}_B}\right] = \underbrace{\gamma\left(-\Delta_{j_A \underline{j}_B} - \Delta_{j_A}\right)}_{(a)} - \underbrace{\gamma\left(-\Delta_{\underline{j}_B} - \Delta_{j_A \underline{j}_B} - \Delta_{j_A}\right)}_{(b)} - \underbrace{\gamma\left(-\Delta_{j_A \underline{j}_B}\right)}_{(c)} + \underbrace{\gamma\left(-\Delta_{\underline{j}_B} - \Delta_{j_A \underline{j}_B}\right)}_{(d)}.$$

In the expression above we are interested in an estimate of (c)*, which is an estimate of* $\gamma(l)$ *for* $l = -\Delta_{j_A \underline{j}_B}$*. Thus, we need to change the end points of the returns (i.e.,* $t_{\underline{j}_B - 1}$ *and* t_{j_A}*), so that the arguments of the function* $\gamma(\cdot)$ *in expressions* (a)*,* (b) *and* (d) *become larger in absolute value than* θ_0*, and hence by virtue of Assumption 1.4, the expressions vanish. This involves an initial guess for* θ_0*, i.e., a guess about the degree of persistence of the noise.*

As the previous example shows, $\gamma(l)$ can be estimated if we choose returns of appropriate length and average over suitable product pairs of non-overlapping returns. For the initial guess for θ_0 we use a result from Hansen & Lunde (2006), who show that for actively traded stocks, returns sampled approximately at one-minute intervals seem to conform to the i.i.d. noise assumption. In our estimation, however, there are cases, for which one minute is not long enough for the dependence to disappear, so we have to use longer periods. We find that even for the most persistent cases, at most six-minute periods are sufficient.

In particular, we implement the estimation of $\gamma(l)$ as follows. We sample returns in tick time, so that there are approximately x seconds between the ticks as described in Barndorff-Nielsen et al. (2006). Denote the number of subgrids thus obtained by S_A and S_B, for asset A and B, respectively. Then for each $s_k = 1, \ldots, S_k$, $k = A, B$, we define the subgrid returns $r_{j_k^s} = p^k_{j_k^s} - p^k_{j_k^s - 1}$, with $\{t_{j_k^s}\}_{j_k^s = 1, \ldots, M_k^s}$ being

the subset of Π_k corresponding to the s_k-subgrid.[7] Assume we want to estimate $\gamma(l)$ for $l = -L, \ldots, L$. For each pair of subgrids of both assets, if $l > 0$ the estimation procedure averages over pairs of non-overlapping returns, such that $t_{j_s^B - 1} - t_{j_s^A} = l$. For $l < 0$, the corresponding condition is $t_{j_s^A - 1} - t_{j_s^B} = -l$. For $l = 0$, we can use returns that fulfil both conditions. Therefore, we can define the estimator

$$\hat{\gamma}(l) = \frac{1}{S_A S_B} \sum_{s_A=1}^{S_A} \sum_{s_B=1}^{S_B} \hat{\gamma}(l)_s,$$

where

$$\hat{\gamma}(l)_s = \begin{cases} -\frac{1}{n_{l,s}} \sum_{j_A^s=1}^{M_A^s} \sum_{j_B^s=1}^{M_B^s} r_{j_A^s} r_{j_B^s} \mathbb{1}_{\{t_{j_B^s - 1} - t_{j_A^s} = l\}} & \text{if} \quad l > 0 \\ -\frac{1}{n_{l,s}} \sum_{j_A^s=1}^{M_A^s} \sum_{j_B^s=1}^{M_B^s} r_{j_A^s} r_{j_B^s} \mathbb{1}_{\{t_{j_A^s - 1} - t_{j_B^s} = 0\} \cup \{t_{j_B^s - 1} - t_{j_A^s} = 0\}} & \text{if} \quad l = 0 \\ -\frac{1}{n_{l,s}} \sum_{j_A^s=1}^{M_A^s} \sum_{j_B^s=1}^{M_B^s} r_{j_A^s} r_{j_B^s} \mathbb{1}_{\{t_{j_A^s - 1} - t_{j_B^s} = -l\}} & \text{if} \quad l < 0 \end{cases}$$

with

$$n_{l,s} = \begin{cases} \sum_{j_A^s=1}^{M_A^s} \sum_{j_B^s=1}^{M_B^s} \mathbb{1}_{\{t_{j_B^s - 1} - t_{j_A^s} = l\}} & \text{if} \quad l > 0 \\ \sum_{j_A^s=1}^{M_A^s} \sum_{j_B^s=1}^{M_B^s} \mathbb{1}_{\{t_{j_A^s - 1} - t_{j_B^s} = 0\} \cup \{t_{j_B^s - 1} - t_{j_A^s} = 0\}} & \text{if} \quad l = 0 \\ \sum_{j_A^s=1}^{M_A^s} \sum_{j_B^s=1}^{M_B^s} \mathbb{1}_{\{t_{j_A^s - 1} - t_{j_B^s} = -l\}} & \text{if} \quad l < 0 \end{cases}$$

In order to be able to judge the statistical significance of our estimates, we derive the first order term in the variance of the estimator under the null hypothesis of independent noise processes, which is a corollary of Lemmas 1.1 and 1.2.

Corollary 1.1. *Under Assumptions 1.1 and 1.3, and if* $\mathbf{\Gamma}(0)$ *is a diagonal matrix, the variance of* $\hat{\gamma}(l)$ *can be approximated by*

- *For* $l = 0$*:*

$$\mathrm{V}\left[\hat{\gamma}(0)\right] \approx \frac{1}{(S_A S_B)^2} \sum_{s=1}^{S_A S_B} \frac{1}{n_{0,s}} (\sigma_{x\ sec,(AA)} \sigma_{x\ sec,(BB)} + 2\omega_{BB} \sigma_{x\ sec,(AA)} + 2\omega_{AA} \sigma_{x\ sec,(BB)} + 5\omega_{AA} \omega_{BB});$$

[7] We will use the shorthand subscript s in place of s^A, s^B to denote a quantity which refers to the subsample pair s^A, s^B, where $s^k = 1, \ldots, S^k$, $k = A, B$.

- *For $l \neq 0$:*

$$\mathrm{V}\left[\hat{\gamma}(l)\right] \approx \frac{1}{(S_A S_B)^2} \sum_{s=1}^{S_A S_B} \frac{1}{n_{l,s}} (\sigma_{x\ sec,(AA)} \sigma_{x\ sec,(BB)} + 2\omega_{BB}\sigma_{x\ sec,(AA)} + 2\omega_{AA}\sigma_{x\ sec,(BB)} + 4\omega_{AA}\omega_{BB}),$$

where $\sigma_{x\ sec,(kk)} = \frac{\int_0^1 \sigma_{kk}(u)du}{23400/x}$ is the x-second average integrated variance, $k = A, B$.

Proof. See the Appendix.

While Theorem 1.3 provides us with a way to bias-correct the CC, the problem remains that the resulting estimator is inconsistent, as we have discussed above. Therefore, we examine the possibility of subsampling, which is studied by Zhang et al. (2005) in the univariate case, who show that it leads to a trade-off between discretization and noise error. If the number of subgrids is chosen to grow with M at an optimal rate, then consistency can be achieved by balancing between these two sources of error. We implement subsampling by choosing a number of subgrids for the base asset, and then for each subgrid computing the CC estimator or its bias corrected version and averaging over the subgrids. We denote these estimators by CC^{sub} and CC^{sub-bc}, respectively.

Choosing the optimal number of subgrids is not straightforward. Here, we derive an expression for the case of uncorrelated (across time) noise. We assume that we have an unbiased estimator of a realized quantity based on M intraday observations, say $Q^{(M)}$. In general, its variance has three distinct parts: a term of order $O\left(M^{-1}\right)$, a term of order $O\left(1\right)$, and a term of order $O\left(M\right)$ (see, e.g. Hansen & Lunde (2006), BNS (2004) and Lemmas 1.1 and 1.2). The $O\left(1\right)$ term does not depend on the sampling frequency and we do not consider it further. The $O\left(M^{-1}\right)$ term is due to discretization, and the $O\left(M\right)$ is due to noise. Therefore, we can express the variance of $Q^{(M)}$ as $\mathrm{V}\left[Q^{(M)}\right] = f(\text{integrated quantities of the efficient price process})/M + Mg(\text{noise moments})$, where the functions $f(\cdot)$ and $g(\cdot)$ are both of order $O\left(1\right)$, e.g. in the case of realized variance with normal noise f is twice the integrated quarticity and g is $12\omega^4$. Now, assume we take q subgrids, meaning that within each subgrid, we compute the realized quantity with returns sampled at q ticks (or regularly spaced intervals). Then, the number of points in each subgrid will be roughly $m = M/q$ and we compute the realized quantity $Q_i^{(m)}$ for each subgrid $i = 1, \ldots, q$. We have

that

$$\mathrm{V}\left[\frac{1}{q}\sum_{i=1}^{q} Q_i^{(m)}\right] = \frac{1}{q^2}\left(\sum_{i=1}^{q}\mathrm{V}\left[Q_i^{(m)}\right] + \sum_{i=1}^{q}\sum_{j\neq i}\mathrm{Cov}\left[Q_i^{(m)}, Q_j^{(m)}\right]\right).$$

The covariance of two realized quantities measured on different subgrids is of the same order as the variance and will depend on the same integrated quantities, so we write it again as $f(\cdot)$. As for the noise, its moments will not appear in the covariance expression since we have assumed uncorrelatedness and no subgrids share observations. Then, what we obtain up to the correct order is

$$\mathrm{V}\left[\frac{1}{q}\sum_{i=1}^{q} Q_i^{(m)}\right] = \underbrace{\frac{1}{q^2}\left(q\frac{f(\cdot)}{m} + q^2\frac{f(\cdot)}{m}\right)}_{distretization} + \underbrace{\frac{1}{q^2}\big(qmg(\cdot)\big)}_{noise}. \tag{1.5}$$

Keeping in mind that $m = M/q$, we have that the discretization term is of order $O\left(\frac{q}{M}\right)$, while the noise term is of order $O\left(\frac{M}{q^2}\right)$. Choosing $q = cM^{2/3}$, where c is a constant, makes both terms of order $O\left(M^{-1/3}\right)$ and the estimator is consistent.
In order to determine the optimal choice of c, we rewrite Equation (1.5) with $q = cM^{2/3}$ and obtain

$$\begin{aligned}\mathrm{V}\left[\frac{1}{q}\sum_{i=1}^{q} Q_i^{(m)}\right] &= \frac{1}{c^2M^{4/3}}\left(\left(cM^{2/3} + c^2M^{4/3}\right)cM^{-1/3}f(\cdot)\right) + \frac{1}{c^2M^{4/3}}\big(Mg(\cdot)\big)\\ &= \frac{f(\cdot)}{M} + \frac{cf(\cdot)}{M^{1/3}} + \frac{g(\cdot)}{c^2M^{1/3}}.\end{aligned}$$

Minimizing this expression with respect to c we obtain

$$c^* = \sqrt[3]{\frac{2g(\cdot)}{f(\cdot)}},$$

i.e., the optimal c is the third root of twice the noise-to-signal ratio. Zhang et al. (2005) arrive at this result in a much more formal setting for the subsampled bias-corrected realized variance. In their example with constant volatility σ, the function $f(\cdot) = \frac{8}{3}\sigma^4$ and $g(\cdot) = 8\omega^4$ (in their notation $8\nu^2$) and hence they obtain

$$c^* = \sqrt[3]{\frac{6\omega^4}{\sigma^4}}.$$

In the bivariate case (e.g. the CC estimator), M corresponds to the asset with less

observations. Thus, for our synchronous trading scenario in Lemma 1.2, $M = M_A$, $f(\cdot) = \int_0^1 (\sigma_{AA}(u)\sigma_{BB}(u) + \sigma_{AB}^2(u))du$ and $g(\cdot) = 2(2\kappa+1)\omega_{AA}\omega_{BB} + 2(4\kappa-1)\omega_{AB}^2$. This might lead to a choice of q, which is slightly higher than optimal, due to the smaller discretization error and larger noise error compared to the more plausible scenario of non-synchronous trading, but it gives us a good approximation.

1.5 Empirical Results and Monte Carlo Study

1.5.1 Empirical Results

In this section, we evaluate the estimators discussed in Sections 1.3 and 1.4 with an application to stock returns and a simulation experiment.

We extract quote data for the 30 DJIA stocks in 2004 from the NYSE Trade and Quotations (TAQ) database. For the data cleaning procedures, we refer the reader to Barndorff-Nielsen et al. (2006). The first step is to estimate the function $\gamma(l)$ based on the non-synchronized observations as described in the previous section. We find evidence that there are significant lead and lag effects among pairs of DJIA stocks, which cannot be explained within the martingale plus i.i.d. noise framework, but can be analyzed within our theoretical setup. This finding must be interpreted carefully with respect to its implications regarding market efficiency. Market efficiency and the incorporation of information into security prices is achieved by trading, which eliminates mispricing by exploiting possible deviations of the price from the fundamental value of the asset. Hence, if transaction costs and/or illiquidity prevent investors from trading profitably on such lead-lag patterns, we cannot claim that prices are inefficient, but rather it takes time for the market to assimilate the new information. For this reason, we refer to these cross correlation patterns as arising from MMS noise. Some evidence of the speed of convergence to efficiency appears in Chordia, Roll & Subrahmanyam (2005). Using a sample of 150 NYSE stocks, they find that order imbalances seem to predict future returns, but it takes between five to sixty minutes for traders to react and undertake trades to remove this dependence. Thus, as they note, "efficiency does not simply congeal after spontaneous combustion".

We also estimate noise variances ω_{AA}^2 and ω_{BB}^2 by subsampling realized variance with one-minute returns and dividing by twice the number of returns (denoted by $\hat{\omega}^2$ as in Barndorff-Nielsen et al. (2006)), and report noise cross-correlation functions computed as $\varrho(l) = \hat{\gamma}(l)/(\hat{\omega}_{AA}\hat{\omega}_{BB})$. Note that these correlation estimates are somewhat

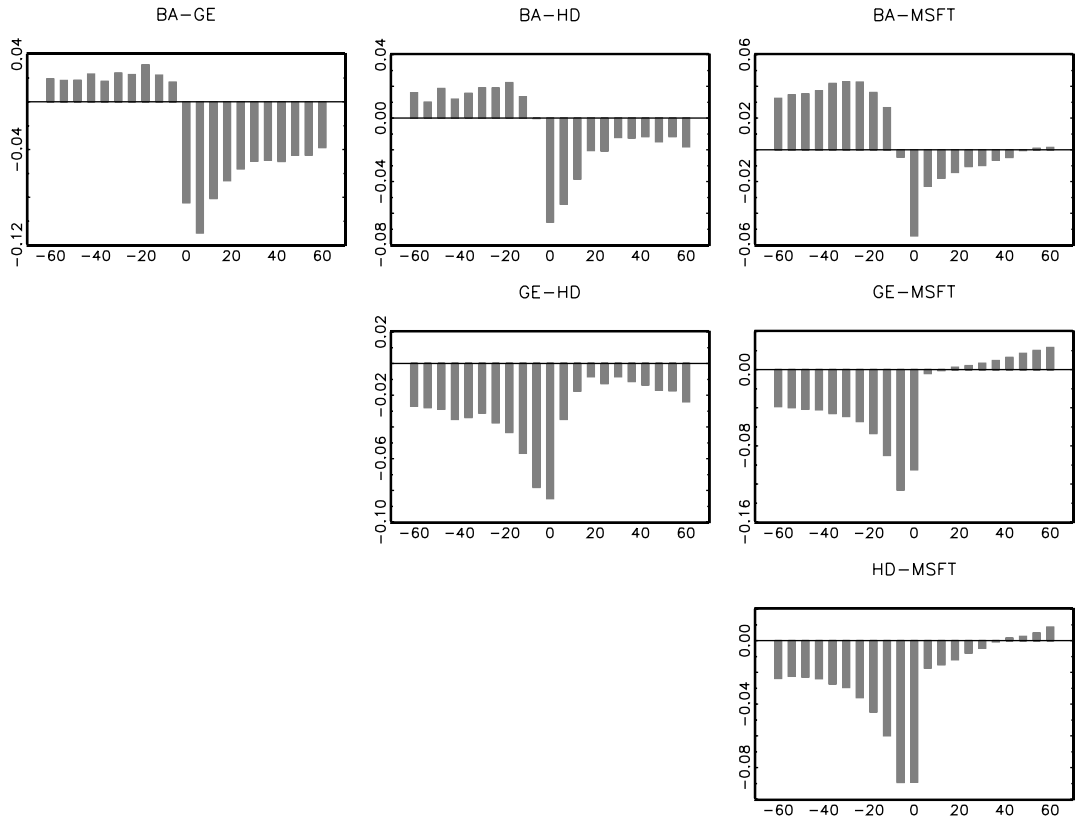

Figure 1.5: Cross-correlation functions $\varrho(l)$ of the noise process of several stock pairs from the DJIA index in 2004. The x-axis corresponds to leads and lags. Every 6^{th} lag (lead) is plotted.

downward biased, since $\hat{\omega}^2_{kk}$ is upward biased by $\int_0^1 \sigma_{kk}(u)du/(2M_k)$, $k = A, B$. We depict part of the results in Figure 1.5. Out of the total of 435 covariance pairs in the 30 DJIA stocks, we find that in most cases the dependence disappears within 60 seconds with the notable exception of the pairs with the NASDAQ stocks – INTC and MSFT. To present our results we have chosen MSFT as one of the two stocks quoted on the NASDAQ, and three stocks quoted on the NYSE: GE – with high quoting activity, and BA and HD – with average quoting activity. Although we observe different patterns, it is evident that the autocorrelation function is not symmetric and hence b^- will in general be different from b^+. As mentioned above b^+ and b^- are chosen in ticks, but dependent on j_A so that the intervals $(t_{\underline{j}_B - b^-}, t_{j_A - 1}]$ and $(t_{j_A}, t_{\overline{j}_B + b^+}]$ are approximately of the same length across j_A, corresponding to the cutoff point θ_0, which is chosen so that the dependence of the noise processes becomes insignificant. We judge the significance from the t-statistics at the 5% significance level, i.e. if |t-stat| < 1.96 we count the lag (lead) as insignificant.

Table 1.1: T-statistics for the significance of the cross covariance function $\gamma(l)$ of the noise process of several stock pairs from the DJIA index in 2004. The t-statistics are computed as $t = \hat{\gamma}(l)/\sqrt{\mathrm{V}\left[\hat{\gamma}(l)\right]}$, where $\mathrm{V}\left[\hat{\gamma}(l)\right]$ is computed as in Corollary 1.1. The integrated variances are estimated by a kernel method (see Barndorff-Nielsen et al. (2006)). Results are reported for every 6^{th} lag (lead) in seconds.

Lead (Lag)	t-statistics					
	BA-GE	BA-HD	BA-MSFT	GE-HD	GE-MSFT	HD-MSFT
-60	2.21	1.67	6.48	-3.01	-8.12	-4.56
-54	2.04	1.05	6.91	-3.12	-8.34	-4.29
-48	2.08	1.96	7.07	-3.24	-8.68	-4.38
-42	2.66	1.26	7.43	-3.93	-8.83	-4.59
-36	1.97	1.65	8.39	-3.82	-9.59	-5.24
-30	2.78	2.00	8.59	-3.50	-10.28	-5.64
-24	2.60	2.00	8.58	-4.19	-11.36	-6.88
-18	3.56	2.36	7.26	-4.88	-13.97	-8.61
-12	2.56	1.42	5.32	-6.35	-18.82	-11.49
-6	1.91	-0.04	-0.96	-8.78	-26.37	-17.21
0	-13.92	-10.03	-13.83	-12.39	-27.83	-21.73
6	-12.61	-5.76	-4.61	-3.98	-0.90	-3.34
12	-9.28	-4.05	-3.61	-1.96	-0.24	-2.92
18	-7.55	-2.15	-2.85	-0.93	0.50	-2.32
24	-6.44	-2.19	-2.12	-1.44	0.85	-1.52
30	-5.70	-1.29	-1.98	-0.94	1.41	-0.92
36	-5.59	-1.32	-1.30	-1.28	2.00	-0.15
42	-5.73	-1.23	-0.99	-1.53	2.67	0.31
48	-5.12	-1.57	-0.11	-1.88	3.57	0.50
54	-5.12	-1.23	0.19	-1.94	4.19	0.93
60	-4.37	-1.91	0.32	-2.72	4.86	1.63

Table 1.1 reports the t-statistics up to lag (lead) 60 for the significance of the cross-covariance function. To estimate the variance of $\hat{\gamma}(l)$ we use the results in Corollary 1.1. While intuition might suggest that for more actively quoted stocks the dependence should vanish more quickly, especially for the pairs with MSFT we find significant dependence patterns beyond 60 seconds. For these pairs we extend our estimation for larger values of l, and for the case with most dependence, GE-MSFT, we find that it takes roughly six minutes for the quote returns to become uncorrelated. In order to assess the performance of the bias-corrected CC estimator, we have plotted covariance signature plots for the six covariance pairs of the four stocks under consideration in Figure 1.6.

As it is evident from the figure, the CC estimator is biased for small values of the sampling frequency due to dependence of the noise. In contrast, the bias-corrected

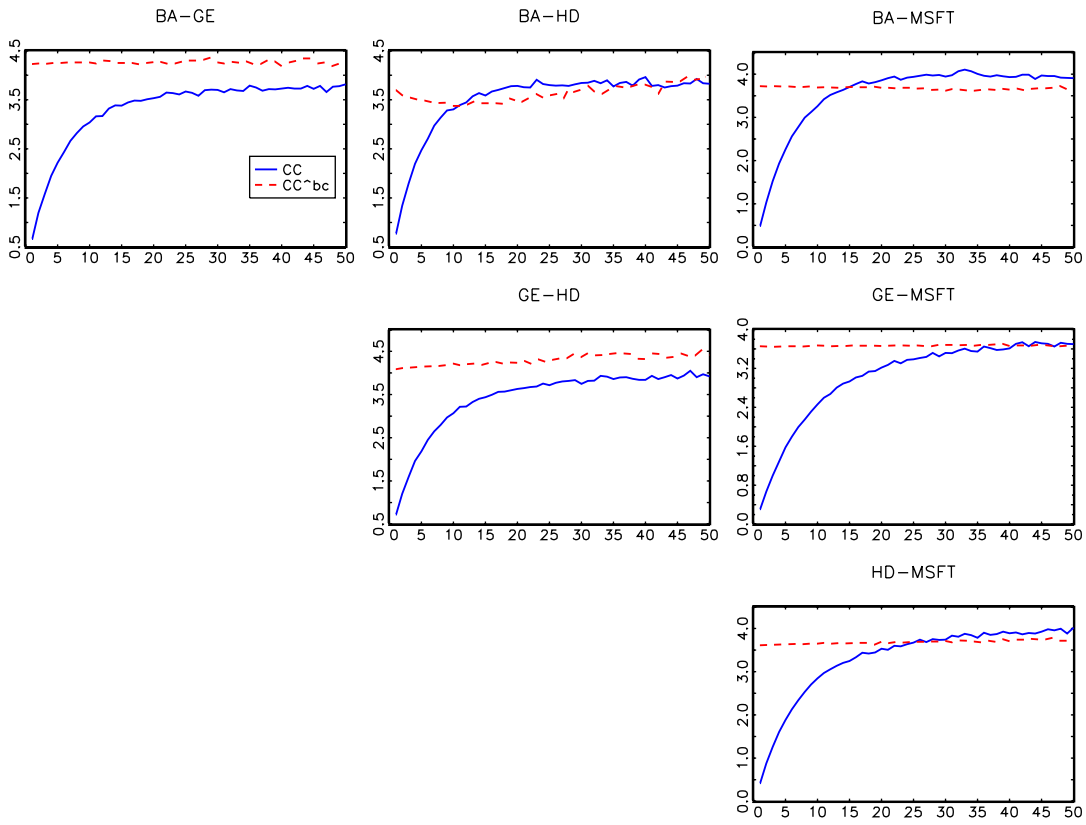

Figure 1.6: Covariance signature plots of the CC (solid line) and CC^{bc} (dashed line) estimator for the daily integrated covariance of several stock pairs from the DJIA index in 2004. The x-axis represents sampling frequency in ticks of the base asset. All values are scaled by 10^5.

estimator looks quite stable for most of the cases, especially when we have a large number of intraday observations, leading to a smaller variance. The effect of subsampling is better revealed by considering simulated data, which we undertake in the following section.

1.5.2 Simulation Design

In this simulation study we examine the bias, standard deviation, and root mean squared error (RMSE) of a set of covariance estimators in order to evaluate the impact of noise and different sampling specifications on their quality. Furthermore, we examine the effect of subsampling when the noise processes are serially cross-correlated. We have aimed at reproducing rich specifications for the price processes with stochastic volatility and correlation, and possibility for leverage.
Specifically, we simulate two univariate price processes $p^{*A}(t)$ and $p^{*B}(t)$ by the following stochastic differential equations with leverage effect, see e.g. Meddahi

(2002):

$$dp^{*k}(t) = \sigma_k(t)\left[\sqrt{1-\lambda_k^2}dW_k^{(1)}(t) + \lambda_k dW_k^{(2)}(t)\right], \qquad k = A, B, \tag{1.6}$$

where $W_A^{(1)}$ and $W_A^{(2)}$, $W_B^{(1)}$ and $W_B^{(2)}$, and $W_A^{(2)}$ and $W_B^{(2)}$ are pairwise independent Brownian motions. The volatilities follow GARCH diffusion processes: $d\sigma_k^2(t) = \kappa_k(\theta_k - \sigma_k^2(t))dt + \omega_k\sigma_k^2(t)dW_k^{(2)}(t)$, for $k = A, B$. The correlation is captured by $d\langle W_A^{(1)}, W_B^{(1)}\rangle_t = \rho(t)dt$. Given this setup, we have that the integrated covariance is given by

$$\mathbf{IC} = \int_0^1 \begin{pmatrix} \sigma_A^2(t) & \bullet \\ \sigma_A(t)\sigma_B(t)\sqrt{(1-\lambda_A^2)(1-\lambda_B^2)}\rho(t) & \sigma_B^2(t) \end{pmatrix} dt$$

Following BNS (2004) the correlation is represented by the anti-Fisher transformation, $\rho(t) = \frac{e^{2x(t)}-1}{e^{2x(t)}+1}$, where $x(t)$ follows the GARCH diffusion $dx(t) = \kappa(\theta - x(t))dt + \omega x(t)dW(t)$ and $W(t)$ is independent from all other Brownian motions.

The next component of the simulation is to generate trading times. Let the time between consecutive observations be denoted as $\mathcal{T}_k$. First, we generate the intertrade (interquote) durations from an exponential distribution with means, e.g. 5 seconds and 10 seconds, respectively. A more realistic sampling scheme would be one with stochastic time-varying intensity, which depends on the level of volatility, fluctuating around a base intensity φ_k^{base}. Then, we can write the density of the durations as $f_{\mathcal{T}_k}(\tau_k) = \varphi_k(t)\exp(-\varphi_k(t)\tau_k)$, with an intensity coefficient given by

$$\varphi_k(t) = \exp\left[\ln\left(\varphi_k^{base}\right) + c_k\left(\sigma_k^2(t) - \overline{\sigma_k^2}\right)\right], \qquad k = A, B,$$

where $\overline{\sigma_k^2}$ is the pathwise average of σ_k^2 and $c_k \geq 0$ is a constant which controls the sensitivity of the intensity to the deviation of σ_k^2 from its mean. Having sampled the true price process at the observation times, we add a noise component, which is modelled either as a Gaussian white noise or as a bivariate ARMA process.

1.5.3 Simulation Results

We now evaluate the relative performance of the following estimators described in Sections1.3 and1.4:

- realized covariance: $RC^{(M)}$, with $M = 32400, 390, 78$;
- realized covariance corrected by one lead and lag: $RC_{AC_1}^{(M)}$, with $M = 32400, 390, 78$;

- Bandi & Russell (2005*b*) optimally sampled realized covariance: $B\&R$;
- realized covariance based on the "Replace All" matching scheme: RC^{ra}; and
- the cumulative covariance estimator and its extensions: CC, CC^{bc}, CC^{sub}, and CC^{sub-bc}.

Due to space considerations, we only present part of the results here. Before considering the case with noise, we want to emphasize again that with non-synchronicity in the absence of noise the CC estimator is unbiased and the most efficient choice. All interpolation-based estimators will suffer from the Epps effect, while the bias-corrected CC will be less efficient because it corrects for noise, which is not present. Furthermore, without noise, the variance of the estimator cannot be reduced by subsampling.

To capture the most important properties found in the data, we present results with stochastic correlation and contemporaneously correlated multivariate normal noise with $\varrho(0) = -0.1$, $\omega_{AA} = 10^{-7}$ and $\omega_{BB} = 2.10^{-7}$, values close to those we obtained empirically. Although this specification is not rich enough to capture lagged cross-correlation effects, it is sufficient to reveal the main properties of the estimators regarding their bias and variance. Under this noise specification, all estimators with the exception of the subsampling ones are inconsistent. The $RC^{(M)}$ is biased in all trading scenarios, while $RC^{(M)}_{AC_1}$ is unbiased under synchronicity. The CC^{bc} is unbiased, while the CC^{sub-bc} is unbiased and consistent under correct choice of q. Table 1.2 contains the results.

In the table we report the bias, standard deviation and RMSE in percent from the true value of $0.49\%^2$ daily integrated covariance. For the regular trading schemes, the realized covariance estimators coincide with the RC^{ra} and CC estimators by construction. With $\omega_{AB} \neq 0$, we see the accumulation of noise in the bias for the realized covariance based estimators. Also, whenever the observation frequency is lower than the sampling frequency, the realized variance estimators coincide, since the higher frequency based ones simply add up zeros. For very large M (the first scenario), the variance of $RC^{(M)}_{AC_1}$ is smaller than the variance of $RC^{(M)}$ (this can be proven theoretically, based on the univariate results of Hansen & Lunde (2006)), but still too high compared to the regular RC based on one-minute returns, due to noise accumulation.

The joint effects of noise and non-synchronicity are captured in the last three trading scenarios. In all three cases, non-synchronicity, reinforced by negative noise correla-

Table 1.2: Bias, standard deviation and RMSE (in percent) of covariance estimators. Case of stochastic correlation and correlated noise.

	Synchronous sampling									Non-synchronous sampling								
	Reg-S (1s)			Reg-S (1m)			Reg-S (5m)			Reg-NS (5,10)			Poisson (1/5,1/10)			Poisson(nh) (1/5,1/10)		
	Bias	St. Dev	RMSE	Bias	St. Dev	RMSE	Bias	St. Dev	RMSE	Bias	St. Dev	RMSE	Bias	St. Dev	RMSE	Bias	St. Dev	RMSE
$RC^{(23400)}$	-1350	114.8	1355	-22.08	25.41	33.66	-9.88	35.91	37.25	-100.0	0.00	100.0	-108.2	15.25	109.3	-106.9	14.35	107.9
$RC^{(390)}$	-22.08	25.41	33.66	-22.08	25.41	33.66	-9.88	35.91	37.25	-1.61	26.17	26.22	-16.39	26.46	31.13	-16.55	26.43	31.19
$RC^{(78)}$	-9.88	35.91	37.25	-9.88	35.91	37.25	-9.88	35.91	37.25	-7.36	35.53	36.28	-9.97	35.82	37.18	-12.92	36.25	38.48
$RC_{AC_1}^{(23400)}$	-4.96	93.26	93.39	-22.08	25.41	33.66	-9.88	35.91	37.25	-100.0	0.00	100.0	-93.88	23.21	96.70	-92.04	22.43	94.74
$RC_{AC_1}^{(390)}$	-11.08	33.25	35.05	-11.08	33.25	35.05	-9.88	35.91	37.25	-9.53	32.54	33.90	-10.20	31.64	33.24	-11.92	32.97	35.06
$RC_{AC_1}^{(78)}$	-5.91	54.12	54.44	-5.91	54.12	54.44	-5.91	54.12	54.44	-8.08	55.08	55.67	-7.73	54.73	55.27	-9.46	55.08	55.88
$B\&R$	-12.41	41.29	43.12	-13.17	35.94	38.28	-10.59	35.60	37.14	-6.78	35.51	36.15	-11.58	33.37	35.32	-9.42	32.34	33.68
RC^{ra}	-1350	114.8	1355	-22.08	25.41	33.66	-9.88	35.91	37.25	-28.52	37.83	47.37	-52.42	37.89	64.68	-50.89	37.50	63.21
CC	-1350	114.8	1355	-22.08	25.41	33.66	-9.88	35.91	37.25	1.39	33.89	33.92	-27.98	32.84	43.14	-26.98	31.33	41.35
CC^{bc}	-5.02	93.26	93.39	-11.04	33.24	35.03	-6.51	53.75	54.14	1.39	33.89	33.92	-0.41	31.40	31.40	-1.48	30.91	30.94
CC^{sub}										0.15	15.37	15.37	-7.12	17.21	18.63	-7.15	15.80	17.34
CC^{sub-bc}										0.15	15.37	15.37	-1.85	17.32	17.42	-2.02	16.35	16.47

Simulation model: $\mathbf{p} = \mathbf{p}^* + \mathbf{u}$, where $\mathbf{p}^*$ is given in (1.6) with $\lambda_A = \lambda_B = 0.5$, σ_k^2 follow GARCH diffusion processes: $d\sigma_k^2(t) = \kappa_k(\theta_k - \sigma_k^2(t))dt + \omega_k\sigma_k^2(t)dW_k^{(2)}(t)$, for $k = A, B$ (stochastic volatility), and $\rho(t) = \frac{e^{2x(t)}-1}{e^{2x(t)}+1}$, where $x(t)$ again follows the GARCH diffusion $dx(t) = \kappa(\theta - x(t))dt + \omega x(t)dW(t)$ (stochastic correlation). The noise is iid normal with contemporaneous correlation: $\mathbf{u} \sim N\left(\mathbf{0}, \begin{pmatrix}\omega_{AA} & \omega_{AB} \\ \omega_{AB} & \omega_{BB}\end{pmatrix}\right)$. The true integrated covariance is $0.49\%^2$.

Note: "Reg-S" stands for regular synchronous observation times, "Reg-NS" – regular non-synchronous observation times (in brackets are given the durations between trades for both assets), "Poisson" – Poisson observation arrivals with intensities for both assets given by the entries in brackets, and "Poisson(nh)" – non-homogeneous Poisson arrivals with intensities given by base components (in brackets) plus volatility-driven components.

tion cause large biases in the RC-based estimators. Even with 5-minute sampling, the bias with Poisson arrivals is -10 to -13%. In the pure non-synchronous case (Reg-NS), the CC and CC^{bc} estimators coincide, since under this noise assumption the parameters b^+ and b^- are nonzero only when there are synchronous observations. Interestingly, in the last two trading scenarios CC^{bc} and CC have roughly the same standard deviation. Since the former is also unbiased, it is the clearly better in terms of RMSE.

In order to see the benefits of subsampling, we choose the optimal number of subgrids q as described in Section1.4, using known quantities to calculate the functions $f(\cdot)$ and $g(\cdot)$. Referring to the variance expression in Lemma 1.2, these functions are given by $f(\cdot) = \int_0^1 (\sigma_{AA}(u)\sigma_{BB}(u)+\sigma_{AB}^2(u))du$ and $g(\cdot) = 2(2\kappa+1)\omega_{AA}\omega_{BB}+2(4\kappa-1)\omega_{AB}^2$. In order to avoid additional noise accumulation through the estimation of κ, we set it equal to one, corresponding to the case of normal noise. For CC^{bc} we use the same optimal q as an approximation, since we do not know explicitly the variance of this estimator. In the last two rows of the table we have the bias, standard deviation and RMSE of the subsampled CC and CC^{bc} with optimal number of subgrids for the non-synchronous trading scenarios. The variance reduction is obvious, and the RMSE's have almost halved for the bias-corrected estimator. In practice, the functions f and g can be estimated, e.g. by means of kernel estimates for the integrated variances and a low-frequency RC estimator for the integrated covariance. The noise variances can be estimated as mentioned above by $\hat{\omega}^2$ and the noise covariance ω_{AB} is readily available as $\hat{\gamma}(0)$. In our simulations, this led to slightly higher optimal q's compared to the infeasible case, since the noise variance estimators are upward biased and the noise-to-signal ratio is overestimated.

In the second simulation experiment, we check the consistency of the subsampling version of the CC^{bc} estimator. Theoretically, we have shown how to choose the optimal number of subgrids in the i.i.d. noise scenario, and that this leads to consistency for unbiased estimators. In order to check if consistency is still obtained with dependent noise, we generate a bivariate ARMA noise process with the same variance and contemporaneous correlation as in the previous simulation, but allow for lead and lag dependence of about one minute. We then generate Poisson sampling times with decreasing intensity. As in the previous simulation, we compute the optimal number of subgrids for the CC^{bc} based on the variance expression under synchronicity and i.i.d. noise in Lemma 1.2. Figure 1.7 plots the standard deviation in percent of the CC^{bc} with and without subsampling against the sampling frequency.

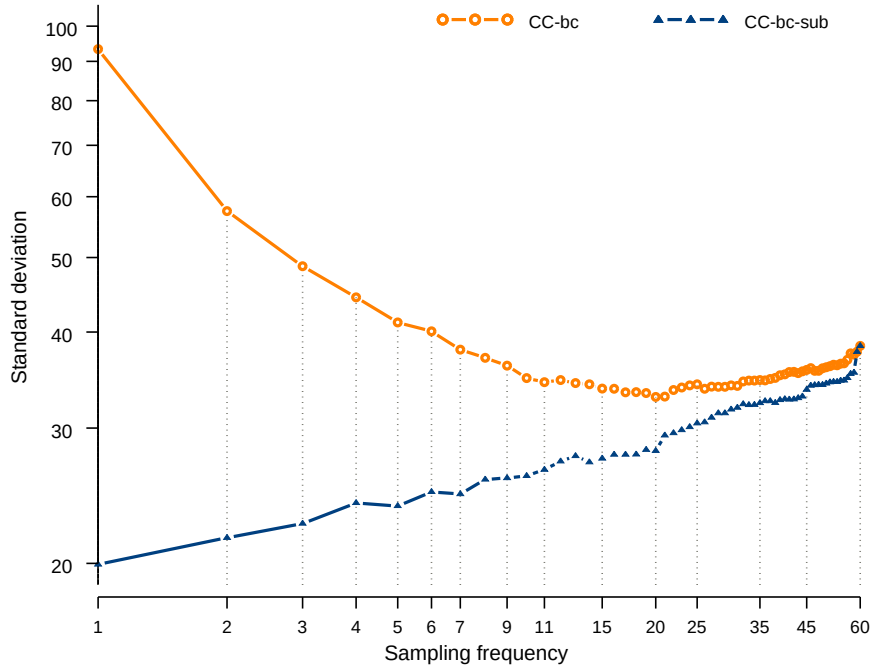

Figure 1.7: Simulation-based standard deviation (in percent) of the CC^{bc} and CC^{sub-bc} estimators across sampling frequency in ticks.

For the subsampling estimator we observe declining standard deviation across the whole range of sampling frequencies. Thus, even with dependent noise, it is still optimal to subsample using all available observations.

1.6 Conclusion

The paper studies the properties of several covariance estimators in the presence of noise and non-synchronicity. In general, the estimators can be separated in two groups: conventional realized covariance estimators based on synchronized observations and estimators based on non-synchronous data.

Non-synchronicity in the trading process of different assets is of fundamental importance whenever high-frequency data is used to estimate covariance. While theoretically realized covariances in the absence of noise are unbiased and consistent, this is no longer true when non-synchronicity is present. Although this effect was noted as early as 1979 by T. Epps, it was not until recently that reliable high-frequency financial data became available and this problem pressed for a solution.

The cumulative covariance estimator by Hayashi & Yoshida (2005) is one possible way of recovering the true integrated covariance when prices are observed without noise. The main focus of the paper was to cast this estimator into the more realistic

world of noisy prices and to examine how it behaves in this new setting. The theoretical results show that the estimator is quite robust to noise and remains unbiased even under independent noise. In a general noise setting, however, it is biased and inconsistent. In this case we propose a bias correction, given that the noise has finite dependence.

In order to apply the bias correction we need an estimate of the cross-covariance function of the noise process. We suggest an approach to estimate this function in calendar time using directly the non-synchronous observations without interpolation. Based on the cross-covariance function, we show how bias-correction can be applied in practice. As the resulting bias-corrected estimator is inconsistent, we show how to choose an optimal number of subgrids and apply subsampling to achieve consistency in the i.i.d. noise case. Under dependent noise, simulations show that subsampling still preserves its merit.

It remains an open question how these high-frequency based covariance and variance estimators can be combined in a variance-covariance matrix which can be used e.g. in portfolio management. Usually this has been achieved by taking realized covariance matrices and possibly bias-correcting them for noise-induced bias. This approach, however, totally neglects the non-synchronicity issue, to which only the diagonal elements of such a matrix are immune. A reasonable approach here would be to apply a procedure similar to the one used by Ledoit, Santa-Clara & Wolf (2003) in their Flex-GARCH methodology. Having constructed a sequence of high-frequency based covariance matrices, they could be modelled dynamically (as in, e.g., Gourieroux, Jasiak & Sufana (2004)) to produce forecasts of multivariate risk, which in turn can be used as input for portfolio and risk optimization problems.

Bibliography

Anderson, T. W. (2003), *An introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, New Jersey.

Bandi, F. M. & Russell, J. R. (2005*a*), Microstructure noise, realized volatility, and optimal sampling. Working paper, Graduate School of Business, The University of Chicago.

Bandi, F. M. & Russell, J. R. (2005*b*), Realized covariation, realized beta, and microstructure noise. Working paper, Graduate School of Business, The University of Chicago.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Estimating quadratic variation using realised variance', *Journal of Applied Econometrics* **17**, 457–477.

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

Barndorff-Nielsen, O., Hansen, P., Lunde, A. & Shephard, N. (2006), Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Working paper, Nuffield College, Oxford.

Chordia, T., Roll, R. & Subrahmanyam, A. (2005), 'Evidence on the speed of convergence to market efficiency', *Journal of Financial Economics* **76**, 271–292.

de Pooter, M., Martens, M. & van Dijk, D. (2006), Predicting the daily covariance matrix for S&P 100 stocks using intraday data - but which frequency to use?. Erasmus University Rotterdam.

Epps, T. (1979), 'Comovements in stock prices in the very short run', *Journal of the American Statistical Association* **74**, 291–298.

Gourieroux, C., Jasiak, J. & Sufana, R. (2004), The wishart autoregressive process of multivariate stochastic volatility. Working Paper, University of Toronto.

Griffin, J. E. & Oomen, R. C. A. (2006), Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Working Paper, University of Warwick.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–218.

Harris, F., McInish, T., Shoesmith, G. & Wood, R. (1995), 'Cointegration, error correction and price discovery on informationally-linked security markets', *Journal of Financial and Quantitative Analysis* **30**, 563–581.

Hayashi, T. & Kusuoka, S. (2004), Nonsynchronous covariation measurement for continuous semimartingales. Preprint Series of the Graduate School of Mathematical Sciences, The University of Tokyo.

Hayashi, T. & Yoshida, N. (2004), 'On covariance estimation for high-frequency financial data', *Proceeding of IASTED/Financial Engineering and Applications 2004* **437-801**, 282–286.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), 'Flexible multivariate GARCH modeling with an application to international stock markets', *Review of Economics and Statistics* **85**, 735–474.

Martens, M. (2004), Estimating unbiased and precise realized covariances. Econometric Institute, Erasmus University Rotterdam.

Meddahi, N. (2002), 'A theoretical comparison between integrated and realized volatility', *Journal of Applied Econometrics* **17**, 479–508.

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

Renò, R. (2001), A closer look at the Epps effect. Università degli Studi di Siena, Working paper n. 335.

Roll, R. (1984), 'A simple implicit measure of the effective bid-ask spread in an efficient market', *Journal of Finance* **39**(4), 1127–1140.

Sheppard, K. (2005), Realized covariance and scrambling. Working paper, University of Oxford.

Zhang, L. (2006), Estimating covariation: Epps effect and microstructure noise. Working Paper.

Zhang, L., Mykland, P. A. & 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.

Appendix: Proofs

Proof of Theorem 1.1. First, we decompose the realized covariance into four components:

$$RC^{(M)} = \sum_{j=1}^{M} \mathbf{r}_j^* \mathbf{r}_j^{*\prime} + \sum_{j=1}^{M} \mathbf{r}_j^* \mathbf{e}_j' + \sum_{j=1}^{M} \mathbf{e}_j \mathbf{r}_j^{*\prime} + \sum_{j=1}^{M} \mathbf{e}_j \mathbf{e}_j'. \tag{A.1.1}$$

For the first term, we have that $\mathrm{E}\left[\sum_{j=1}^{M} \mathbf{r}_j^* \mathbf{r}_j^{*\prime}\right] = \mathbf{IC}$, the second term is Υ_M and the third is Υ_M'. For the fourth term,

$$\begin{aligned} \mathrm{E}\left[\sum_{j=1}^{M} \mathbf{e}_j \mathbf{e}_j'\right] &= M\,\mathrm{E}\left[\mathbf{e}_j \mathbf{e}_j'\right] \\ &= M\,\mathrm{E}\left[(\mathbf{u}_j - \mathbf{u}_{j-1})\left(\mathbf{u}_j' - \mathbf{u}_{j-1}'\right)\right] = 2M(\Gamma(0) - \Gamma(\delta) - \Gamma'(\delta)). \end{aligned}$$

Hence the result follows. □

Proof of Theorem 1.2. The estimator can be decomposed into

$$CC_{(A,B)} = \underbrace{\sum_{j_A=1}^{M_A} r_{j_A}^* r_{\overline{j}_B}^*}_{A} + \underbrace{\sum_{j_A=1}^{M_A} e_{j_A} r_{\overline{j}_B}^*}_{B} + \underbrace{\sum_{j_A=1}^{M_A} r_{j_A}^* e_{\overline{j}_B}}_{C} + \underbrace{\sum_{j_A=1}^{M_A} e_{j_A} e_{\overline{j}_B}}_{D}. \tag{A.1.2}$$

It is important to note that:

$$r_{\overline{j}_B}^* = \left(p_{\overline{j}_B}^{*B} - p_{j_A}^{*B}\right) + \underbrace{\left(p_{j_A}^{*B} - p_{j_A-1}^{*B}\right)}_{\text{true return of asset B over } \left(t_{j_A-1}, t_{j_A}\right]} + \left(p_{j_A-1}^{*B} - p_{\underline{j}_B}^{*B}\right). \tag{A.1.3}$$

We only need to show that $\mathrm{E}[A] = IC_{(A,B)}$, which follows from the decomposition in Equation (A.1.3) and the martingale properties of $\mathbf{p}^*$, since products of non-overlapping returns have zero expectation. The result then follows by taking expectations of (A.1.2). □

Proof of Lemma 1.1. In the hypothesized trading scenario we have that

$$CC_{(A,B)} = \sum_{j_A=1}^{M_A} r_{j_A} \left(\sum_{j_B=2(j_A-1)+1}^{2(j_A-1)+3} r_{j_B} \right)$$

Denote as in the general case but with the explicit dependence of $\overline{j}_B$ on j_A:

$$\sum_{j_B=2(j_A-1)+1}^{2(j_A-1)+3} r_{j_B} = r_{\overline{j}_B(j_A)},$$

which can be decomposed with respect to the overlap with r_{j_A}. Then we have $r_{\overline{j}_B(j_A)} = r_{\overline{j}_B(j_A)}^{(b)} + r_{\overline{j}_B(j_A)}^{(o)} + r_{\overline{j}_B(j_A)}^{(a)}$, where the bracketed superscripts have the following interpretation

b - before r_{j_A}; o - overlapping with r_{j_A}; a - after r_{j_A}.

We have that

$$\begin{aligned} CC_{(A,B)} &= \sum_{j_A=1}^{M_A} r_{j_A} r_{\bar{j}_B(j_A)} = \sum_{j_A=1}^{M_A} \left(r^*_{j_A} + e_{j_A}\right)\left(r^*_{\bar{j}_B(j_A)} + e_{\bar{j}_B(j_A)}\right) \\ &= \sum_{j_A=1}^{M_A} \underbrace{r^*_{j_A} r^*_{\bar{j}_B(j_A)}}_{Y_{j_A}} + \sum_{j_A=1}^{M_A} \underbrace{e_{j_A} r^*_{\bar{j}_B(j_A)}}_{V_{j_A}} + \sum_{j_A=1}^{M_A} \underbrace{r^*_{j_A} e_{\bar{j}_B(j_A)}}_{W_{j_A}} + \sum_{j_A=1}^{M_A} \underbrace{e_{j_A} e_{\bar{j}_B(j_A)}}_{U_{j_A}}. \end{aligned}$$

As none of the sums are correlated with each other (The correlation between $\sum_{j_A=1}^{M_A} V_{j_A}$ and $\sum_{j_A=1}^{M_A} W_{j_A}$ is given in 5. All other combinations have a correlation of zero due to independence of the price and noise processes.), we have that

$$\mathrm{V}\left[CC_{(A,B)}\right] = \mathrm{V}\left[\sum_{j_A=1}^{M_A} Y_{j_A}\right] + \mathrm{V}\left[\sum_{j_A=1}^{M_A} V_{j_A}\right] + \mathrm{V}\left[\sum_{j_A=1}^{M_A} W_{j_A}\right] + \mathrm{V}\left[\sum_{j_A=1}^{M_A} U_{j_A}\right].$$

1. $Y_{j_A} = r^*_{j_A} r^*_{\bar{j}_B(j_A)}$ and we use the equality $\mathrm{V}\left[Y_{j_A}\right] = \mathrm{E}\left[Y^2_{j_A}\right] - \mathrm{E}\left[Y_{j_A}\right]^2$.

$$\begin{aligned} \mathrm{E}\left[Y^2_{j_A}\right] &= \mathrm{E}\left[r^{*2}_{j_A} r^{*2}_{\bar{j}_B(j_A)}\right] = \mathrm{E}\left[r^{*2}_{j_A}\left(r^{*(b)}_{\bar{j}_B(j_A)} + r^{*(o)}_{\bar{j}_B(j_A)} + r^{*(a)}_{\bar{j}_B(j_A)}\right)^2\right] \\ &= \mathrm{E}\left[r^{*2}_{j_A} r^{*(b)^2}_{\bar{j}_B(j_A)}\right] + \mathrm{E}\left[r^{*2}_{j_A} r^{*(o)^2}_{\bar{j}_B(j_A)}\right] + \mathrm{E}\left[r^{*2}_{j_A} r^{*(a)^2}_{\bar{j}_B(j_A)}\right] \\ &= \sigma_{j_A,(AA)}\sigma^{(b)}_{\bar{j}_B(j_A),(BB)} + \psi_{j_A,(AB,AB)} + \sigma^2_{j_A,(AB)} + \sigma_{j_A,(AA)}\sigma^{(a)}_{\bar{j}_B(j_A),(BB)} \\ &= \sigma_{j_A,(AA)}\sigma^{(b)}_{\bar{j}_B(j_A),(BB)} + \sigma_{j_A,(AA)}\sigma_{j_A,(BB)} + 2\sigma^2_{j_A,(AB)} + \sigma_{j_A,(AA)}\sigma^{(a)}_{\bar{j}_B(j_A),(BB)} \end{aligned}$$

$$\mathrm{E}\left[Y_{j_A}\right] = \mathrm{E}\left[r^*_{j_A}\left(r^{*(b)}_{\bar{j}_B(j_A)} + r^{*(o)}_{\bar{j}_B(j_A)} + r^{*(a)}_{\bar{j}_B(j_A)}\right)\right] = \sigma_{j_A,(AB)}$$

$$\begin{aligned} \Rightarrow \mathrm{V}\left[Y_{j_A}\right] &= \sigma_{j_A,(AA)}\sigma^{(b)}_{\bar{j}_B(j_A),(BB)} + \sigma_{j_A,(AA)}\sigma_{j_A,(BB)} + \sigma^2_{j_A,(AB)} + \sigma_{j_A,(AA)}\sigma^{(a)}_{\bar{j}_B(j_A),(BB)} \\ &= \sigma_{j_A,(AA)}\left(\sigma^{(b)}_{\bar{j}_B(j_A),(BB)} + \sigma_{j_A,(BB)} + \sigma^{(a)}_{\bar{j}_B(j_A),(BB)}\right) + \sigma^2_{j_A,(AB)} \\ &= \sigma_{j_A,(AA)}\sigma_{\bar{j}_B(j_A),(BB)} + \sigma^2_{j_A,(AB)}. \end{aligned}$$

Since the elements in the sum are first-order autocorrelated we have

$$\begin{aligned} \mathrm{Cov}\left[Y_{j_A}, Y_{j_A+1}\right] &= \mathrm{E}\left[r^*_{j_A} r^*_{\bar{j}_B(j_A)} r^*_{j_A+1} r^*_{\bar{j}_B(j_A+1)}\right] - \mathrm{E}\left[r^*_{j_A} r^*_{\bar{j}_B(j_A)}\right]\mathrm{E}\left[r^*_{j_A+1} r^*_{\bar{j}_B(j_A+1)}\right] \\ &= \sigma_{j_A,(AB)}\sigma_{j_A+1,(AB)} + \sigma^{(a)}_{\bar{j}_B(j_A),(AB)}\sigma^{(b)}_{\bar{j}_B(j_A+1),(AB)} - \sigma_{j_A,(AB)}\sigma_{j_A+1,(AB)} \\ &= \sigma^{(a)}_{\bar{j}_B(j_A),(AB)}\sigma^{(b)}_{\bar{j}_B(j_A+1),(AB)} = \sigma_{t_{j_A}:t_{\bar{j}_B(j_A)},(AB)}\sigma_{t_{\bar{j}_B(j_A+1)}:t_{j_A},(AB)}. \end{aligned}$$

Since no higher order autocorrelation is present we obtain for the variance of the sum

$$\begin{aligned}\mathrm{V}\left[\sum_{j_A=1}^{M_A} Y_{j_A}\right] &= \sum_{j_A=1}^{M_A}\left(\sigma_{j_A,(AA)}\sigma_{\overline{j}_B(j_A),(BB)} + \sigma^2_{j_A,(AB)}\right) \\ &\quad +2\sum_{j_A=1}^{M_A-1}\sigma_{t_{j_A}:t_{\overline{j}_B(j_A)},(AB)}\sigma_{t_{\underline{j}_B(j_A+1)}:t_{j_A},(AB)}.\end{aligned}$$

2. $V_{j_A} = e_{j_A} r^*_{\overline{j}_B(j_A)}$ and hence

$$\begin{aligned}\mathrm{V}\left[V_{j_A}\right] &= \mathrm{V}\left[r^*_{\overline{j}_B(j_A)}\right]\mathrm{V}\left[u_{j_A} - u_{j_A-1}\right] = 2\omega_{AA}\sigma_{\overline{j}_B(j_A),(BB)} \quad \text{and} \\ \mathrm{Cov}\left[V_{j_A}, V_{j_A+1}\right] &= \mathrm{E}\left[r^*_{\overline{j}_B(j_A)}e_{j_A}\cdot r^*_{\overline{j}_B(j_A+1)}e_{j_A+1}\right] = \mathrm{E}\left[r^*_{\overline{j}_B(j_A)}r^*_{\overline{j}_B(j_A+1)}\right]\mathrm{E}\left[e_{j_A}e_{j_A+1}\right] \\ &= -\omega_{AA}\sigma_{\overline{j}_B(j_A)-1:\overline{j}_B(j_A),(BB)}.\end{aligned}$$

where the first equality follows from the independence of the price and error processes and the second equality we obtain from

$$\begin{aligned}\mathrm{E}\left[r^*_{\overline{j}_B(j_A)}r^*_{\overline{j}_B(j_A+1)}\right] &= \sigma_{\overline{j}_B(j_A)-1:\overline{j}_B(j_A),(BB)} \\ \mathrm{E}\left[e_{j_A}e_{j_A+1}\right] &= \mathrm{E}\left[\left(u_{j_A} - u_{j_A-1}\right)\left(u_{j_A+1} - u_{j_A}\right)\right] = -\omega_{AA}\end{aligned}$$

Since higher order autocovariances are zero, it follows that

$$\mathrm{V}\left[\sum_{j_A=1}^{M_A} V_{j_A}\right] = 2\omega_{AA}\sum_{j_A=1}^{M_A}\sigma_{\overline{j}_B(j_A),(BB)} - 2\omega_{AA}\sum_{j_A=1}^{M_A}\sigma_{\overline{j}_B(j_A)-1:\overline{j}_B(j_A),(BB)} = 2\omega_{AA}\sum_{j_B=1}^{M_B}\sigma_{j_B,(BB)}.$$

3. $W_{j_A} = r^*_{j_A} e_{\overline{j}_B(j_A)}$ and hence

$$\mathrm{V}\left[W_{j_A}\right] = \mathrm{V}\left[r^*_{j_A}\right]\mathrm{V}\left[e_{\overline{j}_B(j_A)}\right] = 2\omega_{BB}\sigma_{j_A,(AA)}.$$

Furthermore we have that

$$\mathrm{Cov}\left[W_{j_A}W_{j_A+h}\right] = \mathrm{E}\left[r^*_{j_A}r^*_{j_A+h}e_{\overline{j}_B(j_A)}e_{\overline{j}_B(j_A+h)}\right] = 0, \quad \forall h \neq 0.$$

It follows that

$$\mathrm{V}\left[\sum_{j_A=1}^{M_A} W_{j_A}\right] = 2\omega_{BB}\sum_{j_A=1}^{M_A}\sigma_{j_A(AA)}.$$

4. $U_{j_A} = e_{j_A}e_{\overline{j}_B(j_A)}$.

$$\mathrm{V}\left[U_{j_A}\right] = \mathrm{V}\left[e_{j_A}e_{\overline{j}_B(j_A)}\right] = \mathrm{V}\left[e_{j_A}\right]\mathrm{V}\left[e_{\overline{j}_B(j_A)}\right] = 4\omega_{AA}\omega_{BB}.$$

Moreover,

$$\mathrm{Cov}\left[U_{j_A}, U_{j_A+h}\right] = \mathrm{E}\left[e_{j_A} e_{j_A+h} e_{\overline{j}_B(j_A)} e_{\overline{j}_B(j_A+h)}\right] = 0, \quad \forall h \neq 0.$$

Therefore

$$\mathrm{V}\left[\sum_{j_A=1}^{M_A} U_{j_A}\right] = 4M_A \omega_{AA} \omega_{BB}.$$

(5. Proof that there is no covariance)

$$\mathrm{Cov}\left[\sum_{j_A=1}^{M_A} V_{j_A}, \sum_{j_A=1}^{M_A} W_{j_A}\right] = \mathrm{E}\left[r^*_{\overline{j}_B(j_A)} e_{j_A} r^*_{j_A+h} e_{\overline{j}_B(j_A+h)}\right]$$

$$= 0, \text{ since } \mathrm{E}\left[e_{j_A} e_{\overline{j}_B(j_A+h)}\right] = 0 \quad \text{and } p \perp\!\!\!\perp u, \quad \forall h.$$

Summing up all the terms we arrive at

$$\begin{aligned}\mathrm{V}\left[CC_{(A,B)}\right] = &\sum_{j_A=1}^{M_A} \left(\sigma_{j_A,(AA)} \sigma_{\overline{j}_B(j_A),(BB)} + \sigma^2_{j_A,(AB)}\right) \\ &+ 2\omega_{AA} \sum_{j_B=1}^{M_B} \sigma_{j_B,(BB)} + 2\omega_{BB} \sum_{j_A=1}^{M_A} \sigma_{j_A,(AA)} \\ &+ 4M_A \omega_{AA} \omega_{BB} + 2 \sum_{j_A=1}^{M_A-1} \sigma_{t_{j_A}:t_{\overline{j}_B(j_A)},(AB)} \sigma_{t_{\underline{j}_B(j_A+1)}:t_{j_A},(AB)}.\end{aligned}$$

□

Proof of Lemma 1.2. Under the assumption that asset A traders every second time asset B does we have that

$$CC_{(A,B)} = \sum_{j_A=1}^{M_A} r_{j_A} \left(\sum_{j_B=2(j_A-1)+1}^{2(j_A-1)+2} r_{j_B} \right) = \sum_{j_A=1}^{M_A} r_{j_A} r_{\overline{j}_B(j_A)},$$

where now $r_{\overline{j}_B(j_A)}$ includes only 2 tick returns of asset B (not 3 like before). Then we use the same decomposition as in the previous Lemma to derive the variance of the estimator. 1. $Y_{j_A} = r^*_{j_A} r^*_{\overline{j}_B(j_A)}$ and since in this scenario returns do not overlap (hence summands are not autocorrelated) we have for the variance of the sum

$$\mathrm{V}\left[\sum_{j_A=1}^{M_A} Y_{j_A}\right] = \sum_{j_A=1}^{M_A} \mathrm{V}\left[Y_{j_A}\right] = \sum_{j_A=1}^{M_A} \psi_{j_A} = \sum_{j_A=1}^{M_A} \left(\sigma_{j_A,(AA)} \sigma_{j_A,(BB)} + \sigma^2_{j_A,(AB)}\right).$$

2. $V_{j_A} = e_{j_A} r^*_{\overline{j}_B(j_A)}$ and due to the independence of the increments of the price process,

the variance of the sum is equal to the sum of the variances

$$\mathrm{V}\left[\sum_{j_A=1}^{M_A} V_{j_A}\right] = \sum_{j_A=1}^{M_A} \mathrm{V}\left[V_{j_A}\right] = \sum_{j_A=1}^{M_A} \mathrm{V}\left[e_{j_A}\right] \mathrm{V}\left[r^*_{\overline{j}_B(j_A)}\right] = 2\omega_{AA} \sum_{j_A=1}^{M_A} \sigma_{\overline{j}_B(j_A),(BB)}.$$

3. $W_{j_A} = e_{\overline{j}_B(j_A)} r^*_{j_A}$ and again we have that

$$\mathrm{V}\left[\sum_{j_A=1}^{A_M} W_{j_A}\right] = \sum_{j_A=1}^{M_A} \mathrm{V}\left[W_{j_A}\right] = \sum_{j_A=1}^{M_A} \mathrm{V}\left[e_{\overline{j}_B(j_A)}\right] \mathrm{V}\left[r^*_{j_A}\right] = 2\omega_{BB} \sum_{j_A=1}^{M_A} \sigma_{j_A,(AA)}.$$

4. $U_{j_A} = e_{j_A} e_{\overline{j}_B(j_A)}$ and hence

$$\mathrm{V}\left[U_{j_A}\right] = \mathrm{V}\left[e_{j_A} e_{\overline{j}_B(j_A)}\right] = \mathrm{E}\left[e^2_{j_A} e^2_{\overline{j}_B(j_A)}\right] - \mathrm{E}\left[e_{j_A} e_{\overline{j}_B(j_A)}\right]^2.$$

Since $e_{j_A} = u_{j_A} - u_{j_A-1}$ and $e_{\overline{j}_B(j_A)} = u_{\overline{j}_B(j_A)} - u_{\underline{j}_B(j_A)}$,

$$\begin{aligned}&\mathrm{E}\left[(u^2_{j_A} - 2u_{j_A}u_{j_A-1} + u^2_{j_A-1})(u^2_{\overline{j}_B(j_A)} - 2u_{\overline{j}_B(j_A)}u_{\underline{j}_B(j_A)} + u^2_{\underline{j}_B(j_A)})\right]\\ &= 2\mu_{4(AA,BB)} + 2\omega_{AA}\omega_{BB} + 4\omega^2_{AB}.\end{aligned}$$

This result follows from Lemma 1 and due to the synchronous trading ($t_{\overline{j}_B(j_A)} = t_{j_A}$ and $t_{\underline{j}_B(j_A)} = t_{j_A-1}$). Next, we have that

$$\mathrm{E}\left[(u_{j_A} - u_{j_A-1})(u_{\overline{j}_B(j_A)} - u_{\underline{j}_B(j_A)})\right] = 2\omega_{AB},$$

and for the variance of U_{j_A} we obtain

$$\mathrm{V}\left[U_{j_A}\right] = 2\mu_{4(AA,BB)} + 2\omega_{AA}\omega_{BB}.$$

Due to the synchronous trading U_{j_A} is first order autocorrelated. Using again a result from Lemma 1 we obtain for the autocovariance

$$\begin{aligned}\mathrm{Cov}\left[e_{j_A} e_{\overline{j}_B(j_A)}, e_{j_A+1} e_{\overline{j}_B(j_A+1)}\right] &= \underbrace{\mathrm{E}\left[e_{j_A} e_{\overline{j}_B(j_A)} e_{j_A+1} e_{\overline{j}_B(j_A+1)}\right]}_{\mu_{4(AA,BB)}+3\omega^2_{AB}}\\ &\quad - \underbrace{\mathrm{E}\left[e_{j_A} e_{\overline{j}_B(j_A)}\right] \mathrm{E}\left[e_{j_A+1} e_{\overline{j}_B(j_A+1)}\right]}_{4\omega^2_{AB}} = \mu_{4(AA,BB)} - \omega^2_{AB}.\end{aligned}$$

Summing up:

$$\begin{aligned}\mathrm{V}\left[\sum_{j_A=1}^{M_A} U_{j_B}\right] &= 2M_A(\mu_{4(AA,BB)} + \omega_{AA}\omega_{BB}) + 2\left(M_A - 1\right)\left(\mu_{4(AA,BB)} - \omega^2_{AB}\right)\\ &= 4M_A\mu_{4(AA,BB)} + 2M_A(\omega_{AA}\omega_{BB} - \omega^2_{AB}) - 2(\mu_{4(AA,BB)} - \omega^2_{AB}).\end{aligned}$$

5. To complete the result we need to compute $2\text{Cov}\left[\sum_{j_A=1}^{M_A} V_{j_A}, \sum_{j_A=1}^{M_A} W_{j_A}\right]$.

$$\text{Cov}\left[V_{j_A}, W_{j_A}\right] = \text{E}\left[r^*_{\overline{j}_B(j_A)} e_{j_A} e_{\overline{j}_B(j_A)} r^*_{j_A}\right] = \text{E}\left[r^*_{\overline{j}_B(j_A)} r^*_{j_A}\right] \text{E}\left[e_{j_A} e_{\overline{j}_B(j_A)}\right] = 2\omega_{AB}\sigma_{j_B,(AB)},$$

and since

$$\text{Cov}\left[V_{j_A}, W_{j_A+h}\right] = 0 \qquad \forall h \neq 0,$$

it follows that

$$2\text{Cov}\left[\sum_{j_A=1}^{M_A} V_{j_A}, \sum_{j_A=1}^{M_A} W_{j_A}\right] = 2\omega_{AB} \sum_{j_A=1}^{M_A} \sigma_{j_A,(AB)}$$

Summing up the 5 terms:

$$\begin{aligned}
\text{V}\left[CC_{(A,B)}\right] = &\sum_{j_A=1}^{M_A} (\sigma_{j_A,(AA)}\sigma_{j_A,(BB)} + \sigma^2_{j_A,(AB)}) + 2\omega_{AA} \sum_{j_B=1}^{M_B} \sigma_{j_B,(BB)} \\
&+ 2\omega_{BB} \sum_{j_A=1}^{M_A} \sigma_{j_A,(AA)} + 4M_A\mu_{4(AA,BB)} + 2M_A(\omega_{AA}\omega_{BB} - \omega^2_{AB}) \\
&- 2(\mu_{4(AA,BB)} - \omega^2_{AB}) + 2\omega_{AB} \sum_{j_A=1}^{M_A} \sigma_{j_A,(AB)}.
\end{aligned}$$

Replacing $\mu_{4(AA,BB)} = \kappa\left(\omega_{AA}\omega_{BB} + 2\omega^2_{AB}\right)$ leads to the result. □

Proof of Theorem 1.3. The running index h walk over indices j_B and we select b^+ and b^- such that for each j_A, $|t_{j_A} - t_{\overline{j}_B+b^+}| > \theta_0$ and $|t_{j_A-1} - t_{\underline{j}_B-b^-}| > \theta_0$. The bias corrected CC estimator is given by

$$\begin{aligned}
CC^{bc}_{(A,B)} &= \sum_{j_A=1}^{M_A} r_{j_A}\left(p^B_{\overline{j}_B+b^+} - p^B_{\underline{j}_B-b^-}\right) \\
&= \underbrace{\sum_{j_A=1}^{M_A} r_{j_A} r_{\overline{j}_B}}_{CC_{(A,B)}} + \sum_{j_A=1}^{M_A} \sum_{h=0}^{b^--1} r_{j_A} r_{\underline{j}_B-h} + \sum_{j_A=1}^{M_A} \sum_{h=1}^{b^+} r_{j_A} r_{\overline{j}_B+h}. \qquad \text{(A.1.4)}
\end{aligned}$$

The last two double sums on the RHS of this equation can be decomposed into sum of four parts $A+B+C+D$, as shown below. We will now prove that (A.1.4) is an unbiased estimator of $IC_{(A,B)}$ by considering each term in turn. First, we note that the first term of the decomposition

$$\text{E}\left[A\right] = \text{E}\left[\sum_{j_A=1}^{M_A} \sum_{h=0}^{b^--1} r^*_{j_A} r^*_{\underline{j}_B-h} + \sum_{j_A=1}^{M_A} \sum_{h=1}^{b} {}^{+} r^*_{j_A} r^*_{\overline{j}_B+h}\right] = 0,$$

since there are no overlapping returns in the sums.
Given that b^+ and b^- are chosen in the way described above and due to Assumption 1.4 we have that

$$\begin{aligned}\mathrm{E}[B] &= \mathrm{E}\left[e_{j_A}(p^{*B}_{\overline{j}_B+b^+} - p^{*B}_{\underline{j}_B-b^-})\right] \\ &= \mathrm{E}\left[e_{j_A}(r^*_{\underline{j}_B-b^-+1} + \ldots + r^*_{\underline{j}_B} + r^*_{\overline{j}_B} + r^*_{\overline{j}_B+1} + \ldots + r^*_{\overline{j}_B+b^+})\right] = 0.\end{aligned}$$

Hence summing from $j_A = 1$ to M_A, keeping the product $e_{j_A} r^*_{\overline{j}_B}$ on the left hand side and putting everything else on the right we obtain

$$b_{M_A} = -\,\mathrm{E}\left[\sum_{j_A=1}^{M_A}\sum_{h=0}^{b^--1} e_{j_A} r^*_{\underline{j}_B-h} - \sum_{j_A=1}^{M_A}\sum_{h=1}^{b^+} e_{j_A} r^*_{\overline{j}_B+h}\right].$$

Similarly, applying the same rearrangement we obtain

$$\begin{aligned}\mathrm{E}[C] &= \mathrm{E}\left[r^*_{j_A}(u^B_{\overline{j}_B+b^+} - u^B_{\underline{j}_B-b^-})\right] \\ &= \mathrm{E}\left[r^*_{j_A}(e_{\underline{j}_B-b^-+1} + \ldots + e_{\underline{j}_B} + e_{\overline{j}_B} + e_{\overline{j}_B+1} + \ldots + e_{\overline{j}_B+b^+})\right] = 0\end{aligned}$$

and hence

$$c_{M_A} = -\,\mathrm{E}\left[\sum_{j_A=1}^{M_A}\sum_{h=0}^{b^--1} r^*_{j_A} e_{\underline{j}_B-h} - \sum_{j_A=1}^{M_A}\sum_{h=1}^{b^+} r^*_{j_A} e_{\overline{j}_B+h}\right].$$

Lastly, we have that for an arbitrary h

$$\begin{aligned}\mathrm{E}\left[e_{j_A} e_{\overline{j}_B+h}\right] &= \mathrm{E}\left[(u^A_{j_A} - u^A_{j_A-1})(u^B_{\overline{j}_B+h} - u^B_{\underline{j}_B+h-1})\right] \\ &= \gamma\left(t_{\overline{j}_B+h} - t_{j_A}\right) - \gamma\left(t_{\underline{j}_B+h-1} - t_{j_A}\right) \\ &\quad - \gamma\left(t_{\overline{j}_B+h} - t_{j_A-1}\right) + \gamma\left(t_{\underline{j}_B+h-1} - t_{j_A-1}\right).\end{aligned}$$

It follows that summing for $h = 1$ to b^+ many terms cancel out and we are left with

$$\begin{aligned}\sum_{h=1}^{b^+}\mathrm{E}\left[e_{j_A} e_{\overline{j}_B+h}\right] &= \underbrace{\left[\gamma\left(t_{\overline{j}_B+b^+} - t_{j_A}\right) - \gamma\left(t_{\overline{j}_B+b^+} - t_{j_A-1}\right)\right]}_{=0 \text{ by Assumption 1.4}} \\ &\quad - \left[\gamma\left(t_{\overline{j}_B} - t_{j_A}\right) - \gamma\left(t_{\overline{j}_B} - t_{j_A-1}\right)\right],\end{aligned}$$

and similarly for the other sum

$$\begin{aligned}\sum_{h=0}^{b^--1}\mathrm{E}\left[e_{j_A} e_{\underline{j}_B-h}\right] &= \underbrace{\left[\gamma\left(t_{\underline{j}_B-b^-} - t_{j_A-1}\right) - \gamma\left(t_{\underline{j}_B-b^-} - t_{j_A}\right)\right]}_{=0 \text{ by Assumption 1.4}} \\ &\quad - \left[\gamma\left(t_{\underline{j}_B} - t_{j_A-1}\right) - \gamma\left(t_{\underline{j}_B} - t_{j_A}\right)\right].\end{aligned}$$

Therefore

$$\mathrm{E}\left[D\right] = \sum_{j_A=1}^{M_A} \mathrm{E}\left[\sum_{h=0}^{b^- -1} e_{j_A} e_{\underline{j}_B - h} + \sum_{h=1}^{b^+} e_{j_A} e_{\overline{j}_B + h}\right] =$$
$$\sum_{j_A=1}^{M_A} -\gamma\left(t_{\underline{j}_B} - t_{j_A - 1}\right) + \gamma\left(t_{\underline{j}_B} - t_{j_A}\right) - \gamma\left(t_{\overline{j}_B} - t_{j_A}\right) + \gamma\left(t_{\overline{j}_B} - t_{j_A - 1}\right) = -d_{M_A}.$$

Combining the results above it follows that

$$\mathrm{E}\left[\sum_{j_A=1}^{M_A}\sum_{h=0}^{b^- -1} r_{j_A} r_{\underline{j}_B - h} + \sum_{j_A=1}^{M_A}\sum_{h=1}^{b^+} r_{j_A} r_{\overline{j}_B + h}\right] = -b_{M_A} - c_{M_A} - d_{M_A},$$

and hence using the result in Theorem 1.2 the estimator (A.1.4) is unbiased. □

Proof of Corollary 1.1. We approximate the variance of the sum of pairs of non-overlapping returns by the sum of the variances. The covariance terms are negligible compared to the variance terms as explained below. As a further approximation we set the variance of each return spanning approximately x seconds equal to the average daily integrated variance per x seconds. Furthermore, for $l = 0$ we approximate $\mu_{4(AA,BB)} = \omega_{AA}\omega_{BB} + 2\omega_{AB}^2$ as in the normal case, in order to avoid the need for estimation of higher noise moments. The variance of a product of non-overlapping returns $r_{j_A^s} r_{j_B^s}$ can be derived as follows

$$\mathrm{V}\left[r_{j_A^s} r_{j_B^s}\right] = \mathrm{V}\left[r^*_{j_A^s} r^*_{j_B^s}\right] + \mathrm{V}\left[r^*_{j_A^s} e_{j_B^s}\right] + \mathrm{V}\left[e_{j_A^s} r^*_{j_B^s}\right] + \mathrm{V}\left[e_{j_A^s} e_{j_B^s}\right],$$

where we have used Assumption 1.3 (i.i.d., exogenous noise). Using the approximation $\mathrm{V}\left[r^*_{j_k^s}\right] = \sigma_{x\ \mathrm{sec},(kk)} = \frac{\int_0^1 \sigma_{kk}(u)du}{23400/x}$, for $k = A, B$ and since $\omega_{AB} = 0$ and the returns share no overlap we arrive at

$$\mathrm{V}\left[r_{j_A^s} r_{j_B^s}\right] = \sigma_{x\ \mathrm{sec},(AA)}\sigma_{x\ \mathrm{sec},(BB)} + 2\omega_{BB}\sigma_{x\ \mathrm{sec},(AA)} + 2\omega_{AA}\sigma_{x\ \mathrm{sec},(BB)} + 4\omega_{AA}\omega_{BB}.$$

For $l \neq 0$ matching return pairs within one subsample pair are independent. For $l = 0$ for each pair satisfying $t_{j_A^s - 1} - t_{j_B^s} = 0$ there is a mirror pair satisfying $t_{j_B^s - 1} - t_{j_A^s} = 0$ and the covariance between them is $\omega_{AA}\omega_{BB}$ (Note that there are $n_{0,s}/2$ such pairs hence we add the covariance only once.). Hence the slightly different expression for this case. We ignore the covariances across subsamples since covariance is only (possibly) nonzero between pairs s_A, s_B and s'_A, s'_B for $s_A \neq s'_A$ and $s_B \neq s'_B$. Thus the number of covariance terms in the sum is at most $\min(S_A, S_B)$, which is negligible compared to the number of variance terms which is $n_{l,s} S_A S_B$. Therefore (e.g. for $l \neq 0$) we have that

$$\mathrm{V}\left[\gamma(l)_s\right] = \frac{1}{n_{l,s}}(\sigma_{x\ \mathrm{sec},(AA)}\sigma_{x\ \mathrm{sec},(BB)} + 2\omega_{BB}\sigma_{x\ \mathrm{sec},(AA)} + 2\omega_{AA}\sigma_{x\ \mathrm{sec},(BB)} + 4\omega_{AA}\omega_{BB}),$$

end the final expression follows readily.
It should be noted that if the integrated covariance is positive as it is usually the case with stocks, ignoring the covariance terms will lead to underestimating the variance of $\hat{\gamma}(l)$ leading to somewhat larger t-statistics. For the bias correction it is most important not to miss significant leads and lags. Therefore, the error due to the approximation is not too harmful. □

Chapter 2

Estimating High-Frequency Based (Co-) Variances: A Unified Approach

2.1 Introduction

This paper presents a novel approach for estimating both the variances and covariances of a multivariate diffusion processes in the presence of market microstructure noise within a unified theoretical framework. Recently, the literature on estimating the variance of irregularly observed high-frequency financial prices has experienced a substantial development with respect to relaxing the usual i.i.d. noise assumption and constructing consistent estimators for the variance of the underlying efficient price process. Leading examples in this field are the realized kernels of Barndorff-Nielsen et al. (2006) and the two-scales realized volatility of Zhang et al. (2005).
The progress in estimating the covariance between two randomly observed diffusion processes has been somewhat more cumbersome, due to the additional complication of non-synchronicity. Nevertheless, recent contributions such as Hayashi & Yoshida (2005) and Corsi & Audrino (2007) have introduced consistent and unbiased estimators for non-synchronously observed processes, when there are no market frictions. Based on these approaches, Griffin & Oomen (2006), Voev & Lunde (2007) and Zhang (2006*b*), among others, consider the properties of such estimators in the presence of measurement noise and propose certain extensions in order to correct for the impact of the noise.
There still does not exist, however, a unified methodology for estimating both the variance and covariance of high-frequency noisy prices, thus allowing for estimation of the whole variance-covariance matrix, which accounts for a wide range of possible noise specifications and non-synchronicity. Our work is intended to fill this gap by

providing precise and unbiased estimators, which are also easy to apply in practice. Furthermore, we obtain estimates for the dependence structure of the noise process, leading to a better understanding of market microstructure frictions on the transaction level. The power of our methodology lies in the ability to separate the variation of the efficient price from the variation of the noise process, which jointly contribute to the variation of the observed noisy price process. This identification is possible since the effect of the noise accumulates (up to a certain extent) by sampling more frequently, while the variation of the true process is constant. To illustrate our approach more intuitively, consider the so-called volatility signature plots introduced by Andersen, Bollerslev, Diebold & Labys (2001) as a graphical tool to study the effects of market microstructure noise on the properties of the realized variance estimator. These plots are an illustration of a particular relationship between the realized variance and the number of sampling points. For each sampling frequency, the expectation of the resulting realized volatility has a component which is constant across the range of sampling frequencies (the integrated variance of the underlying process), and a component which varies with the number of points on the grid (the noise variance which accumulates linearly with the number of returns on the grid). Since the number of sampling points is observable, the true integrated variance can be obtained as the constant of a projection of the realized volatility on the number of returns used to compute it. Based on this simple idea, our approaches lead to efficiency gains of 35% to 50% in a realistic trading scenario. While Curci & Corsi (2006) and Phillips & Yu (2006) use a similar idea for the estimation of the variance, their estimators are different from ours and practically more difficult to implement. Concerning the covariance case with non-synchronicity, our results are to our knowledge, completely new to the literature.
The paper is structured as follows: in Section 2.2 we introduce the notation and the theoretical framework we are working under, Section 2.3 presents our estimation methodology, Section 2.4 contains the results of a simulation study in which we compare our approach to other existing approaches, and Section 2.5 concludes. The Appendix contains detailed results of the simulation study.

2.2 Theoretical Setup

Our basic assumption is that we have irregularly spaced, non-synchronous observations of an n-dimensional continuous time process $\mathbf{p}(t)$, $t \geq 0$, which is a noisy signal for an underlying process $\mathbf{p}^*(t)$:

$$\mathbf{p}(t) = \mathbf{p}^*(t) + \mathbf{u}(t),$$

where $\mathbf{u}(t)$ is the noise term. The elements of $\mathbf{p}$, $\mathbf{p}^*$ and $\mathbf{u}$ are denoted by p^k, p^{*k} and u^k, for $k = 1, \ldots, n$, respectively. As in Barndorff-Nielsen & Shephard (2004), the process $\mathbf{p}^*(t)$ satisfies the following assumption:

Assumption 2.1. *The process $\mathbf{p}^*(t)$ is a multivariate martingale process with stochastic volatility satisfying*

$$\mathbf{p}^*(t) = \int_0^t \mathbf{\Theta}(u) d\mathbf{W}(u)$$

where $\mathbf{\Theta}$ is the spot covolatility process and $\mathbf{W}$ is a vector standard Brownian motion of dimension m. All the elements of $\mathbf{\Theta}(t)$ are cádlág.[1]

Defining the spot covariance as $\mathbf{\Sigma}(t) = \mathbf{\Theta}(t)\mathbf{\Theta}(t)'$, the integrated covariation process of $\mathbf{p}^*$ is given by

$$\mathbf{IC}(t) = \int_0^t \mathbf{\Sigma}(u) du.$$

The diagonal elements of $\mathbf{\Sigma}(t)$ are assumed to be integrable.
Our aim is to estimate the increment of integrated covariance

$$\mathbf{IC}(a, b) = \int_a^b \mathbf{\Sigma}(u) du = \mathbf{IC}(b) - \mathbf{IC}(a).$$

for some predetermined choice of (a, b), e.g., a trading day. Henceforth, we assume that the period of interest is a trading day with $a = 0$ and $b = 1$, and we will omit a and b in the notation.
With respect to the market microstructure noise process, we make the following assumption:

Assumption 2.2.

(*i*) $\mathbf{p}^*(s) \perp\!\!\!\perp \mathbf{u}(t)$, *for all s and t;*

(*ii*) $\mathrm{E}[\mathbf{u}(t)] = 0$ *for all t;*

Under this assumption the noise process can be serially correlated, but is exogenous to the true price process $\mathbf{p}^*$. In the univariate case a similar assumption can be found in Barndorff-Nielsen et al. (2006) and Aït-Sahalia, Mykland & Zhang (2006), among others, who relax the standard assumption of an i.i.d. noise process used in the early realized volatility literature.

[1] The acronym cádlág stands for "continue à droite, limite à gauche". This condition ensures that the integral with respect to $\mathbf{W}$ exists.

In the univariate case, the i.i.d. assumption can be relaxed by postulating serial dependence either in calendar (physical) or in tick time, the latter being more intuitive and also easier to work with. In the multivariate case, i.i.d. noise is still the working assumption used, by e.g. Griffin & Oomen (2006), Bandi & Russell (2005*b*). Recently, Voev & Lunde (2007) have shown that the i.i.d. assumption cannot be sustained empirically, which can be explained by staggered information assimilation in asset prices. While securities, which are more closely followed by analysts and more frequently traded, react faster to new information, slower trading assets take time to adjust to the news, causing lagged correlations across assets.
Voev & Lunde (2007) discuss certain problems of adopting the univariate tick-time dependence assumption in the multivariate case which render the direct implementation of tick-time dependence infeasible, and therefore assume serial cross-dependence in calendar time. In this paper we will follow this approach since it seems to be the most reasonable way to achieve an unified framework for modelling dependent noise processes in the multivariate framework with non-synchronicity. Thus, we complete Assumption 2.2 as follows

Assumption 2.2. *(continued)*

(*iii*) *The noise process* $\mathbf{u}$ *is covariance stationary with autocovariance function given by* $\mathbf{\Gamma}(q) = \mathrm{E}\left[\mathbf{u}(t)\mathbf{u}'(t-q)\right]$.

The (k,l)*-element of* $\mathbf{\Gamma}(q)$*,* $k,l = 1,\ldots,n$ *is denoted by* $\gamma_{k,l}(q)$.

Since it might be of concern that the given assumption does not take trading activity and diurnal seasonality into account, it is worth mentioning that there are other possibilities for defining dependence across assets, which to a certain extent address this issue, but bear some complications. One approach would be to use the pooled arrival process of any two assets as the time scale for defining the cross-correlations between them. This has the disadvantage that for a given asset k, its time-series properties will depend whether we consider it in combination with an asset l or an asset l'. Furthermore, with this approach, we would not be able to define the usual matrix autocorrelation function for multivariate processes, which should satisfy $\gamma_{k,l}(q) = \gamma_{l,k}(-q)$. A partial solution to this problem could be to consider the pooled process of the whole universe of assets under consideration. Such a structure, however, will still depend on the assets included in the universe, and will change from application to application. Apart from this, in the limit, when we include more and more assets, the pooled arrival process will converge to a regular time grid measured in the smallest available time unit (e.g., a second), and hence would be identical to our setup in which we implicitly assume that time is a discrete multiple of a fixed time unit.

While until now we have considered the noise process as a continuous time process, the the noise contamination is actually manifested when transactions or quote updates occur, for example the contamination through the bid-ask bounce effect can be considered to perturb the true price process only at those times when a buy or a sell transaction is carried out. Therefore, it is important to consider its properties at the event times and we will use the notation $u^k(t_j^k) = u^k_{t_j^k}$ and $p^k(t_j^k) = p^k_{t_j^k}$, where t_j^k, $j = 1, \ldots, N^k$ denotes the event (transaction, quotes, etc.) arrival times and N^k is the total number of events for asset $k = 1, \ldots, n$. Under Assumption 2.2, we have, e.g., that $\mathrm{E}\left[u^k_{t_j^k} u^l_{t_{j'}^l}\right] = \gamma_{k,l}(q)$, whenever $t_j^k - t_{j'}^l = q$.
The estimation approach presented in this paper is also applicable under the alternative assumption of dependence defined on the pooled arrival process discussed above, with slight modifications. In particular, if one is only interested in estimating variances, the assumption of tick-time dependence considerably simplifies the estimation.

2.3 Estimation Procedures

If the process p^* were observed directly, a simple and asymptotically error-free estimator for $\mathbf{IC}(a, b)$ is the so-called realized covariance which is the sum of the squares of the increments of the process p^* at the highest available frequency over the interval (a, b). The properties of this estimator under such ideal conditions are derived in Barndorff-Nielsen & Shephard (2004). Two main issues arise for this estimator when used in practice. Firstly, when the separate univariate processes are not observed simultaneously, one has to resort to synchronization techniques in order to define joint observation times for the multivariate process. Such techniques lead to biases in the estimated covariances, which are known as the Epps effect (Epps (1979)). Secondly, the presence of noise leads to biases and inconsistency. The properties of the last-tick interpolation based realized covariances are studied by Zhang (2006*b*), Griffin & Oomen (2006) and Martens (2004), among others. Based on the results of these studies, different approaches are proposed to make the realized covariance robust to market microstructure noise such as calculation of optimal sampling frequencies and lead-lag corrections. More recently, researchers have concentrated on developing sophisticated models which are specifically designed to estimate only the variance of a given asset (variance models) or a single covariance between two assets (covariance models). Concerning the variance models, recent advances include the two-scales realized variance by Zhang et al. (2005), the realized kernels of Barndorff-Nielsen et al. (2006), and the realized range-based variance which has newly been revived by Christensen & Podolskij (2007). With respect to covariance estimation Hayashi

& Yoshida (2005) and Corsi & Audrino (2007) propose an estimator which does not require synchronization of observations and thus accounts for the Epps effect. Griffin & Oomen (2006) study the properties of this estimator under i.i.d. noise, while Voev & Lunde (2007) propose extensions to the Hayashi-Yoshida estimator to make it robust to market microstructure frictions of a general nature.
In our methodology, the variances and covariances are estimated separately, but within the same theoretical framework. Advantages of our estimation procedure are its efficiency, straightforward implementation and robustness to misspecifications of the noise process.

2.3.1 Variance Estimation

We first focus on estimating the integrated variance of a single asset and then we turn to covariance estimation. To separate the variance of the unobservable price process from the variance of the noise component, we use the idea of the volatility signature plot introduced by Andersen, Bollerslev, Diebold & Labys (2001), which is the graphical representation of the realized variance against the sampling frequency at which it was computed. The volatility signature plot depicts the relationship between the realized variance computed with returns sampled on a certain grid and the number of sampling points on the grid for a set of predetermined grids. To gain an intuitive understanding for our estimation procedure, consider the i.i.d. noise case, under which theoretically the noise variance accumulates linearly with the number of sampling points, whereas the integrated variance is constant. Thus, an estimate of the integrated variance can simply be obtained as the intercept of the regression of the realized variances on the number of sampling points on the grid. Under a more general specification of the noise process, as in Assumption 2.2, the realized variances are further affected by the noise autocorrelations, which have to be taken into account in the regression by including appropriate additional regressors. More formally, consider a given asset k with N^k observations (ticks, transactions, quote updates) within the period of interest. To this end, the grid of observations $\{t_j^k\}_{j=1,\ldots,N^k}$ is subdivided into subgrids $\{t_{js+h}^k\}_{j=0,\ldots,\left\lfloor \frac{N^k-h}{s} \right\rfloor}$, where $s = 1, \ldots, S$ and $h = 1, \ldots, s$, which denotes the h-th subgrid for a sampling frequency of s ticks (e.g., with $s = 2$ we can have two subgrids, the first one comprising the ticks $\{t_1^k, t_3^k, t_5^k, \ldots\}$ and the second – the ticks $\{t_2^k, t_4^k, t_6^k, \ldots\}$). For each subgrid, we can define the

corresponding observed and efficient s-tick returns as

$$\begin{aligned} r^k_{t_{js+h}} &= p^k_{t_{(j-1)s+h}} - p^k_{t_{js+h}}, \quad j = 1, \ldots, \left\lfloor \frac{N^k - h}{s} \right\rfloor \\ r^{*k}_{t_{js+h}} &= p^{*k}_{t_{(j-1)s+h}} - p^{*k}_{t_{js+h}}, \quad j = 1, \ldots, \left\lfloor \frac{N^k - h}{s} \right\rfloor, \end{aligned}$$

and the noise returns as

$$e^k_{t_{js+h}} = u^k_{t_{(j-1)s+h}} - u^k_{t_{js+h}}, \quad j = 1, \ldots, \left\lfloor \frac{N^k - h}{s} \right\rfloor.$$

Denote the number of returns for the h-th s-subgrid as $N^k_{h,s} = \left\lfloor \frac{N^k - h}{s} \right\rfloor - 1$. The realized variance of asset k based on this subgrid is defined explicitly as a function of the number of returns on the subgrid:

$$RV^k(N^k_{h,s}) = \sum_{j=1}^{N^k_{h,s}} \left(r^k_{t_{js+h}} \right)^2.$$

To estimate the integrated variance we will exploit the following relationship, which holds under Assumptions 2.1 and 2.2:

$$\begin{aligned} \mathrm{E}\left[RV^k(N^k_{h,s})\right] &= \mathrm{IV}^k + \sum_{j=1}^{N^k_{h,s}} \mathrm{V}\left[e^k_{t_{js+h}}\right] \\ &= \mathrm{IV}^k + 2\sum_{q=1}^{\infty} N^k_{h,s}(q)\left(\gamma_{k,k}(0) - \gamma_{k,k}(q)\right) \\ &\approx \mathrm{IV}^k + 2N^k_{h,s}\gamma_{k,k}(0) - 2\sum_{q=1}^{Q} N^k_{h,s}(q)\gamma_{k,k}(q), \end{aligned} \tag{2.1}$$

where IV^k is the integrated variance of the true price process of asset k, i.e. element (k,k) of the matrix $\mathbf{IC}$ and $N^k_{h,s} = \sum_q N^k_{h,s}(q)$.[2] Thereby, $N^k_{h,s}(q)$ counts the number of q-second returns of asset k for the (h,s)-subgrid given by

$$N^k_{h,s}(q) = \sum_j \mathbb{1}_{\{t^k_{js+h} - t^k_{(j-1)s+h} = q\}}.$$

Note, that these counts need to be considered because we work with irregularly-spaced returns, which under the assumption of an autocovariance function defined on the smallest regular time grid (each second), imply that each $\mathrm{V}\left[e^k_{t_{js+h}}\right]$ depends

[2] A similar result appears in Hansen & Lunde (2006) with endogenous noise and regular sampling.

on the length of the return and thus consists of two elements, namely $\gamma_{k,k}(0)$ and the q-second autocovariance $\gamma_{k,k}(q)$. The approximation in Equation (2.1) results from truncating the autocorrelation function at lag Q. This is reasonable, since for a covariance stationary process the autocovariance function tends to zero for large lags, so that Q has to be chosen appropriately. As we will see below, letting Q be too large leads to more estimation noise, because for large Q's there are relatively few counts $N^k_{h,s}(Q)$. Furthermore, since $N^k_{h,s} = \sum_q N^k_{h,s}(q)$, choosing Q too large yields a singular regressor matrix.
Under the assumption of an i.i.d. noise process we obtain from Equation (2.1) the standard result (as in e.g., Hansen & Lunde (2006)):

$$\mathrm{E}\left[RV^k(N^k_{h,s})\right] \;=\; \mathrm{IV}^k + 2N^k_{h,s}\gamma_{k,k}(0).$$

Equation (2.1) differs to the extent that we have to consider the q-th order autocorrelation of the noise process and we have to count the number of occurrences.
On the basis of the theoretical relationship in Equation (2.1) and the above assumptions, we can easily derive the corresponding pooled OLS regression

$$y_{h,s} = c + \beta' x_{h,s} + \varepsilon_{h,s}, \quad s = 1, \ldots, S, \quad h = 1, \ldots, s \tag{2.2}$$

where $y_{h,s} = RV^k(N^k_{h,s})$ and $x_{h,s}$ is the Q-dimensional vector given by $x_{h,s} = (N^k_{h,s}, N^k_{h,s}(1), \ldots, N^k_{h,s}(Q))'$. In practice, Q has to be chosen appropriately, to reflect the degree of persistence of the noise process in the particular application. In the above regression, one simply regresses the realized variances on $N^k_{h,s}$ and the q-counts $N^k_{h,s}(q)$. The estimated constant $\hat{c}$ is an estimate of the integrated variance IV^k, while $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_Q$ are estimates of $2\gamma_{k,k}(0), -2\gamma_{k,k}(1), \ldots, -2\gamma_{k,k}(Q)$. Hence, as a byproduct of this estimation we can obtain the autocovariance function of the noise process, which can be identified under the assumption that the autocovariance $\gamma_{k,k}(Q)$ vanishes for a large enough Q. For a particular application, one could choose Q in an iterative manner starting from a relatively small value which is increased in each step. The optimal value of Q is the smallest value at which a given criterion (e.g. the gradient of the estimates) no longer changes considerably.

2.3.2 Covariance Estimation

Covariance estimation based on high-frequency data is inherently more challenging than variance estimation, since there is the additional complication of non-synchronicity. As mentioned already, non-synchronicity poses the problem of defining common event times for multiple assets. Typically, last-tick interpolation is employed, in which the last recorded price before a pre-defined observation time is

taken as the observed price at that point of time. This leads to a bias towards zero in the estimated realized covariance as the sampling frequency increases. A solution to this problem is proposed by Hayashi & Yoshida (2005). Considering two assets k and l, the Hayashi-Yoshida (HY) estimator based on all observations is defined as

$$HY^{k,l} = \sum_{j=1}^{N^k} \sum_{j'=1}^{N^l} r^k_{t^k_j} r^l_{t^l_{j'}} \mathbb{1}_{\{(t^k_{j-1}, t^k_j] \cap (t^l_{j'-1}, t^l_{j'}]\}}.$$

As can be seen from the definition, this estimator sums all cross products of overlapping returns of the assets under consideration. We can also base the estimation on the (h, s)- subgrid of asset k in combination with the (h', s')-subgrid of asset l, which we denote by

$$HY^{k,l}(h, s, h', s') = \sum_{j=1}^{N^k_{h,s}} \sum_{j'=1}^{N^l_{h',s'}} r^k_{t^k_{js+h}} r^l_{t^l_{j's'+h'}} \mathbb{1}_{\{(t^k_{(j-1)s+h}, t^k_{js+h}] \cap (t^l_{(j'-1)s'+h'}, t^l_{j's'+h'}]\}}. \quad (2.3)$$

In practice, it is convenient to implement this estimator by picking one of the assets, say k, and determining for each of its tick returns $r^k_{t^k_{js+h}}$, the corresponding return of the other asset which envelops it, i.e. starts before or at $t^k_{(j-1)s+h}$ and spans over at least to t^k_{js+h}. Of course, if one interchanges the assets, the estimator is numerically identical, but with respect to speed of execution, we recommend using the slower trading asset to determine the corresponding enveloping returns of the faster asset. In the following exposition we set the slower asset to be asset k. While the HY estimator is defined using all returns of both assets, effectively, there are at most $\min(N^k_{h,s}, N^l_{h',s'})$ different pairs of returns which contribute to the sum. This arises, because two or more neighboring returns of asset k may happen to be enveloped by the same return of asset l. Due to the summability of log returns, this effectively amounts to only one return pair in the sum of the HY estimator and the noise contaminations cancel against each other. Thus, the amount of noise which accumulates in the sum is a function of such effective pairs, while some of the ticks t^k_{js+h} play no role and are hence irrelevant. In order to determine the number of effective pairs, we introduce the right- and left-continuous counting functions $N^k_{h,s}(t) = \sum_{j=1}^{N^k_{h,s}} \mathbb{1}_{\{t^k_{js+h} \leq t\}}$ and $\breve{N}^k_{h,s}(t) = \sum_{j=1}^{N^k_{h,s}} \mathbb{1}_{\{t^k_{js+h} < t\}}$, $k = 1, \ldots, n$, $s = 1, \ldots, S$,

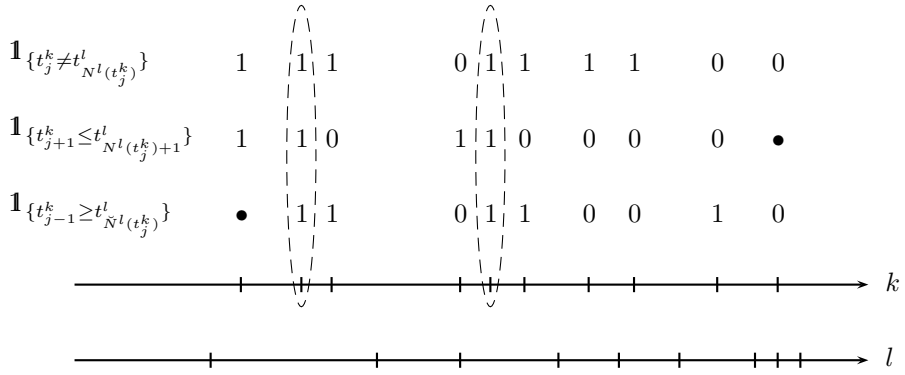

Figure 2.1: A graphical illustration for the identification of irrelevant ticks as described in equation (2.4). The indices h, s, h', s' have been suppressed.

and $h = 1, \ldots, s$. The number of irrelevant ticks t^k_{js+h} can be computed as follows

$$NI^k(h,s,h',s') = \sum_{j=1}^{N^k_{h,s}} \mathbb{1}_{\left\{t^k_{js+h} \neq t^l_{N^l_{h',s'}\left(t^k_{js+h}\right)}\right\}} \mathbb{1}_{\left\{t^k_{(j+1)s+h} \leq t^l_{N^l_{h',s'}\left(t^k_{js+h}\right)+1}\right\}} \mathbb{1}_{\left\{t^k_{(j-1)s+h} \geq t^l_{\check{N}^l_{h',s'}\left(t^k_{js+h}\right)}\right\}} \tag{2.4}$$

Figure 2.1 illustrates graphically how such irrelevant ticks are obtained. A tick t^k_{js+h} is irrelevant if it fulfills three conditions: i.) it is not synchronous with any tick of the other asset, ii.) the next arrival on the pooled process is generated by the same asset, and iii.) the previous arrival on the pooled process is generated by the same asset. If these conditions are satisfied simultaneously, then the impact of the noise at this tick cancels out in the summation of the HY estimator. The number of effective pairs is then given by

$$\tilde{N}^k(h,s,h',s') = N^k_{h,s} - NI^k(h,s,h',s').$$

The HY estimator can then be rewritten as

$$\begin{aligned} HY^{k,l}(h,s,h',s') &= \sum_{j=1}^{N^k_{h,s}} r^k_{t^k_{js+h}} r^l_{t^l_{N^l_{h',s'}\left(t^k_{(j-1)s+h}\right)}:t^l_{\check{N}^l_{h',s'}\left(t^k_{js+h}\right)+1}} \\ &= \sum_{j=1}^{\tilde{N}^k(h,s,h',s')} r^k_{\tilde{t}^k_{js+h}} r^l_{t^l_{N^l_{h',s'}\left(\tilde{t}^k_{(j-1)s+h}\right)}:t^l_{\check{N}^l_{h',s'}\left(\tilde{t}^k_{js+h}\right)+1}}, \end{aligned}$$

where $r^l_{t^l_{j'}:t^l_{i'}}$ denotes the (possibly multiple-tick) return of asset l over the interval $(t^l_{j'}, t^l_{i'})$, and the $\tilde{t}^k_{js+h}$'s denote the relevant ticks of asset k on the (h,s)-subgrid, i.e.,

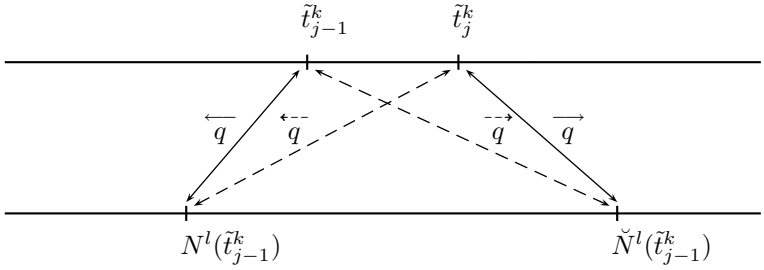

Figure 2.2: Illustration of the definition of $\overleftarrow{q}$, $\overrightarrow{q}$, $\overleftarrow{q}$ and $\overrightarrow{q}$. The indices h, s, h', s' have been suppressed.

the set of all (h, s)-ticks minus the set of ticks fulfilling the condition in equation (2.4).

Each pair $r^k_{\tilde{t}^k_{js+h}} r^l_{t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})}:t^l_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1}}$ can be decomposed as

$$\begin{aligned} r^k_{\tilde{t}^k_{js+h}} r^l_{t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})}:t^l_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1}} &= r^{*k}_{\tilde{t}^k_{js+h}} r^{*l}_{t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})}:t^l_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1}} \\ &\quad + e^k_{\tilde{t}^k_{js+h}} e^l_{t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})}:t^l_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1}} . \end{aligned} \tag{2.5}$$

The first product on the right-hand side of equation (2.5) contributes to the estimation of the integrated covariance, which we would like to measure. The second one is due to noise and we examine it further:

$$\begin{aligned} & e^k_{\tilde{t}^k_{js+h}} e^l_{t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})}:t^l_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1}} \\ &\quad = \left(u^k_{\tilde{t}^k_{js+h}} - u^k_{\tilde{t}^k_{(j-1)s+h}}\right)\left(u^l_{t_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1}} - u^l_{t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})}}\right) \\ &\quad = \gamma_{k,l}\left(\overleftarrow{q}\right) + \gamma_{k,l}\left(\overrightarrow{q}\right) - \gamma_{k,l}\left(\overleftarrow{q}\right) - \gamma_{k,l}\left(\overrightarrow{q}\right), \end{aligned} \tag{2.6}$$

where

$$\begin{aligned} \overleftarrow{q} &= \tilde{t}^k_{(j-1)s+h} - t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})} \\ \overrightarrow{q} &= \tilde{t}^k_{js+h} - t_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1} \\ \overleftarrow{q} &= \tilde{t}^k_{js+h} - t^l_{N^l_{h',s'}(\tilde{t}^k_{(j-1)s+h})} \\ \overrightarrow{q} &= \tilde{t}^k_{(j-1)s+h} - t_{\check{N}^l_{h',s'}(\tilde{t}^k_{js+h})+1} \end{aligned}$$

are the time spans in seconds between the four returns' endpoints and therefore the cross-correlation orders in the autocorrelation function $\gamma_{k,l}(q)$.

Note that we always have $\overleftarrow{q} \geq 0$, $\overrightarrow{q} \leq 0$, $\overleftarrow{\dashrightarrow}{q} > 0$ and $\overrightarrow{\dashrightarrow}{q} < 0$. In Figure 2.2 we illustrate graphically the definition of $\overleftarrow{q}$, $\overrightarrow{q}$, $\overleftarrow{\dashrightarrow}{q}$ and $\overrightarrow{\dashrightarrow}{q}$. The number of $\overleftarrow{q}$ which are of a given length q is given by

$$\overleftarrow{N}^k_{h,s,h',s'}(q) = \sum_{\overleftarrow{q}} \mathbb{1}_{\left\{\overleftarrow{q}=q\right\}} = \sum_j \mathbb{1}_{\left\{\tilde{t}^k_{(j-1)s+h} - t^l_{N^l_{h',s'}\left(\tilde{t}^k_{(j-1)s+h}\right)} = q\right\}}.$$

Similarly, we can define $\overrightarrow{N}^k_{h,s,h',s'}(q)$, $\overleftarrow{\dashrightarrow}{N}^k_{h,s,h',s'}(q)$ and $\overrightarrow{\dashrightarrow}{N}^k_{h,s,h',s'}(q)$. Finally, the number of occurrences of $\gamma_{k,l}(q)$ in the expectation of the HY estimator will be

$$N^k_{h,s,h',s'}(q) = \overleftarrow{N}^k_{h,s,h',s'}(q) + \overrightarrow{N}^k_{h,s,h',s'}(q) - \overleftarrow{\dashrightarrow}{N}^k_{h,s,h',s'}(q) - \overrightarrow{\dashrightarrow}{N}^k_{h,s,h',s'}(q).$$

Then, under Assumptions 2.1 and 2.2 it holds that

$$\begin{aligned} \mathrm{E}\left[HY^{k,l}(h,s,h',s')\right] &= \mathrm{IC}^{k,l} + \sum_{q=-\infty}^{\infty} N^k_{h,s,h',s'}(q)\gamma_{k,l}(q) \\ &\approx \mathrm{IC}^{k,l} + \sum_{q=-Q}^{Q} N^k_{h,s,h',s'}(q)\gamma_{k,l}(q), \end{aligned}$$

where $\mathrm{IC}^{k,l}$ is the integrated covariance of the price processes of assets k and l, i.e. element (k,l) of **IC**. The corresponding pooled OLS regression is

$$\begin{aligned} y_{h,s,h',s'} = c + \beta' x_{h,s,h',s'} + \varepsilon_{h,s,h',s'}, \qquad & s = 1,\ldots,S, \quad h = 1,\ldots,s, \\ & s' = 1,\ldots,S', \quad h' = 1,\ldots,s', \end{aligned} \tag{2.7}$$

where $y_{h,s,h',s'} = HY^{k,l}(h,s,h',s')$ and $x_{h,s,h',s'}$ is the $(2Q+1)$-dimensional vector given by $x_{h,s} = (N^k_{h,s,h',s'}(-Q), N^k_{h,s,h',s'}(-Q+1), \ldots, N^k_{h,s,h',s'}(0), \ldots, N^k_{h,s,h',s'}(Q-1), N^k_{h,s,h',s'}(Q))'$, and Q is chosen suitably.

2.4 Monte Carlo Study

In this section we present the results of a Monte Carlo experiment designed to compare the bias and variance of a number of high-frequency volatility and covolatlity estimators for a broad set of different trading scenarios.

2.4.1 Simulation Setup

We simulate two univariate price processes $p^{*k}(t)$ and $p^{*l}(t)$ with the following stochastic differential equations:

$$dp^{*k}(t) = \sigma_k(t)dW_k, \qquad dp^{*l}(t) = \sigma_l(t)dW_l, \tag{2.8}$$

where $\sigma_k(t)$ and $\sigma_l(t)$ follow GARCH diffusion processes given below and $\langle W_l, W_k \rangle_t = \rho$, i.e., the efficient price processes have stochastic volatility but constant correlation and hence the covariation process is also stochastic. While this setup can be extended by allowing for stochastic correlation, we find this unnecessary here.[3] The volatility processes are modelled as in Andersen & Bollerslev (1998) by

$$\begin{aligned} d\sigma_k^2(t) &= \theta_k(\omega_k - \sigma_k^2(t))dt + \sqrt{2\lambda_k\theta_k}\sigma_k^2(t)dW_k^\sigma(t) \\ d\sigma_l^2(t) &= \theta_l(\omega_l - \sigma_l^2(t))dt + \sqrt{2\lambda_l\theta_l}\sigma_l^2(t)dW_l^\sigma(t) \end{aligned}$$

where $\lambda_k > 0$, $\lambda_l > 0$, $\omega_k > 0$, $\omega_l > 0$, $0 > \theta_k > 1$, $0 > \theta_l > 1$ and the Brownian motions W_k^σ and W_l^σ are independent and also independent of W_k and W_l. Within this framework the integrated covariation matrix is given by

$$\mathbf{IC} = \int_0^1 \begin{pmatrix} \sigma_k^2(t) & \bullet \\ \rho\sigma_k(t)\sigma_l(t) & \sigma_l^2(t) \end{pmatrix} dt$$

The price and volatility processes are generated on a one-second grid for a total of 23400 seconds, corresponding to a typical trading session of 6.5 hours. The parameter values we use are as follows: $\lambda_k = 0.296$, $\lambda_l = 0.480$, $\omega_k = 0.636$, $\omega_l = 0.476$, $\theta_k = 0.035$, $\theta_l = 0.054$. These values have been obtained by Andersen & Bollerslev (1998) for the DM/USD and JPY/USD exchange rates. A similar simulation setup is employed by Barndorff-Nielsen & Shephard (2004), Christensen & Podolskij (2007) and Renò (2001), among others.

The noise processes u_t^k and u_t^l are generated as a bivariate VAR(1) process on the same grid as the price processes:

$$\begin{pmatrix} u_t^k \\ u_t^l \end{pmatrix} = \Phi \begin{pmatrix} u_{t-1}^k \\ u_{t-1}^l \end{pmatrix} + \begin{pmatrix} \varepsilon_t^k \\ \varepsilon_t^l \end{pmatrix},$$

where the matrix $\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$ fulfils the conditions for stationarity of a VAR(1)

[3] Voev & Lunde (2007) find in their simulation study that stochastic vs. constant correlation, in the case of stochastic volatility, does not influence the performance of the covariance estimators.

process and

$$\varepsilon_t \equiv \begin{pmatrix} \varepsilon_t^k \\ \varepsilon_t^l \end{pmatrix} \sim N(0, \Sigma_\varepsilon)$$

is a bivariate white noise process. Obviously, the i.i.d. noise case is obtained for $\Phi = 0$, and the two noise processes are i.i.d. and uncorrelated across assets if, in addition to $\Phi = 0$, Σ_ε is diagonal. An alternative specification to simulate a serially dependent noise process is to model it as an Ornstein-Uhlenbeck process. This possibility is discussed in Aït-Sahalia, Mykland & Zhang (2005), who note that under such an assumption, the noise variance depends on the particular time interval between two observations, i.e., it is of the same order as the integrated variance of the efficient price process $\mathbf{p}^*$. Empirically, however, the market microstructure noise accumulates with the number of observations within a given time interval, and its variance does not go to zero as the time interval shrinks, but rather stays constant. To achieve this one could combine an Ornstein-Uhlenbeck process with an i.i.d. process with constant variance. In this study, instead of adopting this approach which we think will make the simulation intransparent, we approximate this continuous-time structure by a discrete time VAR(1) process on the finest time grid, which we consider a good description of the empirically observed properties of the noise process.

After both the price and the noise processes are generated we obtain the noisy prices

$$\mathbf{p}_t = \mathbf{p}_t^* + \mathbf{u}_t$$

for each $t = 1, \ldots, 23400$ on the second-by-second grid. To obtain different scenarios in terms of trading activity we generate random Poisson sampling times with constant intensities η_k and η_l for asset k and l, respectively.

In our simulation study, we keep the parameters pertaining to the volatility specification fixed, while we vary the parameters Φ, Σ_ε, η_k and η_l to reproduce various noise and trading intensity scenarios. The parameter constellations we consider are presented in Table 2.1. We consider three types of noise: i.i.d., dependent with low persistence, and dependent with high persistence. Each of these specifications is combined with large, moderate and low variances (diagonal elements of Σ_ε) of the white noise process ε_t. Furthermore, we generate observation arrival processes with very high, moderate and low intensities, to obtain a total of 27 scenarios.

Table 2.1: Monte Carlo Simulation Scenarios. We use the following abbreviations: "var." stands for variance, "mod." stands for moderate, "pers." stands for persistence, "int." stands for intensity. The correlation between ε_k and ε_l is set in all scenarios equal to -0.1.

Scenario	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	$\mathrm{V}(\varepsilon_k)$	$\mathrm{V}(\varepsilon_l)$	η_k	η_l
iid, low var., high int.	0	0	0	0	0.0001	0.0002	$\frac{1}{2}$	$\frac{1}{4}$
iid, low var., mod. int.	0	0	0	0	0.0001	0.0002	$\frac{1}{5}$	$\frac{1}{10}$
iid, low var., low int.	0	0	0	0	0.0001	0.0002	$\frac{1}{15}$	$\frac{1}{30}$
iid, mod. var., high int.	0	0	0	0	0.001	0.002	$\frac{1}{2}$	$\frac{1}{4}$
iid, mod. var., mod. int.	0	0	0	0	0.001	0.002	$\frac{1}{5}$	$\frac{1}{10}$
iid, mod. var., low int.	0	0	0	0	0.001	0.002	$\frac{1}{15}$	$\frac{1}{30}$
iid, high var., high int.	0	0	0	0	0.01	0.02	$\frac{1}{2}$	$\frac{1}{4}$
iid, high var., mod. int.	0	0	0	0	0.01	0.02	$\frac{1}{5}$	$\frac{1}{10}$
iid, high var., low int.	0	0	0	0	0.01	0.02	$\frac{1}{15}$	$\frac{1}{30}$
low pers., low var., high int.	0.4	0.1	0.2	0.5	0.0001	0.0002	$\frac{1}{2}$	$\frac{1}{4}$
low pers., low var., mod. int.	0.4	0.1	0.2	0.5	0.0001	0.0002	$\frac{1}{5}$	$\frac{1}{10}$
low pers., low var., low int.	0.4	0.1	0.2	0.5	0.0001	0.0002	$\frac{1}{15}$	$\frac{1}{30}$
low pers., mod. var., high int.	0.4	0.1	0.2	0.5	0.001	0.002	$\frac{1}{2}$	$\frac{1}{4}$
low pers., mod. var., mod. int.	0.4	0.1	0.2	0.5	0.001	0.002	$\frac{1}{5}$	$\frac{1}{10}$
low pers., mod. var., low int.	0.4	0.1	0.2	0.5	0.001	0.002	$\frac{1}{15}$	$\frac{1}{30}$
low pers., high var., high int.	0.4	0.1	0.2	0.5	0.01	0.02	$\frac{1}{2}$	$\frac{1}{4}$
low pers., high var., mod. int.	0.4	0.1	0.2	0.5	0.01	0.02	$\frac{1}{5}$	$\frac{1}{10}$
low pers., high var., low int.	0.4	0.1	0.2	0.5	0.01	0.02	$\frac{1}{15}$	$\frac{1}{30}$
high pers., low var., high int.	0.85	0.15	-0.1	0.85	0.0001	0.0002	$\frac{1}{2}$	$\frac{1}{4}$
high pers., low var., mod. int.	0.85	0.15	-0.1	0.85	0.0001	0.0002	$\frac{1}{5}$	$\frac{1}{10}$
high pers., low var., low int.	0.85	0.15	-0.1	0.85	0.0001	0.0002	$\frac{1}{15}$	$\frac{1}{30}$
high pers., mod. var., high int.	0.85	0.15	-0.1	0.85	0.001	0.002	$\frac{1}{2}$	$\frac{1}{4}$
high pers., mod. var., mod. int.	0.85	0.15	-0.1	0.85	0.001	0.002	$\frac{1}{5}$	$\frac{1}{10}$
high pers., mod. var., low int.	0.85	0.15	-0.1	0.85	0.001	0.002	$\frac{1}{15}$	$\frac{1}{30}$
high pers., high var., high int.	0.85	0.15	-0.1	0.85	0.01	0.02	$\frac{1}{2}$	$\frac{1}{4}$
high pers., high var., mod. int.	0.85	0.15	-0.1	0.85	0.01	0.02	$\frac{1}{5}$	$\frac{1}{10}$
high pers., high var., low int.	0.85	0.15	-0.1	0.85	0.01	0.02	$\frac{1}{15}$	$\frac{1}{30}$

2.4.2 Estimators

In order to compare the performance of our estimation approach against other techniques proposed in the literature, we include a large set of estimators as alternatives. For the variance case we consider the standard realized volatility at different sampling frequencies, including the optimal sampling frequency derived in Bandi & Russell (2005*a*), realized volatility with lag correction, the realized kernels of Barndorff-Nielsen et al. (2006) and the two-scales estimator of Zhang et al. (2005). For the estimation of the integrated covariance we consider the realized covariance computed at different sampling frequencies, including the optimal sampling frequency derived in Bandi & Russell (2005*b*), realized covariance with lead/lag correction, and the

HY estimator along with its subsampled version proposed by Voev & Lunde (2007). Our estimators are the estimated constants in the OLS regressions in equations (2.2) and (2.7) for the integrated variance and covariance, respectively. In our approach there are two parameters that need to be chosen: Q – the number of lags for the (cross) autocovariance function of the noise processes, and S – the number of subsamples. We discuss the choice of these parameters after we setup the notation for the estimators.

The standard realized variance is denoted by $RV(\delta)$, where δ is sampling frequency in seconds, while by $RV_L(\delta)$ we denote the realized variance with a lag correction of L lags. For the realized kernels we use the notation $K^{TH2}(\delta)$ for the modified Tukey-Hanning kernel as described in Barndorff-Nielsen et al. (2006). The two scales realized variance of Zhang et al. (2005) is a combination of an averaged subsampled realized variance at moderate frequencies combined with a very high frequency realized variance correction term and is denoted by $TSRV$. The multi-scale realized variance of Zhang (2006a) is not included as Barndorff-Nielsen et al. (2006) show that its asymptotic distribution is the same as for a realized kernel with a cubic kernel function, which is outperformed by the modified Tukey-Hanning kernel used in this paper.

It should be noted that the TSRV estimator and realized kernels can be constructed differently, i.e., by choosing a different number of subgrids or kernel window, respectively. While we cannot guarantee that these estimators are always constructed optimally, we consider this as a further indication for the robustness of our approach, as our estimators naturally adapt to various noise specifications.

The usual last-tick interpolation realized covariance is denoted by $RC(\delta)$, while its biased corrected version, with L^+ leads and L^- lags, is denoted by $RC_{L^+,L^-}(\delta)$. The Hayashi-Yoshida estimator is denoted as above by HY and its subsampled version based on S subsamples by $HY(S)$. Finally, we denote our estimators by $NV(S,Q)$ in the variance case, and $NV(S,S',Q^+,Q^-)$ in the covariance case. In our Monte Carlo study, we set $S'=1$ for simplicity, i.e., we only consider subsamples of the first asset. The estimators could be improved by subsampling the second asset as well.

For many of the estimators listed above, in order to determine optimal sampling frequencies or optimal numbers of subgrids, one needs to estimate in a first step the second moments of the noise process, as well as the integrated quarticity of the efficient price process, which we denote by IQ^k for asset k. Although there are various estimators for these quantities, we adopt a simple approach following Barndorff-Nielsen et al. (2006). Thus, the noise variances $\gamma_{k,k}(0)$, $\gamma_{l,l}(0)$ are obtained by averaging over subsampled realized variances computed at 60-second grids and

dividing by twice the number of returns, while the integrated quarticity is obtained as the square of the average over realized variance computed at sampling frequency of 20 minutes. While there are currently better estimators for these quantities, this methodology is robust to a fairly large range of noise specifications and delivers reasonable estimates. In order to check whether the results are affected by the fact that we estimate these quantities, we also consider the infeasible versions of the estimators in which we use the true noise variances and integrated quarticity. In the covariance case, we need to estimate $\gamma_{k,l}(0)$ and a quantity which corresponds to the integrated quarticity, which we denote by $\mathrm{IQ}^{k,l}$ and is given by

$$\mathrm{IQ}^{k,l} = \int_0^1 \sigma_{k,k}(t)\sigma_{l,l}(t) + \sigma_{k,l}^2(t)dt,$$

where $\sigma_{k,l}(t)$ is the (k,l)-element of $\mathbf{\Sigma}(t)$.[4] To estimate this quantity we rely on the approach proposed by Bandi & Russell (2005*b*). Having estimated $\gamma_{k,k}(0)$, $\gamma_{l,l}(0)$, IQ^k and $\mathrm{IQ}^{k,l}$, the optimal sampling frequency is given by $\delta^* = \left\lceil \frac{23400}{N^*} \right\rceil$, where N^* is determined by

$$N^* = \begin{cases} \left(\frac{\mathrm{IQ}^k}{\gamma_{k,k}^2(0)}\right)^{\frac{1}{3}}, & \text{in the variance case} \\ \left(\frac{\mathrm{IQ}^{k,l}}{2\gamma_{k,l}^2(0)}\right)^{\frac{1}{3}}, & \text{in the covariance case.} \end{cases} \tag{2.9}$$

For the optimal number of subgrids, we rely on results derived in Zhang et al. (2005) in the variance case and Voev & Lunde (2007) in the covariance case. Thus we determine

$$S^* = c^{1/3}N^{2/3}, \tag{2.10}$$

where N is the total number of observations of the asset under consideration in the variance case, while in the covariance case, N is the number of observations of the slower asset and the optimal constant c is given by

$$c = \begin{cases} \frac{12\gamma_{k,k}^2(0)}{\mathrm{IQ}^k}, & \text{in the variance case} \\ \frac{12\left(\gamma_{k,k}^2(0)\gamma_{l,l}^2(0)+\gamma_{k,l}^2(0)\right)}{\mathrm{IQ}^{k,l}}, & \text{in the covariance case.} \end{cases}$$

2.4.3 Simulation Results

The full results of the Monte Carlo study are not presented here due to space limitations, but are available upon request. Before we present summarized results, it is

[4] Barndorff-Nielsen & Shephard (2004) show that this quantity is the asymptotic variance of the realized covariance estimator.

important to note that our estimators clearly outperform all other considered estimators both in the variance, as well as in the covariance case! Our main competitors are as expected the realized kernel and the TSRV in the variance case, and the Bandi & Russell (2005*b*) realized covariance as well as the HY-type estimators (in some very particular scenarios) in the covariance case.

The considered realized kernel is the only other estimator, apart from our estimator, that delivers unbiased estimates across the range of Monte Carlo scenarios. It is, however, clearly outperformed in the i.i.d. case by the TSRV, while our estimator is not. In order to check whether it could be that the inputs required for the construction of the realized kernels and the TSRV impairs their performance, we computed their infeasible versions by setting the unknown quantities (e.g., the integrated quarticity or noise variance) to their true values. Overall, this did not qualitatively influence the results, implying that the estimates we use to construct the feasible estimators are reasonable.

In the covariance estimation, the subsampled HY estimator performs better than our estimator only in the case of i.i.d. noise with low variance and moderate or low trading intensity. In all other cases, it is severely biased and hence not competitive. The Bandi & Russell (2005*b*) estimator with a first-order lead/lag correction performs quite well and is second only to our estimator the best alternative.

A very nice feature of our approach is that the proposed estimator is very robust and not too sensitive to the choice of the number of subsamples S. What is important, however, is that Q is chosen reasonably, which on the one hand means that it should not be too low (omitted variable problem) in the case of highly persistent noise, and on the other hand not too close to $S(S+1)/2$ (the number of observations in the pooled OLS regression) to assure that the X matrix is not close to being singular. As we do not have a rule to determine S and/or Q according to a theoretically based optimality criterion, we employ a data-driven model selection. In particular we estimate our models for some predetermined set of S and Q values and then select a specification based on an information criterion. In this study we consider the set of values $Q = \{0, 10, 20\}$ (also for the Q^+ and Q^- in the covariance case) and $S = \{S^*, 2S^*, 3S^*\}$, thus obtaining a set of nine estimators for each scenario. It is important to note, that these values are arbitrarily chosen and can lead to certain ill-specified X matrices. In particular, when we have low intensity specifications, the number of observations is small, and hence S^* is small. If Q is then chosen large, then the X matrix is near-singular (number of cases in the OLS regression close to the number of regressors). In our simulations this occurs for the models $NV(S^*, 10)$, $NV(S^*, 20)$ (in the variance case) and $NV(S^*, 1, 10, 10)$, $NV(S^*, 1, 20, 20)$ (in the covariance case) in combination with very low trading intensities (low intensity spec-

ification for asset 2, corresponding to a trade every 30 seconds on average) for which the S^* is too small. Such S, Q-combinations should be avoided and can be identified by a nearly perfect fit in the OLS regression (R-squared very close to one and sum of squared residuals almost equal to zero). While we report results for these estimators in the Appendix, we exclude them from the set of models from which we select the optimal model.

For practical purposes, we propose two criteria for choosing the proper combination of the parameters S and Q, both based on the goodness-of-fit of the regression, given by

$$BIC = \ln\left(\frac{1}{n}\sum_{h,s}\hat{\varepsilon}^2_{h,s}\right) + \frac{p\ln n}{n}, \quad \text{with } n = \frac{S(S+1)}{2} \tag{2.11}$$

$$BIC^* = \ln\left(\frac{1}{n}\sum_{h,s}\frac{\hat{\varepsilon}^2_{h,s}}{s}\right) + \frac{p\ln n}{n}, \quad \text{with } n = S, \tag{2.12}$$

where $\hat{\varepsilon}_{h,s}$ is the pooled OLS regression residual and p is the number of parameters in the regression. The first criterion is the usual Bayesian Information Criterion (BIC) for the pooled regression, while the second one is a modified BIC for a regression over $s = 1, \ldots, S$, where the squared residual for each s is the mean squared residual of the s-block. The modified BIC is motivated by the fact that the number of elements in the s-block, which equally contribute to the estimation, is linearly increasing with s. It accounts for the fact that as s increases, a single (h, s)-observation becomes more noisy and therefore should be counted with an accordingly smaller weight.

Let us have a closer look at the "average" scenario (Table 2.4 in the Appendix) with moderate noise variances (0.001 and 0.002), low noise persistence and moderate trading intensities (1/5 and 1/10), which we believe is rather realistic. In this case the modified BIC criterion selects $NV(3S^*, 10)$ for the variance of the first asset, $NV(3S^*, 0)$ for the variance of the second asset, and $NV(3S^*, 1, 0, 0)$ for the covariance. We note that for this scenario, the modified BIC selects suboptimal models for the first variance, for which the selected model ranks 5th among the NV models with respect to the root mean squared error (RMSE), and for the covariance for which the selected model ranks 3rd among the NV models. Despite this suboptimal choice, the efficiency gain of the selected models compared to the best alternative outside the NV class, in terms of the RMSE, is 40%, 37%, and 47%, for the variance of the first asset, the variance of the second asset, and the covariance, respectively.

In Tables 2.2 and 2.3 we report summary statistics of the ranks of all considered estimators, according to their RMSE, across simulation scenarios. Additionally, the tables contain statistics on the RMSE of all estimators relative to the RMSE of the

estimator selected by the BIC (in Table 2.2) and by the BIC^* (in Table 2.3).

Table 2.2: Mean, median, maximum and minimum of the root mean squared error (RMSE) rankings and of the relative RMSE across simulation scenarios. The relative RMSE is computed relative to the $NV(BIC)$ estimator, which is for each scenario the model with the smallest Bayesian Information Criterion (Equation (2.11)) across all NV estimators.

	Model Ranking				Relative RMSE			
Model	Mean	Median	Min	Max	Mean	Median	Min	Max
				IV[1]				
$RV(5)$	13.9	14	12	14	324.00	102.76	5.67	1587.26
$RV(300)$	7.7	7	2	12	10.93	3.67	1.77	48.39
$RV(900)$	7.2	6	3	11	5.13	3.47	2.29	16.14
$RV(1800)$	7.9	9	4	11	4.55	3.89	2.45	8.63
$RV_1(5)$	11.8	13	3	13	146.71	32.34	1.59	988.54
$RV_1(300)$	5.7	6	3	8	3.81	3.47	2.15	7.96
$RV_1(900)$	8.8	10	3	13	4.82	4.33	2.85	10.41
$RV_1(1800)$	10.3	12	3	14	6.13	5.20	3.47	14.33
$RV(\delta^*)$	6.4	5	3	10	3.85	2.89	1.83	9.30
$RV_1(\delta^*)$	6.7	7	3	9	3.73	3.52	2.64	6.25
$RV_2(\delta^*)$	8.8	9	5	10	4.40	4.29	3.14	7.52
$K^{TH2}(60)$	3.3	3	2	6	2.37	2.06	1.65	4.61
$TSRV$	5.3	2	1	12	7.03	1.39	0.80	63.48
$NV(BIC)$	1.2	1	1	2	1.00	1.00	1.00	1.00
				IV[2]				
$RV(5)$	14.0	14	13	14	379.89	111.34	4.64	2113.11
$RV(300)$	9.1	12	3	12	16.12	5.66	1.54	63.74
$RV(900)$	7.8	6	4	11	6.36	3.53	1.58	21.51
$RV(1800)$	7.8	8	4	10	4.53	3.60	2.17	11.25
$RV_1(5)$	12.9	13	11	13	205.10	70.72	3.91	1241.08
$RV_1(300)$	6.2	7	3	9	4.04	2.87	1.65	10.52
$RV_1(900)$	8.0	9	4	12	4.18	3.92	2.67	8.01
$RV_1(1800)$	8.9	11	3	14	5.00	4.55	3.26	11.37
$RV(\delta^*)$	7.0	8	3	10	4.49	3.35	1.42	11.62
$RV_1(\delta^*)$	6.6	7	3	8	3.47	3.30	1.89	5.42
$RV_2(\delta^*)$	8.3	9	5	10	4.02	3.87	2.21	6.37
$K^{TH2}(60)$	3.1	3	2	5	2.07	2.03	1.46	3.68
$TSRV$	4.2	2	1	12	4.58	1.42	0.85	41.04
$NV(BIC)$	1.2	1	1	2	1.00	1.00	1.00	1.00
				IC				
$RC(5)$	10.8	12	1	14	32.04	4.63	0.96	354.59
$RC(300)$	6.1	7	2	10	4.04	2.22	0.69	21.84
$RC(900)$	6.2	6	2	10	3.35	2.97	0.95	9.37
$RC(1800)$	7.4	8	2	12	3.51	3.27	1.22	6.82
$RC_{1,1}(5)$	10.9	12	4	14	46.91	6.70	0.90	549.92
$RC_{1,1}(300)$	7.2	7	4	11	3.49	3.10	0.95	8.56
$RC_{1,1}(900)$	8.9	11	3	13	4.06	3.79	1.46	8.27
$RC_{1,1}(1800)$	10.1	12	2	14	5.10	4.62	2.03	11.74
$RC(\delta^*)$	7.6	8	2	12	4.64	3.02	0.78	19.38
$RC_{1,1}(\delta^*)$	5.0	4	2	10	3.22	2.10	0.75	8.34
$RC_{2,2}(\delta^*)$	6.1	6	3	10	3.33	2.23	0.76	8.10
HY	10.4	12	2	14	27.13	7.97	0.55	159.16
$HY(S^*)$	6.3	7	1	11	7.67	3.14	0.44	53.74
$NV(BIC)$	2.0	1	1	9	1.00	1.00	1.00	1.00

Table 2.3: Mean, median, maximum and minimum of the root mean squared error (RMSE) rankings and of the relative RMSE across simulation scenarios. The relative RMSE is computed relative to the $NV(BIC^*)$ estimator, which is for each scenario the model with the smallest modified Bayesian Information Criterion (Equation (2.12)) across all NV estimators.

	Model Ranking				Relative RMSE			
Model	Mean	Median	Min	Max	Mean	Median	Min	Max
	IV[1]							
$RV(5)$	13.9	14	12	14	322.98	102.76	5.67	1383.74
$RV(300)$	7.7	7	2	12	10.80	3.67	1.07	45.97
$RV(900)$	7.2	6	3	11	5.06	3.47	1.38	15.64
$RV(1800)$	7.9	9	4	11	4.48	3.89	1.95	8.63
$RV_1(5)$	11.8	13	3	13	142.67	32.34	1.59	891.90
$RV_1(300)$	5.7	6	3	8	3.76	3.47	1.42	7.93
$RV_1(900)$	8.8	10	3	13	4.75	4.57	2.38	10.41
$RV_1(1800)$	10.3	12	3	14	6.04	5.20	3.19	14.33
$RV(\delta^*)$	6.4	5	3	10	3.82	2.89	1.10	9.25
$RV_1(\delta^*)$	6.7	7	3	9	3.69	3.51	1.59	6.25
$RV_2(\delta^*)$	8.8	9	5	10	4.35	4.28	1.89	7.52
$K^{TH2}(60)$	3.3	3	2	6	2.33	2.11	1.23	4.61
$TSRV$	5.3	2	1	12	6.70	1.08	0.80	55.34
$NV(BIC^*)$	1.2	1	1	2	1.00	1.00	1.00	1.00
	IV[2]							
$RV(5)$	14.0	14	13	14	385.82	123.71	4.64	2086.80
$RV(300)$	9.1	12	3	12	16.27	5.72	1.54	62.95
$RV(900)$	7.8	6	4	11	6.42	3.53	1.58	21.24
$RV(1800)$	7.8	8	4	10	4.57	3.60	2.17	11.11
$RV_1(5)$	12.9	13	11	13	206.57	71.81	3.91	1225.62
$RV_1(300)$	6.2	7	3	9	4.08	2.87	1.65	10.39
$RV_1(900)$	8.0	9	4	12	4.22	3.92	2.67	8.01
$RV_1(1800)$	8.9	11	3	14	5.05	4.89	3.26	11.37
$RV(\delta^*)$	7.0	8	3	10	4.53	3.35	1.42	11.47
$RV_1(\delta^*)$	6.6	7	3	8	3.51	3.30	1.89	5.35
$RV_2(\delta^*)$	8.3	9	5	10	4.07	3.87	2.21	6.37
$K^{TH2}(60)$	3.1	3	2	5	2.09	2.03	1.46	3.68
$TSRV$	4.2	2	1	12	4.59	1.42	0.85	40.53
$NV(BIC^*)$	1.2	1	1	2	1.00	1.00	1.00	1.00
	IC							
$RC(5)$	10.8	12	1	14	30.22	4.63	0.96	362.70
$RC(300)$	6.2	7	2	10	3.87	2.22	1.37	22.34
$RC(900)$	6.2	6	2	10	3.32	2.97	1.91	9.58
$RC(1800)$	7.4	8	2	12	3.55	3.22	2.20	6.82
$RC_{1,1}(5)$	10.9	12	4	14	42.79	6.24	1.34	562.50
$RC_{1,1}(300)$	7.2	7	4	11	3.46	3.10	1.91	8.75
$RC_{1,1}(900)$	8.9	11	3	13	4.12	3.79	2.60	8.27
$RC_{1,1}(1800)$	10.1	12	2	14	5.24	4.88	2.58	11.74
$RC(\delta^*)$	7.7	8	3	12	4.54	2.91	1.04	19.83
$RC_{1,1}(\delta^*)$	5.2	4	2	10	3.17	2.21	1.19	8.53
$RC_{2,2}(\delta^*)$	6.3	6	3	10	3.29	2.33	1.42	8.13
HY	10.5	12	2	14	24.63	7.97	0.80	118.20
$HY(S^*)$	6.3	7	1	11	7.17	3.14	0.79	54.97
$NV(BIC^*)$	1.3	1	1	3	1.00	1.00	1.00	1.00

An alternative model selection strategy could rely on a procedure, in which one starts from a high value of Q which is sequentially reduced. Following this procedure, it can be detected whether the estimates change significantly as Q becomes smaller. If this is the case, there is an indication for the presence of persistence in the noise processes and consequently Q should be chosen preferably a bit too high rather than too low. Whenever Q is chosen to be large, it is beneficial to choose S large as well, since as mentioned above and explicit in the simulation results, one cannot choose S too poorly by choosing it too large, which also alleviates the discussed near-singularity problem.

From Tables 2.2 and 2.3 it is clear that our models are dominating and that the proposed selection criteria, although not perfect are doing a very good job. In particular, the modified BIC, which is motivated by the pooled form of the regression, is more robust and selects better models. Considering the model ranking for all 28 variance models, including all of our nine specifications, we observe that there is at least one estimator from the NV class that outperforms all others in *each* Monte Carlo scenario in the variance case, although it is not always selected by the BIC or the modified BIC! This presents possibilities for improvement of the selection criteria, which we consider to be a possible area of further research.

In the covariance case we observe a similar pattern with three exceptions: the $HY(S^*)$ ranks first in the iid., low variance, moderate intensity and iid., low variance, low intensity scenarios, while the $RC(5)$ ranks first in the low persistence, low variance and high intensity scenario. The last case is a coincidence, in which the noise induced positive bias cancels almost exactly against the negative bias caused by the Epps effect.

An interesting issue which has to be addressed is whether the separate unconstrained estimation of variances and covariances leads to a well-defined covariance matrix. In the Appendix, we have added in the last row of the tables for the covariance estimators the number of cases (out of the 1000 replications) for which the matrix resulting from the particular covariance estimator combined with the two corresponding variance estimators was non-positive definite. In most scenarios and for most estimators this never happens. We observe, however, that for the estimators for which the choice of S and Q leads to singularity problems, which we discussed above, in the worst case this occurs up to 35% of the cases, which is caused by the very large variance of the particular covariance and/or variance estimators. For the estimators chosen by the BIC and modified BIC criteria, though, a non-positive definite matrix is obtained only in one case in only one scenario, i.e., once out of 27000 cases. We note that the problem of obtaining a positive definite covariance matrix estimate is a serious issue, which has to be addressed and analyzed in a higher dimensional

setup. The results we present here, are just a first indication of the performance of our estimators with respect to this issue, which needs to be pursued further.

2.5 Conclusion

The paper introduces a unified framework for the estimation of integrated second moments of irregularly observed asset prices contaminated by market microstructure noise. The estimation is performed under fairly weak assumptions on the dependence structure of the noise processes in a simple OLS regression framework. This approach allows for a robust estimation of the whole covariance matrix of asset returns in applications with large number of assets. Moreover, we can identify the dependence structure of the noise process, which sheds light on market microstructure properties. We derive the OLS regressions theoretically for the variance and covariance case and perform an extensive Monte Carlo study to compare the performance of our estimators against the most recent and commonly used approaches in the extant literature. The results are unequivocal: our estimators clearly dominate the other approaches across a comprehensive range of trading scenarios.
Promising directions for further research are on the one hand a more in-depth analysis of a model selection criterion based on the statistical properties of our estimators, relaxing the assumption of noise exogeneity, and on the other hand an empirical application with a large number of assets, e.g., in the field of asset pricing or risk management.

Bibliography

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *Review of Financial Studies* **18**(2), 351–416.

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2006), Ultra high frequency volatility estimation with dependent microstructure noise. Working Paper, Princeton University.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Bandi, F. M. & Russell, J. R. (2005*a*), Microstructure noise, realized volatility, and optimal sampling. Working paper, Graduate School of Business, The University of Chicago.

Bandi, F. M. & Russell, J. R. (2005*b*), Realized covariation, realized beta, and microstructure noise. Working paper, Graduate School of Business, The University of Chicago.

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

Barndorff-Nielsen, O., Hansen, P., Lunde, A. & Shephard, N. (2006), Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Working paper, Nuffield College, Oxford.

Christensen, K. & Podolskij, M. (2007), 'Realized range-based estimation of integrated variance', *Journal of Econometrics* **141**, 323–349.

Corsi, F. & Audrino, F. (2007), Realized correlation tick-by-tick. Working paper, University of Lugano.

Curci, G. & Corsi, F. (2006), Discrete sine transform for multi-scales realized volatility measures. Working Paper, University of Lugano.

Epps, T. (1979), 'Comovements in stock prices in the very short run', *Journal of the American Statistical Association* **74**, 291–298.

Griffin, J. E. & Oomen, R. C. A. (2006), Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Working Paper, University of Warwick.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–218.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Martens, M. (2004), Estimating unbiased and precise realized covariances. Econometric Institute, Erasmus University Rotterdam.

Phillips, P. C. & Yu, J. (2006), 'Comment on "realized variance and market microstructure noise" by peter r. hansen and asger lunde', *Journal of Business and Economic Statistics* **24**, 202–208.

Renò, R. (2001), A closer look at the Epps effect. Università degli Studi di Siena, Working paper n. 335.

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

Zhang, L. (2006*a*), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* **12**, 1019–1043.

Zhang, L. (2006*b*), Estimating covariation: Epps effect and microstructure noise. Working Paper.

Zhang, L., Mykland, P. A. & 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.

Appendix

Table 2.4: Monte Carlo simulation results for scenario: **low pers., mod. var., mod. int.** Each cell entry consists of the mean and the standard deviation (in parentheses) over 1000 Monte Carlo replications. δ^* denotes the optimal sampling frequency as in Equation (2.9). S^* denotes the optimal number of subgrids as in Equation (2.10).

Model \ True	IV^1	IV^2	$\gamma_1(0)$	$\gamma_1(1)$	$\gamma_1(2)$	$\gamma_2(0)$	$\gamma_2(1)$	$\gamma_2(2)$
	0.6281	0.5234	0.0012	0.0005	0.0002	0.0028	0.0014	0.0008
$RV(5)$	8.1545	10.4098						
	(0.2501)	(0.4165)						
$RV(300)$	0.8046	0.9499						
	(0.1334)	(0.1660)						
$RV(900)$	0.6639	0.6371						
	(0.1768)	(0.1838)						
$RV(1800)$	0.6098	0.5510						
	(0.2413)	(0.2206)						
$RV_1(5)$	3.2950	6.7987						
	(0.1952)	(0.3617)						
$RV_1(300)$	0.6152	0.5240						
	(0.1875)	(0.1915)						
$RV_1(900)$	0.6048	0.5067						
	(0.2972)	(0.2639)						
$RV_1(1800)$	0.5837	0.5000						
	(0.4059)	(0.3542)						
$RV(\delta^*)$	0.7229	0.6769	0.0020			0.0034		
	(0.1665)	(0.1874)	(0.0001)			(0.0002)		
$RV_1(\delta^*)$	0.5981	0.5146	0.0020			0.0034		
	(0.2290)	(0.2343)	(0.0001)			(0.0002)		
$RV_2(\delta^*)$	0.6171	0.5330	0.0020			0.0034		
	(0.2806)	(0.2994)	(0.0001)			(0.0002)		
$K^{TH2}(60)$	0.6113	0.5139	0.0020			0.0034		
	(0.1553)	(0.1545)	(0.0001)			(0.0002)		
$TSRV$	0.7391	0.6230	0.0020			0.0034		
	(0.0627)	(0.0780)	(0.0001)			(0.0002)		
$NV(S^*,0)$	0.7453	0.6315	0.0011			0.0026		
	(0.0521)	(0.0662)	(0.0000)			(0.0001)		
$NV(2S^*,0)$	0.6722	0.5652	0.0012			0.0026		
	(0.0559)	(0.0685)	(0.0000)			(0.0001)		
$NV(3S^*,0)$	0.6511	0.5463	0.0012			0.0026		
	(0.0643)	(0.0766)	(0.0000)			(0.0001)		
$NV(S^*,10)$	0.6253	0.5213	0.0012	0.0001	0.0026	0.0028	-0.0084	0.0075
	(0.0719)	(0.0890)	(0.0001)	(0.0062)	(0.0339)	(0.0002)	(0.5014)	(0.9626)
$NV(2S^*,10)$	0.6219	0.5201	0.0012	0.0000	0.0026	0.0028	-0.0041	0.0057
	(0.0686)	(0.0813)	(0.0001)	(0.0060)	(0.0316)	(0.0001)	(0.4789)	(0.9682)
$NV(3S^*,10)$	0.6207	0.5193	0.0012	0.0000	0.0021	0.0028	-0.0014	0.0076
	(0.0767)	(0.0888)	(0.0001)	(0.0059)	(0.0302)	(0.0002)	(0.4730)	(0.9754)
$NV(S^*,20)$	0.6235	0.5220	0.0012	-0.0002	0.0046	0.0028	0.0793	0.0002
	(0.0896)	(0.1059)	(0.0001)	(0.0154)	(0.0740)	(0.0002)	(2.8558)	(2.1718)
$NV(2S^*,20)$	0.6210	0.5200	0.0012	-0.0001	0.0040	0.0028	0.0683	-0.0012
	(0.0731)	(0.0863)	(0.0001)	(0.0144)	(0.0636)	(0.0002)	(2.6919)	(2.0950)
$NV(3S^*,20)$	0.6201	0.5191	0.0012	0.0001	0.0030	0.0028	0.0672	-0.0030
	(0.0809)	(0.0937)	(0.0001)	(0.0141)	(0.0596)	(0.0002)	(2.6625)	(2.1380)

Table 2.4 ***(cont'd)*****:** Monte Carlo simulation results for scenario: **low pers., mod. var., mod. int.** Each cell entry consists of the mean and the standard deviation (in parentheses) over 1000 Monte Carlo replications. δ^* denotes the optimal sampling frequency as in Equation (2.9). S^* denotes the optimal number of subgrids as in Equation (2.10). #NonPD stands for the number of non-positive definite matrices out of the 1000 Monte Carlo replications.

Model \ True	IC 0.2005	$\gamma_{12}(-2)$ 0.0005	$\gamma_{12}(-1)$ 0.0006	$\gamma_{12}(0)$ 0.0001	$\gamma_{12}(1)$ 0.0002	$\gamma_{12}(2)$ 0.0001	#NonPD
$RC(5)$	0.3405 (0.1459)						
$RC(300)$	0.2121 (0.1091)						
$RC(900)$	0.1964 (0.1432)						
$RC(1800)$	0.1893 (0.1820)						
$RC_{1,1}(5)$	0.4742 (0.1671)						
$RC_{1,1}(300)$	0.1965 (0.1487)						
$RC_{1,1}(900)$	0.1929 (0.2129)						
$RC_{1,1}(1800)$	0.1956 (0.2950)						
$RC(\delta^*)$	0.2434 (0.1148)			0.0001 (0.0000)			
$RC_{1,1}(\delta^*)$	0.2034 (0.1126)			0.0001 (0.0000)			
$RC_{2,2}(\delta^*)$	0.1966 (0.1241)			0.0001 (0.0000)			
HY	0.7097 (0.1683)						
$HY(S^*)$	0.3830 (0.0594)						
$NV(S^*,1,0,0)$	0.2238 (0.0505)			0.0006 (0.0002)			0
$NV(2S^*,1,0,0)$	0.2106 (0.0515)			0.0006 (0.0002)			0
$NV(3S^*,1,0,0)$	0.2067 (0.0576)			0.0006 (0.0002)			0
$NV(S^*,1,10,10)$	0.2004 (0.0622)	0.0003 (0.0052)	0.0003 (0.0044)	0.0000 (0.0014)	0.0005 (0.0049)	0.0005 (0.0053)	0
$NV(2S^*,1,10,10)$	0.2003 (0.0588)	0.0002 (0.0024)	0.0002 (0.0026)	0.0001 (0.0006)	0.0004 (0.0023)	0.0004 (0.0029)	0
$NV(3S^*,1,10,10)$	0.2002 (0.0646)	0.0002 (0.0020)	0.0003 (0.0020)	0.0001 (0.0005)	0.0004 (0.0019)	0.0003 (0.0022)	0
$NV(S^*,1,20,20)$	0.2012 (0.0842)	0.0016 (0.0244)	0.0014 (0.0227)	0.0002 (0.0023)	-0.0015 (0.0298)	-0.0006 (0.0268)	0
$NV(2S^*,1,20,20)$	0.2002 (0.0650)	0.0003 (0.0050)	0.0004 (0.0049)	0.0001 (0.0006)	0.0003 (0.0050)	0.0002 (0.0053)	0
$NV(3S^*,1,20,20)$	0.2000 (0.0700)	0.0003 (0.0039)	0.0004 (0.0038)	0.0001 (0.0005)	0.0003 (0.0038)	0.0002 (0.0040)	0

Chapter 3

Dynamic Modelling of Large Dimensional Covariance Matrices

3.1 Introduction

Modelling and forecasting the variances and covariances of returns of financial assets is crucial for financial management and portfolio selection and re-balancing. Recently this branch of the econometric literature has grown at a very fast pace. One of the simplest methods used is the sample covariance matrix. A stylized fact, however, is that there is a serial dependence in the second moments of returns. Thus, more sophisticated models had to be developed which incorporate this property, as well as other well-known features of financial return distributions such as leptokurtosis or the so-called "leverage effect". This led to the development of the univariate GARCH processes and their extension - the multivariate GARCH (MGARCH) models (for a comprehensive review see Bauwens, Laurent & Rombouts (2006)), which include also the modelling of covariances. One of the most severe drawbacks of the MGARCH models, however, is the difficulty of handling dimensions higher than 4 or 5 (or with very restrictive assumptions). Another more practically oriented field of research deals with the problem of how to reduce the noise inherent in simpler covariance estimators such as the sample covariance matrix. Techniques have been developed to "shrink" the sample covariance (SC) matrix, thereby reducing its extreme values in order to mitigate the effect of the so-called error maximization noted by Michaud (1989). One of the shrinkage estimators used among practitioners is the Black-Litterman model (Black & Litterman (1992)). This model uses a prior which reflects an investor's beliefs about securities returns and combines it with implied equilibrium expected returns to obtain a posterior distribution, whose variance is a combination of the covariance matrix of implied returns and the confidence of the investor's views (which reflect the prior covariance). Further, Ledoit & Wolf

(2003) and Ledoit & Wolf (2004) use shrinkage methods to combine a SC matrix with a more structured estimator (e.g a matrix with equal pairwise correlations, or a factor estimator). The idea is to combine an asymptotically unbiased estimator having a large variance with a biased estimator, which is considerably less noisy. So the shrinkage actually amounts to optimizing in terms of the well-known trade-off between bias and variance.

Recently, with the availability of high-quality transaction databases, the technique of realized variance and covariance (RC) gained popularity. A very comprehensive treatment of volatility modelling with focus on forecasting appears in Andersen, Bollerslev, Christoffersen & Diebold (2006). Andersen, Bollerslev, Diebold & Ebens (2001), among others, have shown that there is a long-range persistence (long memory) in daily realized volatilities, which allows one to obtain good forecasts by means of fractionally integrated ARMA processes. At the monthly level, we find that the autocorrelations decline quite quickly to zero, which led us to choose standard ARMA models for fitting and forecasting.

The aim of this paper is to compare the forecasting performance of a set of models, which are suitable to handle large dimensional covariance matrices. Letting H denote the set of considered models, we have $H = \{s, ss, rm, rc, src, drc, dsrc\}$, where the first two models are based on the sample covariance matrix, the third model is a RiskMetrics™ exponentially weighted moving average (EWMA) estimator developed by J.P. Morgan (1996), the fourth and the fifth represent simple forecasts based on the realized and on the shrunk realized covariance matrix, and the last two models employ dynamic modelling of the RC and shrunk RC, respectively. We judge the performance of the models by looking at their ability to forecast individual variance and covariance series by employing a battery of Diebold-Mariano (Diebold & Mariano (1995)) tests. Of course, if we have good forecasts for the individual series, then the whole covariance matrix will also be well forecast. The practical relevance of a good forecast can be seen by considering an investor who faces an optimization problem to determine the weights of some portfolio constituents. One of the crucial inputs in this problem is a forecast of future movements and co-movements in asset returns. Our contribution is to propose a methodology which improves upon the sample covariance estimator and is easy to implement even for very large portfolios. We show that in some sense these models are more flexible than the MGARCH models, although this comes at the expense of some complications.

The remainder of the paper is organized as follows: Section 3.2 sets up the notation and describes the forecasting models, Section 3.3 presents the data set used to compare the forecasting performance of the models, Section 3.4 discusses the results on the forecast evaluation and Section 3.5 concludes the paper.

3.2 Forecasting models

In this section we describe each of the covariance forecasting models. First, we introduce some notation and description of the forecasting methodology. We concentrate on one-step ahead forecasts of covariance matrices of N stocks, and consider the monthly frequency. The information is updated every period and a new forecast is formed. Thus, each new forecast incorporates the newest information which has become available. Such a strategy might describe an active long-run investor, who revises and rebalances her portfolio every month. Let the multivariate price process be defined as $\mathbf{P} = \{\mathbf{P}_t(\omega), t \in (-\infty, \infty), \omega \in \Omega\}$, where Ω is an outcome space.[1] The portfolio is set up at $t = 0$ and updated at each $t = 1, 2, \ldots, \bar{T}$, where $\bar{T}$ is the end of the investment period. The frequency of the observations in our application is daily, which we refer to as intra-periods. In this setup, we can formally define the information set at each time $t \geq 0$ as a filtration $\mathcal{F}_t = \sigma(\mathbf{P}_s(\omega), s \in \mathcal{T})$ generated by $\mathbf{P}$, with $\mathcal{T} = \{s : s = -L + \frac{j}{M}, j = 0, 1, \ldots, (L+t)M\}$, M – the number of intra-periods within each period[2] and L – the number of periods, for which price data is available, before the investment period. It is important to note that not all information is considered in the forecasts based on the sample covariance matrix. For these models only the lower frequency monthly sampling is needed. Furthermore, we define the monthly returns as $\mathbf{r}_t = \ln(\mathbf{P}_t) - \ln(\mathbf{P}_{t-1})$, where $\mathbf{P}_t$ is the realization of the price process at time t, and the j^{th} intra-period return by $\mathbf{r}_{t+\frac{j}{M}} = \ln\left(\mathbf{P}_{t+\frac{j}{M}}\right) - \ln\left(\mathbf{P}_{t+\frac{j-1}{M}}\right)$. The realized covariance at time $t+1$ is given by:

$$\mathbf{\Sigma}_{t+1}^{RC} = \sum_{j=1}^{M} \mathbf{r}_{t+\frac{j}{M}} \mathbf{r}'_{t+\frac{j}{M}}. \tag{3.1}$$

Assessing the performance of variance forecasts has been quite problematic, since the true covariance matrix $\mathbf{\Sigma}_t$ is not directly observable. This has long been a hurdle in evaluating GARCH models. Traditionally, the squared daily return was used as a measure of the daily variance. Although this is an unbiased estimator, it has a very large estimation error due to the large idiosyncratic noise component of daily returns. Thus a good model may be evaluated as poor, simply because the target is measured with a large error. In an important paper, Andersen & Bollerslev (1998) showed that GARCH models actually provide good forecasts when the target to which they are compared is estimated more precisely, by means of sum of squared intradaily returns. Since then, it has become a practice to take the realized variance

[1] Of course, in reality the price process could not have started in the infinite past. Since we are interested in when the process became observable, and not in its beginning, we leave the latter unspecified.

[2] This number is not necessarily the same for all periods and should be denoted more precisely by $M(t)$. This is not done in the text to avoid cluttering of the notation.

as the relevant measure for comparing forecasting performance. In this spirit, we use the realized monthly covariance in place of the true matrix. Thus we will assess a given forecast $\hat{\mathbf{\Sigma}}^{(h)}_{t+1|t}$, $h \in H$ by its deviation from $\mathbf{\Sigma}^{RC}_{t+1}$.

3.2.1 A sample covariance forecast

In this section we describe a forecasting strategy based on the sample covariance matrix, which will serve as a benchmark. The sample covariance is a consistent estimator for the true population covariance under weak assumptions. We use a rolling window scheme and define the forecast as:

$$\hat{\mathbf{\Sigma}}^{(s)}_{t+1|t} = \frac{1}{T} \sum_{s=t-T+1}^{t} (\mathbf{r}_s - \bar{\mathbf{r}}_{t,T})(\mathbf{r}_s - \bar{\mathbf{r}}_{t,T})', \tag{3.2}$$

where for each t, $\bar{\mathbf{r}}_{t,T}$ is the sample mean of the return vector $\mathbf{r}$ over the last T observations. We will denote the sample covariance matrix at time t by $\mathbf{\Sigma}^{SC}_t$. For T we choose a value of 60, which with monthly data corresponds to a time span of five years. As the near future is of the highest importance in volatility forecasting, this number might seem too large. Too small a number of periods, however, would lead to a large variance of the estimator, therefore other authors (e.g. Ledoit & Wolf (2004)) have also chosen 60 months as a balance between precision and relevance of the data. A problem of this approach, as simple as it is, is that new information is given the same weight as very old information. Another obvious oversimplification is that we do not account for the serial dependence present in the second moments of financial returns.

3.2.2 A shrinkage sample covariance forecast

In this section we briefly present the shrinkage estimator, proposed by Ledoit & Wolf (2003), in order to give an idea of the shrinkage principle.
The shrinkage estimator of the covariance matrix $\mathbf{\Sigma}_t$ is defined as a weighted linear combination of some shrinkage target F_t and the sample covariance matrix, where the weights are chosen in an optimal way. More formally, the estimator is given by

$$\mathbf{\Sigma}^{SS}_t = \hat{\alpha}^*_t F_t + (1 - \hat{\alpha}^*_t)\mathbf{\Sigma}^{SC}_t, \tag{3.3}$$

$\hat{\alpha}^*_t \in [0, 1]$ is an estimate of the optimal shrinkage constant α^*_t.
The shrinking intensity is chosen to be optimal with respect to a loss function defined as a quadratic distance between the true and the estimated covariance matrices based on the Frobenius norm. The Frobenius norm of an $N \times N$ symmetric matrix Z with

elements $(z_{ij})_{i,j=1,\ldots,N}$ is defined by

$$\|Z\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} z_{ij}^2. \tag{3.4}$$

The quadratic loss function is the Frobenius norm of the difference between $\mathbf{\Sigma}_t^{SS}$ and the true covariance matrix:

$$L(\alpha_t) = \left\|\alpha_t F_t + (1-\alpha_t)\mathbf{\Sigma}_t^{SC} - \mathbf{\Sigma}_t\right\|^2. \tag{3.5}$$

The optimal shrinkage constant is defined as the value of α which minimizes the expected value of the loss function (i.e. the risk) in expression (3.5):

$$\alpha_t^* = \underset{\alpha_t}{\operatorname{argmin}} \operatorname{E}\left[L(\alpha_t)\right]. \tag{3.6}$$

For an arbitrary shrinkage target F and a consistent covariance estimator S, Ledoit & Wolf (2003) show that

$$\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\operatorname{Var}\left[s_{ij}\right] - \operatorname{Cov}\left[f_{ij}, s_{ij}\right]\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\operatorname{Var}\left[f_{ij} - s_{ij}\right] + (\phi_{ij} - \sigma_{ij})^2\right)}, \tag{3.7}$$

where f_{ij} is a typical element of the sample shrinkage target, s_{ij} – of the covariance estimator, σ_{ij} – of the true covariance matrix, and ϕ_{ij} – of the population shrinkage target Φ. Further they prove that this optimal value is asymptotically constant over T and can be written as[3]:

$$\kappa_t = \frac{\pi_t - \rho_t}{\nu_t}. \tag{3.8}$$

In the formula above, π_t is the sum of the asymptotic variances of the entries of the sample covariance matrix scaled by $\sqrt{T}$:

$$\pi_t = \sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{AVar}\left[\sqrt{T}s_{ij,t}\right],$$

ρ_t is the sum of asymptotic covariances of the elements of the shrinkage target with the elements of the sample covariance matrix scaled by $\sqrt{T}$:

$$\rho_t = \sum_{i=1}^{N}\sum_{j=1}^{N} \operatorname{ACov}\left[\sqrt{T}f_{ij,t}, \sqrt{T}s_{ij,t}\right],$$

[3]In their paper the formula appears without the subscript t. By adding it here we want to emphasize that these variables are changing over time.

and ν_t measures the misspecification of the shrinkage target:

$$\nu_t = \sum_{i=1}^{N}\sum_{j=1}^{N}(\phi_{ij,t} - \sigma_{ij,t})^2.$$

Following their formulation and assumptions, $\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Var}\left[\sqrt{T}(f_{ij} - s_{ij})\right]$ converges to a positive limit, and so $\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Var}\left[f_{ij} - s_{ij}\right] = \mathrm{O}(1/T)$. Using this result and the $\sqrt{T}$ convergence in distribution of the elements of the sample covariance matrix, Ledoit & Wolf (2003) show that the optimal shrinkage constant is given by:

$$\alpha_t^* = \frac{1}{T}\frac{\pi_t - \rho_t}{\nu_t} + O\left(\frac{1}{T^2}\right). \tag{3.9}$$

Since α^* is unobservable, it has to be estimated. Ledoit & Wolf (2004) propose a consistent estimator of α^* for the case where the shrinkage target is a matrix in which all pairwise correlations are equal to the same constant. This constant is the average value of all pairwise correlations from the sample correlation matrix. The covariance matrix resulting from combining this correlation matrix with the sample variances, known as the equicorrelated matrix, is the shrinkage target. The equicorrelated matrix is a sensible shrinkage target as it involves only a small number of free parameters (hence less estimation noise). Thus the elements of the sample covariance matrix, which incorporate a lot of estimation error and hence can take rather extreme values are "shrunk" towards a much less noisy average. Using the equicorrelated matrix as the shrinkage target F_t in equation (3.3) the forecast is given by

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(ss)} = \mathbf{\Sigma}_t^{SS}. \tag{3.10}$$

3.2.3 A RiskMetrics™ forecast

The RiskMetrics™ forecasting methodology is a modification of the sample covariance matrix, in which observations which are further in the past are given exponentially smaller weights, determined by a factor λ. For the generic $(i,j), i,j = 1,\ldots,N$ element of the EWMA covariance matrix $\mathbf{\Sigma}_t^{RM}$ we have:

$$\sigma_{ij,t}^{RM} = (1-\lambda)\sum_{s=1}^{t}\lambda^{s-1}\left(r_{i,s} - \bar{r}_i\right)\left(r_{j,s} - \bar{r}_j\right), \tag{3.11}$$

where $\bar{r}_i = \frac{1}{t}\sum_{s=1}^{t} r_{i,s}$. Again, the forecast is given by:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(rm)} = \mathbf{\Sigma}_t^{RM}. \tag{3.12}$$

Methods to choose the optimal λ are discussed in J.P. Morgan (1996). In this paper we set $\lambda = 0.97$, the value used by J.P. Morgan for monthly (co)volatility forecasts. Note that contrary to the sample covariance matrix, for which we use a rolling window scheme, in the RiskMetrics approach we use at each t all the available observations from the beginning of the observation period up to t. Since in the RiskMetrics approach the weights decrease exponentially, the observations which are further away in the past are given relatively smaller weights and hence do not influence the estimate as much as in the sample covariance matrix.

3.2.4 A simple realized covariance forecast

The realized covariance estimator was already defined in expression (3.1). Its univariate and multivariate properties have been studied among others, by Barndorff-Nielsen & Shephard (2004) and by Andersen, Bollerslev, Diebold & Labys (2003). In the limit, when $M \to \infty$, Barndorff-Nielsen & Shephard (2004) have shown that realized covariance is an error-free measure for the integrated covariation of a very broad class of stochastic volatility models. In the empirical part we compute monthly realized covariance by using daily returns (see also French, Schwert & Stambaugh (1987)). The simple forecast is defined by:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(rc)} = \mathbf{\Sigma}_t^{RC}. \tag{3.13}$$

Thus an investor who uses this strategy simply computes the realized covariance at the end of each month and then uses it as his best guess about the true covariance matrix of the next month. A nice feature of this method is that it only uses recent information which is of most value for the forecast but imposes a very simple and restrictive time dependence. Practically equation (3.13) states that all variances and covariances follow a random walk process. However, as we shall see later, the estimated series of monthly variances and covariances show weak stationarity.

3.2.5 A shrinkage realized covariance forecast

Although the estimator discussed in the previous section is asymptotically error-free, in practice one cannot record observations continuously. A much more serious problem is the fact that at very high frequencies, the martingale assumption needed for the convergence of the realized covariances to the integrated covariation is no

longer satisfied. At trade-by-trade frequencies, market microstructure affects the price process and results in microstructure noise induced autocorrelations in returns and hence biased variance estimates. Methods to account for this bias and correct the estimates have been developed by Hansen & Lunde (2006), Oomen (2005), Aït-Sahalia et al. (2005), Bandi & Russell (2005*a*), Zhang et al. (2005), and Voev & Lunde (2007), among others. At low frequencies the impact of market microstructure noise can be significantly mitigated, but this comes at the price of higher variance of the estimator. Since we are using daily returns, market microstructure is not an issue. Thus we will suggest a possible way to reduce variance. Again as in Section 3.2.2, we will try to find a compromise between bias and variance applying the shrinkage methodology. The estimator looks very much like the one in expression (3.3). In this case we have:

$$\mathbf{\Sigma}_t^{SRC} = \hat{\alpha}_t^* F_t + (1 - \hat{\alpha}_t^*)\mathbf{\Sigma}_t^{RC}, \tag{3.14}$$

where now F_t is the equicorrelated matrix, constructed from the realized covariance matrix $\mathbf{\Sigma}_t^{RC}$ in the same fashion as the equicorrelated matrix constructed from the sample covariance matrix, as explained in Section 3.2.2. Similarly to the previous section, the forecast is simply

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(src)} = \mathbf{\Sigma}_t^{SRC}. \tag{3.15}$$

Since the realized covariance is a consistent estimator, we can still apply formula (3.7) taking into account the different rate of convergence. In order to compute the estimates for the variances and covariances, we need a theory for the distribution of the realized covariance, which is developed in Barndorff-Nielsen & Shephard (2004), who provide asymptotic distribution results for the realized covariation matrix of continuous stochastic volatility semimartingales ($\mathcal{SVSM}^c$). Assuming that the log price process $\ln \mathbf{P} \in \mathcal{SVSM}^c$, we can decompose it as $\ln \mathbf{P} = a^* + m^*$, where a^* is a process with continuous finite variation paths and m^* is a local martingale. Furthermore, under the condition that m^* is a multivariate stochastic volatility process, it can be defined as $m^*(t) = \int_0^t \Theta(u)\mathrm{d}w(u)$, where Θ is the spot covolatility process and w is a vector standard Brownian motion. Then the spot covariance is defined as:

$$\Sigma(t) = \Theta(t)\Theta(t)', \tag{3.16}$$

assuming that (for all $t < \infty$)

$$\int_0^t \Sigma_{kl}(u)\mathrm{d}u < \infty, \quad k, l = 1, \ldots, N, \tag{3.17}$$

where $\Sigma_{kl}(t)$ is the (k, l) element of the $\Sigma(t)$ process. Having laid this notation we will now interpret the "true" covariance matrix as:

$$\mathbf{\Sigma}_{t+1} = \int_t^{t+1} \Sigma(u)\mathrm{d}u. \tag{3.18}$$

Thus the covariance matrix at time $t+1$ is the increment of the integrated covariance matrix of the continuous local martingale from time t to time $t + 1$. The realized covariance as defined in expression (3.1) consistently estimates $\mathbf{\Sigma}_{t+1}$ as given in equation (3.18). Furthermore, Barndorff-Nielsen & Shephard (2004) show that under a set of regularity conditions the realized covariation matrix follows asymptotically, as $M \to \infty$, the normal law with $N \times N$ matrix of means $\int_t^{t+1} \Sigma(u)\mathrm{d}u$. The asymptotic covariance of

$$\sqrt{M}\left\{\mathbf{\Sigma}_{t+1}^{RC} - \int_t^{t+1} \Sigma(u)\mathrm{d}u\right\}$$

is Ω_{t+1}, a $N^2 \times N^2$ array with elements

$$\Omega_{t+1} = \left\{\int_t^{t+1} \left\{\Sigma_{kk'}(u)\Sigma_{ll'}(u) + \Sigma_{kl'}(u)\Sigma_{lk'}(u)\right\}\mathrm{d}u\right\}_{k,k',l,l'=1,\ldots,N}.$$

Of course, this matrix is singular due to the equality of the covariances in the integrated covariance matrix. This can easily be avoided by considering only its unique lower triangular elements, but for our purposes it will be more convenient to work with the full matrix. The result above is not useful for inference, since the matrix Ω_{t+1} is not known. Barndorff-Nielsen & Shephard (2004) show that a consistent, positive semi-definite estimator is given by a random $N^2 \times N^2$ matrix:

$$H_{t+1} = \sum_{j=1}^{M} x_{j,t+1}x'_{j,t+1} - \frac{1}{2}\sum_{j=1}^{M-1}\left(x_{j,t+1}x'_{j+1,t+1} + x_{j+1,t+1}x'_{j,t+1}\right), \tag{3.19}$$

where $x_{j,t+1} = vec\left(\mathbf{r}_{t+\frac{j}{M}}\mathbf{r}'_{t+\frac{j}{M}}\right)$ and the *vec* operator stacks the columns of a matrix into a vector. It holds that $MH_{t+1} \xrightarrow{p} \Omega_{t+1}$ with $M \to \infty$.
With the knowledge of this matrix, we can combine the asymptotic results for the

realized covariance, with the result in equation (3.7) to compute the estimates for π_t, ρ_t and ν_t.

For the equicorrelated matrix F we have that[4] $f_{ij} = \bar{r}\sqrt{\sigma_{ii}^{(RC)}\sigma_{jj}^{(RC)}}$, where $\bar{r}$ is the average value of all pairwise correlations, implied by the realized covariance matrix, and $\sigma_{ij}^{(RC)}$ is the (i,j) element of the realized covariance matrix. Thus Φ, the population equicorrelated matrix, has a typical element $\phi_{ij} = \bar{\varrho}\sqrt{\sigma_{ii}\sigma_{jj}}$, where σ_{ij} is the (i,j) of the true covariance matrix Σ and $\bar{\varrho}$ is the average correlation implied by it. Substituting $\sigma_{ij}^{(RC)}$ for s_{ij} in equation (3.7) and multiplying by M gives for the optimal shrinkage intensity:

$$M\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathrm{Var}\left[\sqrt{M}\sigma_{ij}^{(RC)}\right] - \mathrm{Cov}\left[\sqrt{M}f_{ij}, \sqrt{M}\sigma_{ij}^{(RC)}\right]\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathrm{Var}\left[f_{ij} - \sigma_{ij}^{(RC)}\right] + (\phi_{ij} - \sigma_{ij})^2\right)}. \tag{3.20}$$

Note that this equation resembles expression (3.8). The only difference is the scaling by $\sqrt{M}$ instead of $\sqrt{T}$, which is due to the $\sqrt{M}$ convergence. In this case π_t, the first summand in the numerator, is simply the sum of all diagonal elements of Ω_t. By using the definition of the equicorrelated matrix, it can be shown that the second term, ρ_t, can be written as (suppressing the index t):

$$\rho = \sum_{i=1}^{N}\mathrm{AVar}\left[\sqrt{M}\sigma_{ii}^{(RC)}\right] + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\mathrm{ACov}\left[\sqrt{M}\bar{r}\sqrt{\sigma_{ii}^{(RC)}\sigma_{jj}^{(RC)}}, \sqrt{M}\sigma_{ij}^{(RC)}\right]. \tag{3.21}$$

Applying the delta method the second term can be expressed as[5]

$$\frac{\bar{r}}{2}\left(\sqrt{\frac{\sigma_{jj}^{(RC)}}{\sigma_{ii}^{(RC)}}}\mathrm{ACov}\left[\sqrt{M}\sigma_{ii}^{(RC)}, \sqrt{M}\sigma_{ij}^{(RC)}\right] + \sqrt{\frac{\sigma_{ii}^{(RC)}}{\sigma_{jj}^{(RC)}}}\mathrm{ACov}\left[\sqrt{M}\sigma_{jj}^{(RC)}, \sqrt{M}\sigma_{ij}^{(RC)}\right]\right).$$

From this expression we see that ρ also involves summing properly scaled terms of the Ω matrix. In the denominator of equation (3.20), the first term is of order $\mathrm{O}(1/M)$, and the second one is consistently estimated by $\hat{\nu} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(f_{ij} - \sigma_{ij}^{(RC)}\right)^2$.

Since we have a consistent estimator for Ω, we can now also estimate π and ρ. In particular, we have

[4] In the following exposition, the time index is suppressed for notational convenience.

[5] cf. Ledoit & Wolf (2004)

$$\begin{aligned}\hat{\pi} &= \sum_{i=1}^{N}\sum_{j=1}^{N} h_{ij,ij} \\ \hat{\rho} &= \sum_{i=1}^{N} h_{ii,ii} + \frac{\bar{r}}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{\frac{\sigma_{jj}^{(RC)}}{\sigma_{ii}^{(RC)}}}h_{ii,ij} + \sqrt{\frac{\sigma_{ii}^{(RC)}}{\sigma_{jj}^{(RC)}}}h_{jj,ij},\end{aligned}$$

where $h_{kl,k'l'}$ is the element of H which estimates the corresponding element of Ω. Thus we can estimate κ_t by $\hat{\kappa}_t = \frac{\hat{\pi}_t - \hat{\rho}_t}{\hat{\gamma}_t}$ and the estimator for the optimal shrinkage constant is:

$$\hat{\alpha}_t^* = \max\left\{0, \min\left\{\frac{\hat{\kappa}_t}{M}, 1\right\}\right\}. \tag{3.22}$$

The estimated optimal shrinkage constants for our dataset range from 0.0205 to 0.2494 with a mean of 0.0562.

3.2.6 Dynamic realized covariance forecasts

This model is an alternative to the one in Section 3.2.4. The most popular models for time varying variances and covariances are the GARCH models. The most significant problem of these models is the large number of parameters in large dimensional systems. The recent DCC models of Tse & Tsui (2002) and Engle (2002) propose a way to mitigate this problem by using the restriction that all correlations obey the same dynamics. Recently Gourieroux et al. (2004) have suggested an interesting alternative – the WAR (Wishart autoregressive) model, which has certain advantages over the GARCH models, e.g. smaller number of parameters, easy construction of non-linear forecasts, simple verification of stationarity conditions, etc. Even quite parsimonious models, however, have a number of parameters of the order $N(N+1)/2$. With $N = 15$ this means more than 120 parameters, which would be infeasible for estimation. We therefore suggest a simple approach in which all variance and covariance series are modelled univariately as ARMA processes and individual forecasts are made, which are then combined into a forecast of the whole matrix. This approach can also be extended by including lags of squared returns which can be interpreted as a kind of ARCH-terms. A theoretical drawback of this model, is that such a methodology does not guarantee the positive definiteness of the forecast matrix. It turns out that this problem could be quite severe, especially if we include functions of lagged returns in the specification. Hence we propose two possible solutions. First, if the above mentioned problem occurs relatively rarely, then in these cases we can define the forecast as in Section 3.2.4, which would ensure

that all forecast matrices are positive definite. More precisely, instead of assuming a random walk process for the realized covariance series (as in Section 3.2.4) we now model each of them as ARMAX$(p, q, 1)$[6] processes as follows:

$$\sigma_{ij,t}^{(RC)} = \omega + \sum_{s=1}^{p} \varphi_s \sigma_{ij,t-s}^{(RC)} + \sum_{u=0}^{q} \theta_u \varepsilon_{ij,t-u} + \alpha r_{i,t-1} r_{j,t-1}, \tag{3.23}$$

with $\theta_0 = 1$ and $\varepsilon_{ij,t}$ – a Gaussian white noise process. The model easily extends to an ARMAX(p, q, k) specification with k lags of crossproducts. The parameters φ_s, θ_u and α are estimated by maximum likelihood starting at $t = 100$ and the forecasts $\hat{\sigma}_{ij,t+1|t}^{(RC)}$ are collected in a matrix $\mathbf{\Sigma}_{t+1}^{DRC}$. At time $t+1$ the new information is taken into account and the procedure is repeated. The best model for each series is selected by minimizing the Akaike information criterion (AIC).
In this case the forecast is:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(drc)} = \begin{cases} \mathbf{\Sigma}_{t+1}^{DRC}, & \text{if } \mathbf{\Sigma}_{t+1}^{DRC} \text{ is positive definite} \\ \mathbf{\Sigma}_{t}^{RC}, & \text{otherwise.} \end{cases} \tag{3.24}$$

A more robust solution is to factorize the sequence of realized covariance matrices into their Cholesky decompositions, model the dynamics and forecast the Cholesky series and then reconstruct the variance and covariance forecasts. This ensures the positive definiteness of the resulting forecast. In this case the Cholesky series are modelled like in equation (3.23), the forecasts are collected in a lower triangular matrix $\mathbf{C}_{t+1}$ and the covariance forecast is given by:

$$\hat{\mathbf{\Sigma}}_{t+1|t}^{(drc-Chol)} = \mathbf{C}_{t+1}\mathbf{C}_{t+1}'. \tag{3.25}$$

Analogously, we can use these two strategies to model dynamically the series of shrunk variance covariance matrices which defines the forecasts $\mathbf{\Sigma}_{t+1|t}^{(dsrc)}$ and $\mathbf{\Sigma}_{t+1|t}^{(dsrc-Chol)}$.

3.3 Data

The data we have used consists of 15 stocks from the current composition of the Dow Jones Industrial Average index from 01.01.1980 to 31.12.2002. The stocks are Alcoa (NYSE ticker symbol: AA), American Express Company (AXP), Boeing Company (BA), Caterpillar Inc. (CAT), Coca-Cola Company (KO), Eastman Kodak (EK), General Electric Company (GE), General Motors Corporation (GM), Hewlett-Packard Company (HPQ), International Business Machines (IBM), McDonald's Corporation (MCD), Philip Morris Companies Incorporated (MO), Procter &

[6]The last parameter shows the number of lags of the X variable.

Gamble (PG), United Technologies Corporation (UTX) and Walt Disney Company (DIS). The reason that we have considered only 15 stocks is due to fact that the realized covariance matrices are of full rank only if $M > N$, where M is the number of intra-period observations used to construct the realized covariance, in our case number of daily returns used to construct each monthly realized covariance. Usually there are 21 trading days per month, but some months have had fewer trading days (e.g. September 2001). With intradaily data this problem would not be of importance, since then we can easily have hundreds of observations within a day. Such datasets are already common, but they still do not cover large periods of time. Nevertheless, the dynamic properties of daily realized volatilities, covariances and correlations are studied by e.g. Andersen, Bollerslev, Diebold & Ebens (2001) and Andersen, Bollerslev, Diebold & Labys (2001). It has been shown that there is a long-range persistence, which allows for construction of good forecasts by means of ARFIMA processes.

All the stocks are traded on the NYSE and we take the daily closing prices and monthly closing prices to construct corresponding returns. The data is adjusted for splits and dividends. We find the typical properties of financial returns: negative skewness (with the exception of PG), leptokurtosis and non-normality. The average (across stocks) mean daily return is 0.05% and the average daily standard deviation is 1.9 %. From the daily data log monthly returns are constructed by using the opening price of the first trading day of the month and the closing price of the last day. These returns are then used to construct rolling window sample covariance matrices, used in the first two forecasting models.

3.4 Results

In this section we present and discuss the results on the performance of the forecasting models described in Section 3.2.

In order to asses the forecasting performance, we employ Diebold-Mariano tests for each of the variance and covariance series. Then we measure the deviation of the forecast as a matrix from its target by using again the Frobenius norm, which gives an overall idea of the comparative performance of the models. Of course, if the individual series are well forecast, so will be the matrix. As a target or "true" covariance matrix, we choose the realized covariance matrix. First, we present some graphical results. Out of the total of 120 variance and covariance forecast series, Figure 3.1 plots 9 representative cases, for the sample covariance and the RiskMetrics™ model, against the realized series. The name, which appears above each block in the figure, represents either a variance series (e.g. EK), or a covariance

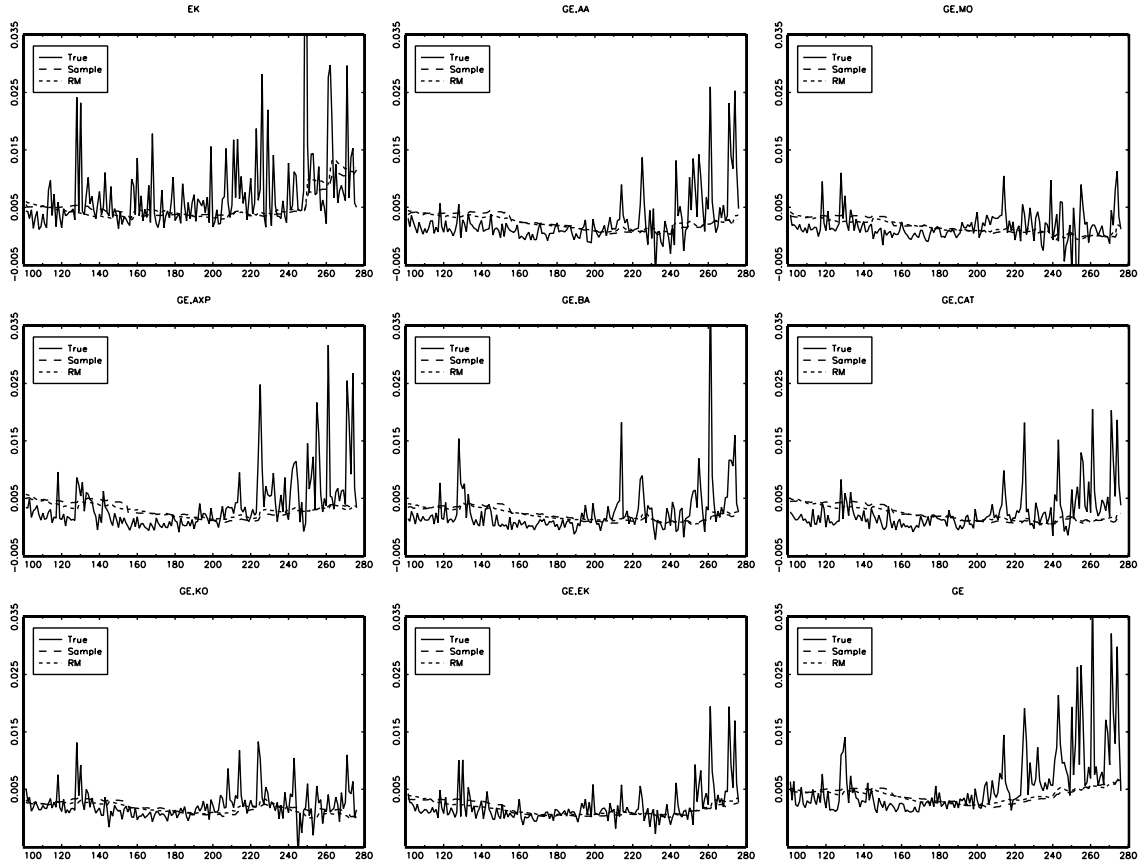

Figure 3.1: Comparison of the sample covariance based (Sample) and Riskmetrics™ (RM) forecast against the realized covariance (True).

one (e.g. GE,AA).

Both forecasts are quite close, and as can be seen, they cannot account properly for the variation in the series. As the tests show, however, the Riskmetrics™ fares better and is the best model among the sample based ones. It is already an acknowledged fact that financial returns have the property of volatility clustering. This feature is also clearly evident in the figure, where periods of low and high volatility can be easily distinguished, which suggests that variances and covariances tend to exhibit positive autocorrelation. Figure 3.2 shows the autocorrelation functions for the same 9 series of realized (co)variances.

The figure clearly shows that there is some positive serial dependence, which usually dies out quickly, suggesting stationarity of the series. Stationarity is also confirmed by running Augmented Dickey-Fuller (ADF) tests, which reject the presence of an unit root in all series at the 1% significance level.

The observed dependence patterns suggest the idea of modelling the variance and covariance series as well as their shrunk versions as ARMA processes. This resulted in a few cases in which the matrix forecast was not positive definite (16 out of 176

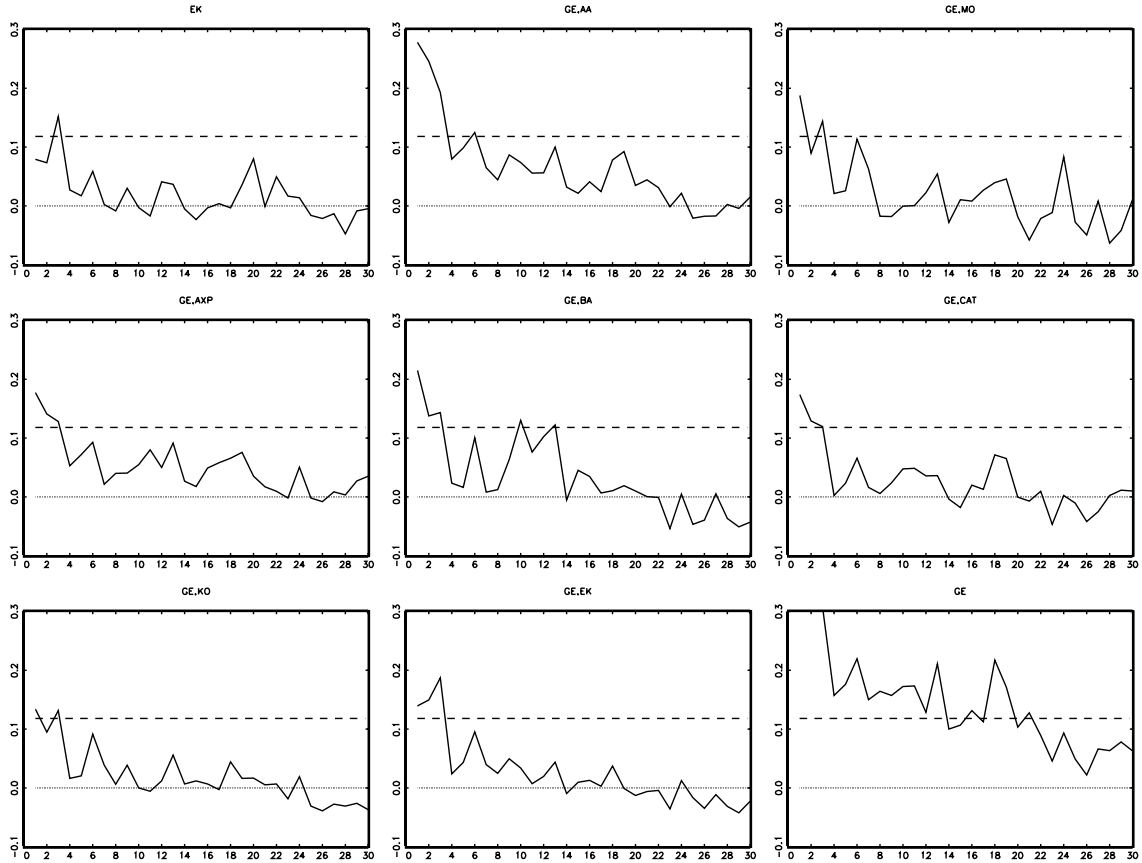

Figure 3.2: Autocorrelation functions of the realized variance and covariance series. The dashed line represents the upper 95% confidence bound.

for the original series and 8 out of 176 for the shrunk series). Thus the forecast in expression (3.24) seems to be reasonable and as we shall see later, compares well to the sample covariance based models. In a GARCH framework, the conditional variance equation includes not only lags of the variance, but also lags of squared innovations (shocks). When mean returns are themselves unpredictable (the usual approach is to model the mean equation as an ARMA process), the shock is simply the return. This fact led us to include lags of squared returns (for the variance series) and cross-products (for the covariance series) as in the ARMAX$(p, q, 1)$ model in equation (3.23). This added flexibility, however, comes at the price of a drastic increase of the non-positive definite forecasts (108 and 96 out of 176, respectively). Thus the forecast in equation (3.24) comes quite close to the simple realized and shrunk realized covariance models in Sections 3.2.4 and 3.2.5, respectively. A solution to this issue is to decompose the matrices into their lower triangular Cholesky factors, forecast the Cholesky series, and then reconstruct the matrix. This leads to the forecasting formula in equation (3.25), which defines the $drc - Chol$ and $dsrc - Chol$ forecasting models for the simple realized and shrunk realized covari-

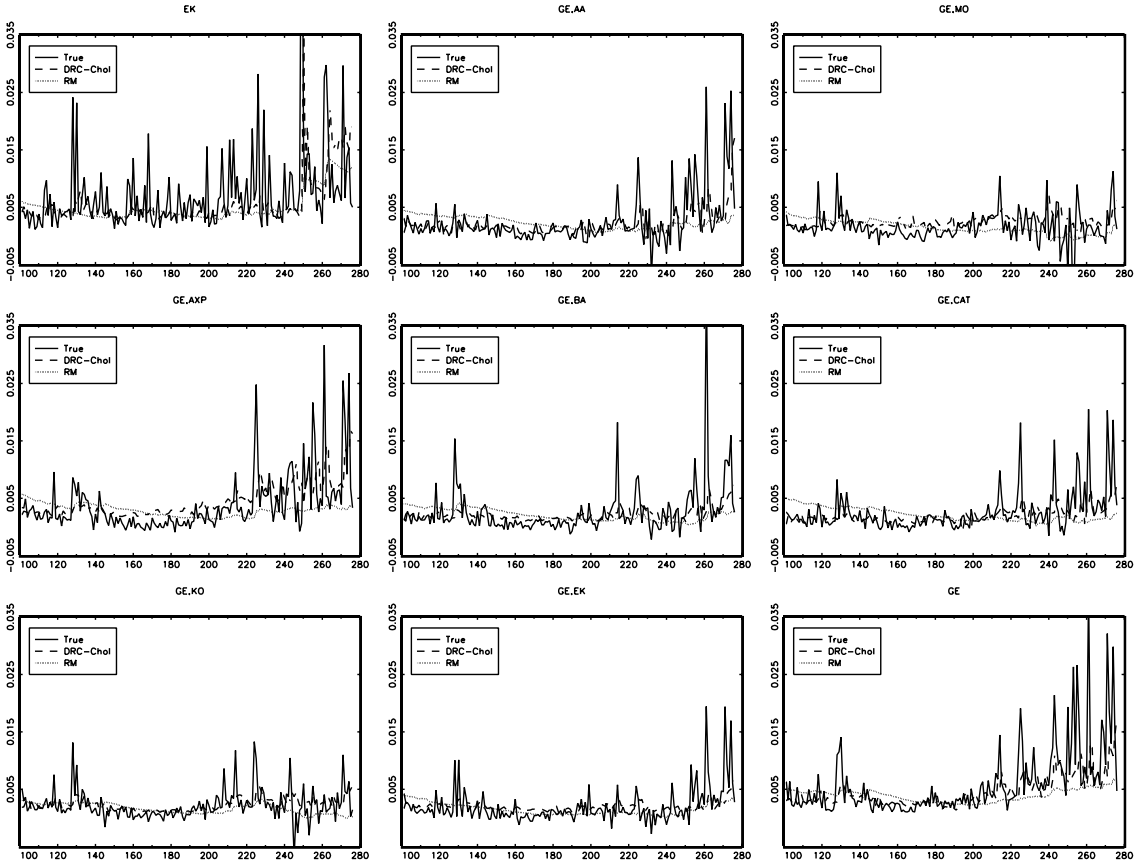

Figure 3.3: Comparison of the Riskmetrics™ forecast (RM) and the dynamic realized covariance forecast based on Cholesky series (DRC-Chol) against the realized covariance (True).

ance case, respectively. A drawback of this approach is that the Cholesky series do not have an intuitive interpretation. They are simply used as a tool to constrain the forecasts to satisfy the complicated restrictions implied by the positive definiteness requirement. Another drawback is that the Cholesky decomposition involves non-linear transformations of the original series. Thus, if one can adequately forecast the nonlinear transformation, this does not immediately mean that applying the inverse transformation to the forecast will produce a good forecast of the initial series. So there is a trade-off between the possibility of including more information in the forecast and obtaining positive definite matrices on the one hand, and the distortions caused by the non-linearity of the transformation on the other. It turns out that in our case the beneficial effects outweigh the negative ones. Figure 3.3 shows the $drc - Chol$ and the RiskMetrics™ forecast for the same 9 variance and covariance series.

From the figure it is evident that the dynamic forecasts track the true series much closer than the RiskMetrics™ forecasts, especially at the end of the period when the (co)volatilities were more volatile. The $dsrc - Chol$ forecast looks quite similar to

the $drc - Chol$ (due to the usually small shrinkage constants), but as we shall see later the forecasts are in fact somewhat better.

Turning to the statistical comparison of the forecasting methods, we first briefly present the Diebold-Mariano testing framework as in Harvey, Leyborne & Newbold (1997). Suppose a pair of l-step ahead forecasts h_1 and h_2, $h_1, h_2 \in H$ have produced errors (e_{1t}, e_{2t}), $t = 1, \ldots, T$. The null hypothesis of equality of forecasts is based on some function $g(e)$ of the forecast errors and has the form $\mathrm{E}\left[g(e_{1t}) - g(e_{2t})\right] = 0$. Defining the loss differential $d_t = g(e_{1t}) - g(e_{2t})$ and its average $\bar{d} = T^{-1} \sum_{t=1}^{T} d_t$, the authors note that "the series d_t is likely to be autocorrelated. Indeed, for optimal l-steps ahead forecasts, the sequence of forecast errors follows a moving average process of order $(l-1)$. Thus result can be expected to hold approximately for any reasonably well-conceived set of forecasts." Consequently, it can be shown that the variance of $\bar{d}$ is, asymptotically,

$$\mathrm{Var}\left[\bar{d}\right] \approx T^{-1} \left[\gamma_0 + 2 \sum_{k=1}^{l-1} \gamma_k\right], \tag{3.26}$$

where γ_k is the k^{th} autocovariance of d_t. The Diebold-Mariano test statistic is:

$$S_1 = \left[\widehat{\mathrm{Var}}\left[\bar{d}\right]\right]^{-1/2} \bar{d}, \tag{3.27}$$

where $\widehat{\mathrm{Var}}\left[\bar{d}\right]$ is obtained from equation (3.26) by substituting for γ_0 and γ_k the sample variance and autocovariances of d_t, respectively. Tests are then based on the asymptotic normality of the test statistic. Noting that we only consider 1-step ahead forecasts in this paper, the series d_t should not be autocorrelated. As already noted above, this is expected to hold for any *reasonably* constructed forecasts. Actually, however, the sample based forecasts are not really *reasonable* in the sense that they do not account for the serial dependence of the process they are supposed to forecast. Thus, the degree of autocorrelation in the d_t series, when either h_1 or h_2 is a sample based forecast, will correspond to the degree of dependence in the series to be forecast. For this reason, ignoring autocovariances in the construction of the Diebold-Mariano tests will lead to an error in the test statistic. To correct for this we include in $\widehat{\mathrm{Var}}\left[\bar{d}\right]$ the first k significant autocorrelations for each of the 120 series. Table 3.1 summarizes the results of the Diebold-Mariano tests carried out pairwise between all models for all 120 series. The first entry in each cell of the table shows the number of series (out of 120) for which the model in the corresponding column outperforms the model in the corresponding row. The second entry corresponds to the number of significant outperformances according to the Diebold-Mariano tests at the 5% significance level. Hence, the table is in a sense symmetric, as the number of times model h_1 outperforms model h_2 plus the number of times model h_2 outperforms

model h_1 (given by the first number in each cell) sum up to 120 – the total number of series. This is not the case, only for the pairs highlighted in bold, because the 15 variance series are unchanged in their respective "shrunk" versions.[7] Thus, in these cases there are only 105 covariance series forecasts to be compared.

Table 3.1: Results from the Diebold-Mariano tests. Due to the definition of the shrinkage target, the first numbers in the pairs highlighted in bold do not sum up to 120, since the variance series are unchanged in their respective "shrunk" versions. Thus, in these cases there are only 105 series forecasts to be compared.

	s	*ss*	*rm*	*rc*	*src*	*drc*	*dsrc*	*drc–Chol*	*dsrc–Chol*
s	-	**85/28**	106/50	14/1	16/1	47/20	89/37	93/49	100/55
ss	**20/0**	-	106/47	14/1	16/1	47/20	89/37	92/49	100/55
rm	14/0	14/0	-	7/1	11/1	37/7	73/29	85/33	89/37
rc	106/60	106/61	113/69	-	**105/86**	119/59	120/88	115/80	117/88
src	104/55	104/56	109/69	**0/0**	-	119/50	120/86	114/77	117/85
drc	73/12	73/12	83/26	1/0	1/0	-	**104/31**	98/47	103/58
dscr	31/3	31/3	47/8	0/0	0/0	**1/0**	-	69/28	83/35
drc (Chol)	27/8	28/8	35/10	5/1	6/1	22/7	51/12	-	91/19
dsrc (Chol)	20/7	20/7	31/8	3/1	3/1	17/6	37/11	29/3	-

At first glance one can notice that the worst performing models are the *rc* and *src* models. Among the sample based forecasts the RiskMetrics™ is the one which delivers the best performances. The comparison between the sample and the shrinkage sample forecasts shows that shrinking has indeed improved upon the sample covariance matrix. This holds also for the realized covariance matrix. Here, the result is reinforced by the fact that shrinking also increases the probability of obtaining a positive definite forecast. In fact, the quite poor performance of the *drc* model is not due to the poor forecasting of the series themselves, but due to the large error, introduced by taking the previous realized covariance matrix, in case of a non-positive definite forecast (see equation (3.24)). Even though this only happens in 16 out of 176 cases, it is enough to distort the forecast considerably. The main result of this paper, however, arises from the comparison of the dynamic models with the sample based ones, which can be drawn by considering the last three columns of the table. For most of the series the dynamic models provide better forecasts, which results in smaller errors in the covariance matrix forecasts, as will be shown later. Despite the fact that the number of significant outperformances is not strikingly high (due to the small number of periods for evaluation), it is still clear that the dynamic models outperform decisively even the best model among the sample based ones. Furthermore, as noted earlier, the forecasts using the Cholesky decomposition appear to be better compared to those which model the variance and covariance series directly.

[7] By shrinking towards the equicorrelated matrix, the variances do not change.

This result comes mainly as a consequence of the considerable explanatory power of the lagged shocks in addition to the lagged (co)variances, which could not have been utilized had not we assured the positive definiteness of the forecasts.
In order to understand better the benefits from modelling the variance and covariance series dynamically, we shall consider an alternative (but closely related) measure of forecasting error. In section 3.2.2 it was shown how the Frobenuis norm can be used as a measure of distance between two matrices. Here we will utilize this concept again by considering the following definition of the forecast error in terms of a matrix forecast:

$$e_t^{(h)} = \left\| \hat{\mathbf{\Sigma}}_{t|t-1}^{(h)} - \mathbf{\Sigma}_t^{RC} \right\|^2, \quad h \in H. \tag{3.28}$$

The root mean squared prediction errors (RMSPE) are collected in Table 3.2.

Table 3.2: Root mean squared prediction errors and corresponding ranks of the forecasting models based on the Frobenius norm.

Model	RMSPE	Relative rank
s	0.06021	7
ss	0.06016	6
rm	0.05887	4
rc	0.06835	9
src	0.06766	8
drc	0.06004	5
$dscr$	0.05749	1
drc (Chol)	0.05854	3
$dsrc$ (Chol)	0.05799	2

The ranking of the models according to this table is quite similar to the one following from Table 3.1. The only difference is that now the $dsrc$ model appears to be somewhat better than the $dsrc - Chol$, which is most probably due to chance, since as we saw earlier the latter model forecasts most of the series better. As a conclusion, we can state again that in general, the dynamic models outperform the sample covariance based ones.

3.5 Conclusion

Volatility forecasting is crucial for portfolio management, option pricing and other fields of financial economics. Starting with Engle (1982) a new class of econometric models was developed to account for the typical characteristics of financial returns volatility. This class of models grew rapidly and numerous extensions were proposed.

In the late 1980's these models were extended to handle not only volatilities, but also covariance matrices. The main practical problem of these models is the large number of parameters to be estimated, if one decides to include more than a few assets in the specification. Partial solutions to this "curse of dimensionality" were proposed, which imposes restrictions on the system dynamics. Still, modelling and forecasting return covariance matrices remains a challenge. This paper proposes a methodology which is more flexible than the traditional sample covariance based models and at the same time is capable of handling a large number of assets. Although conceptually this methodology is more elaborate than the above mentioned traditional models, it is easily applicable in practice and actually requires shorter historical samples, but with a higher frequency. The gains come from the fact that with high-frequency observations, the latent volatility comes close to being observable. This enables the construction of realized variance and covariance series, which can be modelled and forecast on the basis of their dynamic properties. Additionally, we show that shrinking, which has been shown to improve upon the sample covariance matrix, can also be helpful in reducing the error in the realized covariance matrices. A practical drawback which appears in this framework is that the so constructed forecasts are not always positive definite. One possible solution to this is to use the Cholesky decomposition as a method of incorporating the positive definiteness requirement in the forecast.

The paper shows that on the monthly frequency, this approach produces better forecasts based on results from Diebold-Mariano tests. The possible gains from a better forecast are, e.g., construction of mean-variance efficient portfolios. Providing a more accurate forecast of future asset comovements will result in better balanced portfolios. These gains will be most probably higher and more pronounced if intradaily returns are used for the construction of daily realized covariance matrices, which remains a possible avenue for further research. It has been shown (e.g. by Andersen, Bollerslev, Diebold & Ebens (2001)) that realized daily volatilities and correlations exhibit high persistence. Since by incorporating intra-daily information these realized measures are also quite precise, this serial dependence can be exploited for volatility forecasting. A possible extension of the methodological framework suggested in the paper could be modelling the realized series in a vector ARMA system, in order to analyze volatility spillovers across stocks, industries or markets, which however would again involve a large number of parameters.

A closely related area of research is concerned with the methods for evaluation of covariance matrix forecasts. In this paper we have used purely statistical evaluation tools based on a symmetric loss function. An asymmetric measure in this case may have more economic meaning, since it is quite plausible to assume that if a portfolio

variance has been overestimated, the consequences are less adverse than if it has been underestimated. In a multivariate context Byström (2002) uses as an evaluation measure of forecasting performance the profits generated by a simulated trading of portfolio of rainbow options. The prices of such options depend on the correlation between the underlying assets. Thus the agents who forecast the correlations more precisely should have higher profits on average.

Further, the models presented in this paper can be extended by introducing the possibility of asymmetric reaction of (co)volatilities to previous shocks (leverage). This can be achieved by introducing some kind of asymmetry in equation (3.23), e.g., by including products of absolute shocks or products of indicator functions for positivity of the shocks.

Bibliography

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *Review of Financial Studies* **18**(2), 351–416.

Andersen, T., Bollerslev, T., Christoffersen, P. F. & Diebold, F. X. (2006), Volatility forecasting, *in* G. Elliott, C. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting', 1 edn, Elsevier, chapter 15.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Bandi, F. M. & Russell, J. R. (2005), Microstructure noise, realized volatility, and optimal sampling. Working paper, Graduate School of Business, The University of Chicago.

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

Bauwens, L., Laurent, S. & Rombouts, J. (2006), 'Multivariate garch models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Black, F. & Litterman, R. (1992), 'Global portfolio optimization', *Financial Analysts Journal* **48**(5), 28–43.

Byström, H. (2002), 'Using simulated currency rainbow options to evaluate covariance matrix forecasts', *Journal of International Financial Markets, Institutions and Money* **12**, 216–230.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**(3), 253–263.

Engle, R. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models', *Journal of Business and Economic Statistics* **20**, 339–350.

French, K. R., Schwert, G. W. & Stambaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 3–29.

Gourieroux, C., Jasiak, J. & Sufana, R. (2004), The wishart autoregressive process of multivariate stochastic volatility. Working Paper, University of Toronto.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–218.

Harvey, D., Leyborne, S. & Newbold, P. (1997), 'Testing the equality of prediction mean squared errors', *International Journal of Forecasting* **13**, 281–291.

Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit, O. & Wolf, M. (2004), 'Honey, i shrunk the sample covariance matrix', *Journal of Portfolio Management* **31**, 110–119.

Michaud, R. O. (1989), 'The markowitz optimization enigma: Is 'optimized' optimal?', *Financial Analysts Journal* **45**(1), 31–42.

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

Tse, Y. & Tsui, A. (2002), 'A multivariate generalized auto-regressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business and Economic Statistics* **20**, 351–362.

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

Zhang, L., Mykland, P. A. & 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.

Complete Bibliography

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *Review of Financial Studies* **18**(2), 351–416.

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2006), Ultra high frequency volatility estimation with dependent microstructure noise. Working Paper, Princeton University.

Andersen, T., Bollerslev, T., Christoffersen, P. F. & Diebold, F. X. (2006), Volatility forecasting, *in* G. Elliott, C. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting', 1 edn, Elsevier, chapter 15.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Anderson, T. W. (2003), *An introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, New Jersey.

Bandi, F. M. & Russell, J. R. (2005*a*), Microstructure noise, realized volatility, and optimal sampling. Working paper, Graduate School of Business, The University of Chicago.

Bandi, F. M. & Russell, J. R. (2005*b*), Realized covariation, realized beta, and microstructure noise. Working paper, Graduate School of Business, The University of Chicago.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Estimating quadratic variation using realised variance', *Journal of Applied Econometrics* **17**, 457–477.

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

Barndorff-Nielsen, O., Hansen, P., Lunde, A. & Shephard, N. (2006), Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Working paper, Nuffield College, Oxford.

Bauwens, L., Laurent, S. & Rombouts, J. (2006), 'Multivariate garch models: a survey', *Journal of Applied Econometrics* **21**, 79–109.

Black, F. & Litterman, R. (1992), 'Global portfolio optimization', *Financial Analysts Journal* **48**(5), 28–43.

Byström, H. (2002), 'Using simulated currency rainbow options to evaluate covariance matrix forecasts', *Journal of International Financial Markets, Institutions and Money* **12**, 216–230.

Chordia, T., Roll, R. & Subrahmanyam, A. (2005), 'Evidence on the speed of convergence to market efficiency', *Journal of Financial Economics* **76**, 271–292.

Christensen, K. & Podolskij, M. (2007), 'Realized range-based estimation of integrated variance', *Journal of Econometrics* **141**, 323–349.

Corsi, F. & Audrino, F. (2007), Realized correlation tick-by-tick. Working paper, University of Lugano.

Curci, G. & Corsi, F. (2006), Discrete sine transform for multi-scales realized volatility measures. Working Paper, University of Lugano.

de Pooter, M., Martens, M. & van Dijk, D. (2006), Predicting the daily covariance matrix for S&P 100 stocks using intraday data - but which frequency to use?. Erasmus University Rotterdam.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**(3), 253–263.

Engle, R. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models', *Journal of Business and Economic Statistics* **20**, 339–350.

Epps, T. (1979), 'Comovements in stock prices in the very short run', *Journal of the American Statistical Association* **74**, 291–298.

French, K. R., Schwert, G. W. & Stambaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 3–29.

Gourieroux, C., Jasiak, J. & Sufana, R. (2004), The wishart autoregressive process of multivariate stochastic volatility. Working Paper, University of Toronto.

Griffin, J. E. & Oomen, R. C. A. (2006), Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Working Paper, University of Warwick.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–218.

Harris, F., McInish, T., Shoesmith, G. & Wood, R. (1995), 'Cointegration, error correction and price discovery on informationally-linked security markets', *Journal of Financial and Quantitative Analysis* **30**, 563–581.

Harvey, D., Leyborne, S. & Newbold, P. (1997), 'Testing the equality of prediction mean squared errors', *International Journal of Forecasting* **13**, 281–291.

Hayashi, T. & Kusuoka, S. (2004), Nonsynchronous covariation measurement for continuous semimartingales. Preprint Series of the Graduate School of Mathematical Sciences, The University of Tokyo.

Hayashi, T. & Yoshida, N. (2004), 'On covariance estimation for high-frequency financial data', *Proceeding of IASTED/Financial Engineering and Applications 2004* **437-801**, 282–286.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), 'Flexible multivariate GARCH modeling with an application to international stock markets', *Review of Economics and Statistics* **85**, 735–474.

Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit, O. & Wolf, M. (2004), 'Honey, i shrunk the sample covariance matrix', *Journal of Portfolio Management* **31**, 110–119.

Martens, M. (2004), Estimating unbiased and precise realized covariances. Econometric Institute, Erasmus University Rotterdam.

Meddahi, N. (2002), 'A theoretical comparison between integrated and realized volatility', *Journal of Applied Econometrics* **17**, 479–508.

Michaud, R. O. (1989), 'The markowitz optimization enigma: Is 'optimized' optimal?', *Financial Analysts Journal* **45**(1), 31–42.

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

Phillips, P. C. & Yu, J. (2006), 'Comment on "realized variance and market microstructure noise" by peter r. hansen and asger lunde', *Journal of Business and Economic Statistics* **24**, 202–208.

Renò, R. (2001), A closer look at the Epps effect. Università degli Studi di Siena, Working paper n. 335.

Roll, R. (1984), 'A simple implicit measure of the effective bid-ask spread in an efficient market', *Journal of Finance* **39**(4), 1127–1140.

Sheppard, K. (2005), Realized covariance and scrambling. Working paper, University of Oxford.

Tse, Y. & Tsui, A. (2002), 'A multivariate generalized auto-regressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business and Economic Statistics* **20**, 351–362.

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

Zhang, L. (2006*a*), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* **12**, 1019–1043.

Zhang, L. (2006*b*), Estimating covariation: Epps effect and microstructure noise. Working Paper.

Zhang, L., Mykland, P. A. & 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.

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema

Three Essays on Estimation and Dynamic Modelling of Multivariate Market Risks using High Frequency Financial Data

ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt.[8] Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Konstanz, den 2. Januar 2008

(Valeri Voev)

[8] Siehe hierzu die Abgrenzung zu Kapiteln 1 und 2 auf der folgenden Seite.

Abgrenzung

Ich versichere hiermit, dass ich Kapitel 3 der vorliegenden Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.
Kapitel 1 entstammt einer gemeinsamen Arbeit mit Herrn Prof. Asger Lunde (Aarhus School of Business). Meine individuelle Leistung bei der Erstellung dieser Arbeit ist 90%.
Kapitel 2 entstammt einer gemeinsamen Arbeit mit Herrn Ingmar Nolte (Universität Konstanz). Meine individuelle Leistung bei der Erstellung dieser Arbeit ist 50%.