Wavelet-based Realized Variation and Covariation Theory

Jozef Baruník

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Supervisor:

prof. Ing. Miloslav Vošvrda, CSc.

To my family...

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 24, 2011

Signature

English Abstract

The study of volatility and covariation has become one of the most active and successful areas of research in time series econometrics and economic forecasting in recent decades. This dissertation contains a complete theory for realized variation and covariation estimation, generalizing current knowledge and taking the estimation into the time-frequency domain for the first time. The first part of the theory presents a wavelet-based realized variation theory, while the second part introduces its multivariate counterpart, a wavelet-based realized covariation theory. The results generalize the popular realized volatility framework by bringing robustness to noise as well as jumps and the ability to measure realized variation and covariation not only in the time domain, but also in the frequency domain. The theory is also tested in a numerical study of the small sample performance of the estimators and compared to other popular realized variation estimators under different simulation settings with changing noise as well as jump level. The results reveal that our wavelet-based theory is able to estimate the realized measures with the greatest precision. Another notable contribution lies in the application of the presented theory. Our time-frequency estimators not only produce more efficient estimates, but also decompose the realized variation and covariation into arbitrarily chosen investment horizons. The results thus provide a better understanding of the dynamics of stock markets. In the last part, the theory is also used to build a long memory forecasting model based on the decomposed measures. Wavelet-based estimators carry the highest information content for the volatility, covariation and correlations forecasts when compared to other estimators. Moreover, decomposition of realized covariation to its continuous part, individual jumps and co-jumps improves the covariance forecasts significantly.

Czech Abstract

Kvadratická variace a kovariace se během několika posledních desetiletí staly jedněmi z nejfrekventovanějších, ale také nejúspěšněji studovaných témat ekonometrie časových řad. Tato disertace obsahuje kompletní teorii odhadu realizované variace a kovariace. Tato teorie je zobecněním současného stavu poznání v dané oblasti, přičemž vlastním přínosem je odhad veličin v časově frekvenční doméně. Zatímco první část teorie je věnována jednorozměrnému odhadu realizované variace pomocí waveletů, druhá část přináší vícerozměrný protějšek této teorie: odhad realizované kovariace pomocí waveletů. Konkrétními přínosy k již známým přístupům k realizované variaci je jednak robustifikace šumu, který nově nemusí být nutně Gaussovský a může obsahovat skoky, dále pak možnost měřit realizovanou variaci a kovariaci nejen v časové, ale i frekvenční doméně. Teorie je ověřena pomocí numerické studie zkoumající výkonnost z ní odvozených odhadů na malých vzorcích a srovnávající tyto odhady s ostatními užívanými odhady realizované variace, přičemž odhady jsou srovnány pomocí simulace při různých úrovních šumu a velikosti skoků. Výsledky studie ukazují, že tyto nové odhady dosahují nejlepších výsledků a jsou tedy dobře použitelné pro odhad realizované volatility (druhé odmocniny realizované variace). Za zmínku stojí ještě jedna aplikace v práci dosažených výsledků: rozklad realizované variace a kovariace podle libovolně zvolených investičních horizontů - výsledky práce tedy mohou kromě přesnějšího odhadování sloužit i k lepšímu porozumění dynamiky akciových trhů. V poslední části je zkonstruován model pro predikci volatility, kovariace a korelací. Použití odhadů pomocí naší waveletové realizované teorie zlepší predikční schopnosti. Odhad individuálních skoků a společných skoků pomocí naší teorie dále zlepší predikční schopnosti realizované kovariace.

Contents

Acknowledgments

Looking back in time, I would never have imagined that four years of Ph.D. study would turn into such an enjoyable journey. Here, I would like to express my gratitude to those who made this happen. Without them, I would never have been able to learn so quickly and enjoy it so much.

Most of all, I have benefited from the positive attitude and enthusiasm of my colleague Lukáš Vácha and my supervisor Miloslav Vošvrda. I started working with Lukáš on our first project exactly four years ago, and from the very beginning I could feel a huge benefit from our collaboration. Lukáš is also co-author of the theory presented in this thesis. As my supervisor, Miloslav has provided me with razor-sharp guidance which has shaped my professional life, always helping me to cross my own borders. The research presented in this dissertation also greatly benefited from the help of Dr. Najzar from the Faculty of Mathematics and Physics at Charles University in Prague, who was willing to provide us with a solid mathematical background for wavelet theory and also to help us formalize some of the proofs. The support from the Czech Science Foundation under Grant 402/09/H045 and Centre of research MSMT LC06075 is gratefully acknowledged.

I am saving the last paragraph for the people whom most researchers owe the most – family. I am grateful to Lucia, as her patience, love and support has no bounds, as well as to my parents for the unconditional love they provided me with.

Prague, September 2011
Jozef Baruník

Nontechnical Summary

The research of financial market volatility and co-volatility has been progressing rapidly in recent decades. These topics are critical to fundamental risk and asset pricing and have important applications to risk management and portfolio allocation. More generally, research on financial markets is a highly empirical discipline that has become one of the most active fields in the social sciences. Beyond the motivation of high expected profits, or perhaps the high willingness of market participants to pay for this kind of research, there are deep intellectual reasons driving this huge interest in financial markets. Over the past 30 years, financial econometricians have uncovered fascinating properties and regularities of asset returns. The market itself has been reminding us of the crucial properties that help us to understand its behavior: events like sudden stock market crashes are a good example. Black Monday on October 19, 1987, prompted researchers to sit up and notice that the underlying distribution of asset returns is far from normal.

Regardless of the huge amount of commotion present in the data, the most promising part of the research was the predictability of the second distributional moment of returns. Volatility clustering, persistence and its time variation are directly observable even to random bystanders. Thus, a vast literature emerged on the phenomenon, and time-varying volatility approaches became a traditional benchmark. The most direct way to capture the dynamics of second moments is to write a process for the volatility of returns conditional on past returns and other available information. Robert Engle was the first to propose this approach to volatility modeling by allowing for conditional heteroskedasticity to the autoregressive process of returns (the ARCH model). His seminal contribution, which changed the world of stock market modeling, was recognized in 2003 when he was awarded the Nobel Memorial Prize in Economic Sciences. With this award, interest in financial market volatility increased even further. An alternative approach that has been attracting interest in recent years permits the conditional mean and variance of financial returns to depend on a latent state that cannot be observed directly; this so-called regime-switching approach was pioneered by James Hamilton. The technical difficulty of this approach is that in order to capture the complexity of asset markets one would like to allow for many possible states in volatility, resulting in an inapplicable model with a very large number of parameters to be estimated.

In recent years, the study of high-frequency financial data has brought new development into the field. Explosive growth in the availability of such data has made the unobservable theoretical process of volatility suddenly observable. As noted by Nour Meddahi, Per Mykland and Neil Shephard in the editorial of the special issue of the *Journal of Econometrics* on realized volatility modeling published in January 2011, the concept of: *"Realized Volatility is emblematic of this development, in that it was the earliest estimator which took advantage of the data in a non-parametric fashion."* Thus, with the availability of high-frequency data, modeling of the second distributional moment of returns has turned into simple use of nonparametric measures. The important result of the field is the finding that the returns standardized by the realized volatility are asymptotically normally distributed, as implied by the normal-mixture hypothesis. As both the data availability and the theoretical results have grown, it has become possible to find answers to more complex questions, up to the point where *"Realized Volatility"* is now as much the name of a paradigm as the name of an estimator.

The present dissertation contributes to this fascinating research in several ways. It offers a new complete theory generalizing the popular realized volatility as well as covariance measures. The theory introduces innovative estimators robust to noise as well as jumps in the financial stock markets and brings a considerable improvement to estimates of the realized variance-covariance matrix in terms of precision. The theory can also be used to disentangle jumps and co-jumps from the continuous part of the processes. Moreover, the newly developed estimators are able to decompose the realized measures into several investment horizons, providing much general understanding of the stock markets. This last theoretical novelty allows volatility and covariation to be studied in the time-frequency domain. In addition, the dissertation contains a small sample study of the properties of the proposed theory as well as the small sample performance of the estimators in the forecasting exercise under different conditions, confirming the theoretical results. Another notable contribution to the research field lies in the application of the derived theory. General time-frequency estimators yield not only more efficient estimates, but also insights into the decomposed realized variance, covariance or its transformations, correlation and portfolio beta. Thus, the theory helps to improve our understanding of stock market generating processes. The dissertation is divided into two parts presenting the theory and application on univariate as well as multivariate results.

After the necessary introduction to the theory of quadratic variation and realized variation measurement, the first part of the dissertation defines the wavelet-based realized variation theory. Standing on our theoretical results proposing the Wavelet Representation Theorem, which extends the well-known Martingale Representation Theorem, the estimator of wavelet-based realized variation is defined together with its theoretical properties. Using wavelets, the estimator is able to consistently estimate jumps from the price process. It is robust to noise and it generates an unbiased consistent estimator of the *true* underlying variance. The theoretical part also contains an important discussion of the similarities between wavelet theory and stochastic processes.

To support the theory, a numerical study of the small sample performance of the estimators is carried out. In this study, we compare our estimators to several of the most popular estimators, namely, realized variance, bipower variation, two-scale realized volatility and realized kernels. The wavelet-based estimator proves to have the lowest bias of all the estimators in the jump-diffusion model with stochastic volatility as well as the fractional stochastic volatility model simulated with different levels of noise and numbers of jumps. While all the other estimators suffer from substantial bias caused either by jumps or by noise, our theory proves to hold its properties. As predictability of volatility is of interest to researchers as well as practitioners, a numerical study of the behavior of the forecasts is also carried out. Again, our theory proves to be the most powerful in forecasting volatility under the different simulation settings.

While the first chapters of Part I derive the theory and show its power on a small sample study, the last chapter uses the theory to decompose the empirical volatility. By studying the statistical properties of unconditional daily log-return distributions standardized by volatility estimated using the different estimators we find that standardization by our wavelet-based estimator brings the returns close to the Gaussian normal distribution. All the other estimators are affected by the presence of jumps in the data. The differences are economically significant, as we find that the average volatility estimated using our wavelet-based theory is 6.34% lower than the volatility estimated with the standard estimator.

Furthermore, we decompose the realized volatility into several intraday horizons. Here we note that the theory is able to decompose the realized measures into any arbitrary investment horizon, i.e., from 1-minute up to 1-month, when estimating monthly measures. In our analysis performed on forex data, we limit ourselves to illustrating the theory on the decomposition of daily realized measures. Specifically, we decompose the realized volatility into investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes and 40–80 minutes, and the rest (80 minutes up to 1 day). The analysis uncovers interesting dynamics. Most of the action in the stock markets comes from higher frequencies. We find that on average, about 52% of the volatility of the forex markets examined is created on the 5–10 minute investment horizon, approximately 25% comes from the 10–20 minute investment horizon, and only 12%, 6% and 5% correspond to the horizons of 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day), respectively. Note that by adding the contributions of the different investment horizons we always get 100%.

The last part of the univariate empirical analysis is devoted to the forecasting of realized volatility. One of the issues with the interpretation of wavelets in economic applications is that they behave like a filter. Thus wavelets can hardly be used for forecasting economic time series most of the time. But in the realized measures, we only use wavelets to decompose the daily variation of the returns using intraday information, while forecasting daily volatility. We build a new forecasting model based on an ARFIMA type

model using the decomposition provided by our theory. In-sample as well as more important out-of-sample forecasts show that our theory is able to forecast volatility with the lowest error. Concluding the empirical findings, we show that our wavelet-based theory brings a significant improvement to volatility estimation and forecasting. It also offers a new method of time-frequency modeling of realized volatility which helps us to better understand the dynamics of stock market behavior. Specifically, our theory uncovers that most of the volatility is created on higher frequencies.

The second part of this dissertation follows the structure of the first part closely. After the necessary introduction of the generalized multivariate framework for modeling the covariation structure between processes, we build a new, wavelet-based realized covariation theory by extending the findings from the univariate part, and we define the wavelet-based realized estimator of covariance together with its properties. We use wavelets to disentangle jumps from co-jumps, which is crucial in the study of multivariate dependencies. Having defined the estimators of variance and covariance, we also define the transformations of interest for portfolio theory: the wavelet-based realized correlation measure and the wavelet-based realized beta. Similarly to the univariate findings, the presented theory provides a new type of multivariate estimators in the time-frequency domain.

To support the theoretical results, we again run a numerical study of the small sample behavior of the estimators. In the study, we simulate prices using a bivariate jump-diffusion stochastic volatility process and compare the performance of the wavelet-based realized covariation and correlation estimators with the popular realized covariance, bipower realized covariance, two-scale realized covariance and multivariate kernels estimators. The study proves that in this generalized setting as well, our wavelet-based realized theory is able to outperform other methods of estimation, as it displays the lowest bias under different amounts of simulated noise and jumps in the bivariate process.

The last chapter applies the multivariate theory and studies the decomposition of integrated covariation, correlation and beta on the forex markets. Our estimator is able to separate jumps, co-jumps and *true* covariation from the data. It is also robust to the Epps effect caused by noise in the data. The results suggest that understanding jumps and co-jumps in a multivariate setting may be crucial for studying dependencies. While individual jumps bring some bias to the covariance, co-jumps introduce large bias into the covariation measure. The impact on correlation is even more crucial. Individual jumps in the processes bring large downward bias to the correlation measure, while co-jumps introduce upward bias with a smaller magnitude.

The empirical part also contains an interesting study of multivariate unconditional volatility distributions and their decomposition into several investment horizons. While the multivariate volatilities show strong dependence, a volatility-in-correlation effect suggests that the standard mean-variance efficiency calculations based on constant correla-

tions are misguided. Our results have significant economic value, as a wrong assumption about the dependence process will have a direct impact on the portfolio valuation. The dynamics of the decomposed dependencies reveal interesting results as well. Our wavelet-based realized theory generates a more precise correlation measure with narrower confidence intervals than the standard realized correlations. A study of the temporal dependence in the decomposed correlations reveals that a similar heterogeneous autoregression type model as in the univariate case should be used for forecasting.

Similarly to the univariate part, therefore, we build a forecasting model for covariation and correlation based on wavelet decomposition. Our model again outperforms all other models in-sample as well as out-of-sample. Finally, the wavelet-based realized beta estimator proves to be more precise, with narrower confidence intervals than the realized beta. All these results shed more light on the dependence, thus improving the applications in portfolio theory.

In conclusion, this dissertation presents a new theoretical framework generalizing the popular concept of realized variance and covariance. The work also contributes to the literature by providing interesting empirical findings from our time-frequency realized measures.

Part I

Wavelet-based realized variation theory

Chapter 1

Introduction

Volatility of asset returns has become one of the primary concerns in financial econometrics research over the past decade. The main reason for this is the greatly expanding availability of high-frequency data, which has made the unobservable part of stochastic volatility models observable. The increasingly popular *Realized Volatility* approach was pioneering work which took advantage of the data in a nonparametric fashion, but as both theoretical insights and data availability have grown rapidly in the past decade, *Realized Volatility* has become as much the name of a paradigm shift as the name of an estimator.

The popularity of realized volatility is mainly due to its two distinct implications for practical estimation and forecasting. The first relates to the measurement of realizations of the latent volatility process without the need for any assumptions about the explicit model. The second brings the possibility of modeling volatility directly through standard time series econometrics with discretely sampled daily data, while effectively extracting information from intraday high-frequency data.

The most fundamental result in realized variation states that it provides a consistent nonparametric estimate of price variability over a given time interval. The formalized theory is presented by Andersen et al. (2003). While these authors provide a unified framework for modeling, Zhou (1996) was one of the first to provide a formal assessment of the relationship between cumulative squared intraday returns and the underlying return variance. The pioneering work by Olsen & Associates on the use of high-frequency data, summarized by Dacorogna et al. (2001), produced milestone results for many of the more recent empirical developments in realized variation. A vast quantity of literature on several aspects of estimating volatility has emerged in the wake of these fundamental contributions. Rather than providing a tedious literature review here, we will introduce the main findings of the literature gradually in the text while discussing the aspects they develop.

Our work builds on the popular *Realized Volatility* approach, bringing even more insights to the theory. While most time series models are set in the time domain, we enrich

the analysis by the frequency domain. This is enabled by the use of the continuous wavelet transform. It is a logical step to take, as the stock markets are believed to be driven by heterogeneous investment horizons. In our work, we ask if wavelet decomposition can improve our understanding of volatility series and hence improve volatility forecasting and risk management.

Another very appealing feature of wavelets is that they can be embedded into stochastic processes, as shown by Antoniou and Gustafson (1999). Thus we can conveniently use them to extend the theory of realized measures. One of the issues with the interpretation of wavelets in economic applications is that they behave like a filter. Thus wavelets can hardly be used for forecasting in econometrics. But in the realized measures, we use wavelets only to decompose the daily variation of the returns using intraday information. Moreover, the approach suggests constructing a model from the wavelet decomposition.

We are not the first to use this idea. Several attempts to use wavelets in the estimation of realized variation have emerged in the past few years. Hø g and Lunde (2003) were the first to suggest a wavelet estimator of realized variance. Capobianco (2004), for example, proposes to use a wavelet transform as a comparable estimator of quadratic variation. Subbotin (2008) uses wavelets to decompose volatility into a multi-horizon scale. Next, Nielsen and Frederiksen (2008) compare the finite sample properties of three integrated variance estimators, i.e., realized variance, Fourier and wavelet estimators. They consider several processes generating time series with a long memory, jump processes as well as bid-ask bounce. Gençay et al. (2010) mention the possible use of wavelet multiresolution analysis to decompose realized variance in their paper, while they concentrate on developing much more complicated structures of variance modeling in different regimes through wavelet-domain hidden Markov models.

One remarkable exception which fully completes the current literature on using wavelets in realized variation theory is the work of Fan and Wang (2007), who were the first to use the wavelet-based realized variance estimator and also the methodology for the estimation of jumps from the data. In our work, we generalize the results of Fan and Wang (2007) in several ways. Instead of using the Discrete Wavelet Transform we use the Maximum Overlap Discrete Wavelet Transform, which is a more efficient estimator and is not restricted to sample sizes that are powers of two. We also use the Daubechies family of wavelets instead of the Haar type. Moreover, in the next chapters of this thesis, we will introduce a generalization of this approach to covariation estimation. In the next section, we will define the wavelet-based framework for the estimation of realized variance.

Thus in our work, we propose a theory for wavelet-based estimation of realized variation. Generalization of the Martingale Representation Theorem to our Wavelet Representation Theorem gives us the power to decompose the return processes into several investment horizons in continuous time. We show that asymptotically, the wavelet decomposition is the same as the realized volatility estimator. Connecting it with the result of Zhang et al. (2005), who introduced a two-scale realized volatility estimator robust to noise into the literature, we arrive at an estimator which is robust to both jumps and noise. Based on this result, we present a complete wavelet-based realized variation theory generalizing the realized measures. Moreover, we use wavelets for jump detection.

To study the small sample behavior we run a large numerical study showing that the asymptotic property holds under various settings, and our wavelet-based estimator also proves to have the lowest forecast bias.

After the theory is derived, we apply our estimator to the modeling of currency futures volatility. The results show that daily returns standardized by the volatility estimated by our methodology are much closer to the standard normal distribution than any other realized volatility estimator. Moreover, we take advantage of the wavelet-based estimator and decompose the volatility series into a jump component and several investment horizons. The decomposition reveals interesting results. Most of the volatility in the stock markets comes from high frequencies. Based on this decomposition, we also build an ARFIMA-based forecasting model, which proves to have the best forecasts of volatility.

The first part of this dissertation is organized in chapters. The first chapter introduces the theory of quadratic variation, continuous-time price processes and standard variance measures. It also discusses the effects of microstructure noise and jumps in variance estimation. The second chapter contains the complete wavelet-based realized variance theory, while all of the proofs are relegated to a technical appendix (Appendix 11A). After presenting the introduction to wavelet theory, we derive a generalization of the Martingale Representation Theorem called the Wavelet Representation Theorem. Based on this result, we build our wavelet-based realized variance estimator and derive its properties. Chapter 4 tests the theory in a numerical study and compares the small sample behavior of the wavelet-based estimator with other popular estimators, while assuming different processes driving the stock market with different amounts of noise and jumps. Specifically, we consider jump-diffusion stochastic volatility and fractional stochastic volatility. The chapter concludes with a numerical study assessing the forecasting performance of the estimators. The last chapter of this part applies the presented theory, decomposes the empirical volatility of forex stock markets and finally uses the decomposition for forecasting.

Chapter 2

Quadratic return variation

2.1 Continuous-time no-arbitrage price processes

We begin with a description of the framework used for studying processes in a continuous-time no-arbitrage setting. A proper characterization of the price process as well as the introduction of definitions are central to the variation measures we rely on through this work. The concepts introduced in the following section will thus be cornerstones of the rest of the thesis.

Asset returns are commonly assumed to contain a predictable component, which compensates the investor for the risk of holding the security, and an unobservable shock, which cannot be predicted using the available information. The common assumption of no arbitrage implies that stock market return innovations have a higher magnitude than the mean return. Thus, over an infinitesimal time interval, we do not need to specify the conditional mean return in order to be able to identify the conditional return variation. This result has important implications for the approach to modeling and measuring variation in continuous time.

Consider a univariate risky logarithmic asset price process p_t defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The price process evolves in continuous time over the interval $[0, T]$, where T is a finite positive integer. Further, consider the natural information filtration, an increasing family of σ-fields $(\mathcal{F}_t)_{t\in[0,T]} \subseteq \mathcal{F}$, which satisfies the usual conditions. Information set $\mathcal{F}_t$ contains the full history up to time t of the realized values of the asset price and other relevant state variables.

Following Andersen et al. (2003), we define the continuously compounded asset return over the $[t-h, t]$ time interval, $0 \leq h \leq t \leq T$, by $r_{t,h} = p_t - p_{t-h}$. A special case of the continuously compounded return, the cumulative return process from $t = 0$ up to time t, $r_t = (r_t)_{t\in[0,T]}$, is then $r_t \equiv r_{t,t} = p_t - p_0$ and inherits all the main properties of p_t. These definitions imply a simple relation between the period-by-period and the cumulative returns that we use repeatedly in the further text: $r_{t,h} = r_t - r_{t-h}$, $0 \leq h \leq t \leq T$.

Moreover, the asset price process is also assumed to remain almost surely (henceforth

$a.s.$) strictly positive and finite so that p_t and r_t are well defined over $[0,T]\,(a.s.)$. r_t has only a countable number of jump points over $[0,T]$, and both the price process and the return process are squared integrable. Defining $r_{t-} \equiv \lim_{\tau\to t,\tau<t} r_\tau$ and $r_{t+} \equiv \lim_{\tau\to t,\tau>t} r_\tau$ uniquely determines the right-continuous, left-limit (càdlàg[1]) version of the process, for which $r_t = r_{t+}(a.s.)$, and the left-continuous, right-limit (càglàd)[2] version, for which $r_t = r_{t-}(a.s.)$, for all $t \in [0,T]$. In what follows, we assume without loss of generality that we work with the càdlàg version of the return processes.

The jumps in the cumulative price process and hence the return process are $\Delta r_t = r_t - r_{t-}$, $0 \le t \le T$. At continuity points in r_t, we have $\Delta r_t = 0$. Moreover, $P\left(\Delta r_t \neq 0\right) = 0$ for any arbitrary chosen $t \in [0,T]$. This assumption does not imply that jumps are rare and we also need to assume that the jump process does not explode. The nonexplosive jump process will be called a *regular* process with a finite number of jumps.

After the necessary introductions to the price process have been made, we can invoke the final standard assumptions to complete the definition of the continuous-time no-arbitrage price process. In a frictionless market with no arbitrage and finite-expected returns, p_t must constitute a special semi-martingale (Back, 1991). A fundamental result of stochastic integration theory states that such processes permit a unique canonical decomposition (e.g. Protter, 1992).

Proposition 1 *Return decomposition*
Any arbitrage-free logarithmic price process $(p_t)_{t\in[0,T]}$ subject to the regularity conditions (outlined above) may be uniquely represented by

$$r_t \equiv p_t - p_0 = \mu_t + M_t = \mu_t + M_t^C + M_t^J, \tag{2.1}$$

where μ_t is a predictable and finite-variation process, M_t is a local martingale that may be further decomposed to M_t^C, a continuous sample path, infinite variation local martingale component, and M_t^J, a compensated jump martingale. By definition, $\mu_0 \equiv M_0 \equiv M_0^C \equiv M_0^J \equiv 0$, which implies that $r_t \equiv p_t$.

The instantaneous return r_t, can thus be uniquely decomposed into a predictable and integrable mean (expected return) component and a local martingale innovation. It is natural first to concentrate on the behavior of the martingale component of the decomposition 2.1. However, we would need a continuous record of the data to be able to observe M_t. Consequently, we have to focus on measures that represent the variation over a discrete time interval, suggesting a natural notion of variation based on the quadratic variation process for the local martingale component. Over a discrete time interval, the decomposition 2.1 becomes $r_{t,h} = \mu_{t,h} + M_{t,h}^C + M_{t,h}^J$, where $\mu_{t,h} = \mu_t - \mu_{t-h}$, $M_{t,h}^C = M_t^C - M_{t-h}^C$ and $M_{t,h}^J = M_t^J - M_{t-h}^J$.

Generally, let r_t denote any semi-martingale process. The unique *quadratic variation process*, $[r,r]_t$, $t \in [0,T]$, associated with r_t is formally defined as:

$$[r,r]_t = r_t^2 - 2\int_0^t r_{s-} dr_s, \tag{2.2}$$

[1]From French "continu à droite, limite à gauche".
[2]From French "continu à gauche, limite à droite".

and $r_{t-} = \lim_{s \uparrow t} r_s$ is well-defined (Protter, 1992). If the finite variation process μ_t from the decomposition 2.1 is continuous, then its quadratic variation is zero. Thus predictable component does not affect the quadratic variation of the return $(r_t)_{t\in[0,T]}$. Thus, quadratic variation of the return process over $[t-h,t]$ can be obtained from (e.g. Andersen et al., 2003 and Barndorff-Nielsen and Shephard, 2002b):

Definition 1 *Quadratic return variation*
The quadratic return variation of $(r_t)_{t\in[0,T]}$ *over* $[t-h,t]$, *for* $0 \leq h \leq t \leq T$, *is*

$$QV_{t,h} = [r,r]_t - [r,r]_{t-h} = \left[M^C, M^C\right]_t - \left[M^C, M^C\right]_{t-h} + \sum_{t-h<s\leq t} \delta M_s^2 \quad (2.3)$$

$$= \left[M^C, M^C\right]_t - \left[M^C, M^C\right]_{t-h} + \sum_{t-h<s\leq t} \delta r_s^2 \quad (2.4)$$

Most continuous-time models for asset returns can be cast within the very general setting of Proposition 1. Quadratic variation provides a framework to study the major object of interest in financial econometrics: the model-implied return variation (as well as its square root called volatility). While the integral representations for continuous sample path semi-martingales corresponding to Proposition 1 are rather abstract, the continuous-time models in the theoretical asset and derivatives pricing literature are frequently assumed to have continuous sample paths with the corresponding diffusion processes given in the form of stochastic differential equations (SDE henceforth). This assumption can be made using the following result without loss of generality (Protter, 1992).

Proposition 2 *Martingale representation theorem*
For any univariate, square-integrable, continuous sample path, logarithmic price process $(p_t)_{t\in[0,T]}$ *which is not locally riskless, there exists a representation such that over* $[t-h,t]$, *for all* $0 \leq h \leq t \leq T$

$$r_{t,h} = \mu_{t,h} + M_{t,h} = \int_{t-h}^{t} \mu_s ds + \int_{t-h}^{t} \sigma_s dW_s, \quad (2.5)$$

where μ_s *is an integrable, predictable and finite-variation stochastic process,* σ_s *is a strictly positive càdlàg stochastic process satisfying*

$$P\left[\int_{t-h}^{t} \sigma_s^2 ds < \infty\right] = 1,$$

and W_t *is a standard Brownian motion.*

As an example, we consider the Black and Scholes (1973) model, which is a special case of the setting discussed. Conditional mean process μ is constant, the continuous martingale M^C is a standard Brownian motion process, and the jump martingale M^J is zero:

$$dp_t = \mu dt + \sigma dW_t. \quad (2.6)$$

In this case, the quadratic variation over $[t-h,t]$ for $0 \leq h \leq t \leq T$ simplifies to

$$QV_{t,h} = \int_{t-h}^{t} \sigma^2 ds = \sigma^2 h, \tag{2.7}$$

thus the return variation is constant over any time interval of length h.

Another commonly used example in the financial literature is the following jump-diffusion model of Merton (1976a):

$$dp_t = (\mu - \lambda\bar{\xi})dt + \sigma dW_t + \xi_t d_{q_t}, \tag{2.8}$$

where q is a Poisson process uncorrelated with W and governed by the constant jump intensity λ. The magnitude of the jump in the return process is controlled by factor $\xi_t \sim N(\bar{\xi}, \sigma_\xi^2)$. The quadratic return variation of this process over $[t-h,t]$ for $0 \leq h \leq t \leq T$ is

$$QV_{t,h} = \int_{t-h}^{t} \sigma^2 ds + \sum_{t-h \leq s \leq t} J_s^2 = \sigma^2 h + \sum_{t-h \leq s \leq t} J_s^2, \tag{2.9}$$

where $J_t = \xi_t dq_t$ is non-zero only if a jump occurs. When compared to the previous example represented by 2.6, the return variation is also constant, but it contains jump variation.

Finally, a very general class of stochastic volatility models that we will use in this work can be defined by the following jump-diffusion model:

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t, \tag{2.10}$$

where q is a constant-intensity Poisson process with the same magnitude as process 2.8. Process 2.10 may be characterized as a Brownian semi-martingale with finite jump activity, and is also a special case of decomposition 1. Its quadratic return variation over $[t-h,t]$ for $0 \leq h \leq t \leq T$ is

$$QV_{t,h} = \underbrace{\int_{t-h}^{t} \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \leq s \leq t} J_s^2}_{\text{Jump Variation}} . \tag{2.11}$$

Thus generally, we will consider the quadratic variation to contain two parts – a so-called *Integrated Variance* part and variation of jumps.

2.2 Realized variance measurement

Model-free measures of return variation based only on realizations of the return have a long history in the literature. For example, French et al. (1987) use daily return observations to compute monthly realized variance estimates. More recently, the availability of transaction data has made it possible to refine earlier measures of historical volatility

into the notion of *realized variance*, a consistent estimator of quadratic variation, as popularized mainly by Andersen et al. (2003). In fact, the quadratic variation of the process has become observable with the availability of high-frequency data, and this approach to variation measurement has started ongoing new research into volatility modeling based on general distributional assumptions. Let us introduce the estimator.

Definition 2 *Realized variance*
The realized variance over $[t-h,t]$*, for* $0 \leq h \leq t \leq T$*, is defined by*

$$\widehat{RV}_{t,h} = \sum_{i=1}^{n} r^2_{t-h+\left(\frac{i}{n}\right)h}, \tag{2.12}$$

where n *is the number of observations in* $[t-h,t]$*.*

The realized variance is simply the second sample moment of the return process over a fixed interval of length h, scaled by the number of observations n so that it provides a variance measure calibrated to the h-period measurement interval. Semi-martingale theory ensures that the realized variance measure $\widehat{RV}_{t,h}$ described by 2.12 converges to the return quadratic variation QV described by 2.3. Details of this important result as well as other theoretical properties can be found in Andersen and Bollerslev (1998), Andersen et al. (2001, 2003) and Barndorff-Nielsen and Shephard (2001, 2002a,b).

Proposition 3 *Realized variance as an unbiased variance estimator*
If the return process $(r_t)_{t\in[0,T]}$ *is square-integrable and* $\mu_t \equiv 0$*, then for any value of* $n \geq 1$ *and* $h > 0$*,*

$$E\left[RV_{t,h} \middle| \mathcal{F}_t\right] = E\left[M^2_{t,h} \middle| \mathcal{F}_t\right] = E\left[\widehat{RV}_{t,h} \middle| \mathcal{F}_t\right] \tag{2.13}$$

Proposition 4 *Consistency of realized variance*
The realized variance provides a consistent nonparametric measure of the variance,

$$\text{plim}_{n\to\infty} \widehat{RV}_{t,h} = RV_{t,h}, \quad 0 \leq h \leq t \leq T, \tag{2.14}$$

where the convergence is uniform in probability.

Thus as such, the *ex-post* realized variance, $\widehat{RV}_{t,h}$, is an unbiased estimator of the *ex-ante* expected variance $RV_{t,h}$. With increasing sampling frequency $n \to \infty$, the realized variance is also a consistent estimator of the variance over any fixed-length time interval, $h > 0$. While the realized variance is of direct interest as an indicator of return variability, it can also provide an indication of the underlying return distribution itself. The returns $r_{t,h}$, conditional on the variance and the mean return over the interval $[t-h,t]$, will be Gaussian.

Proposition 5 *Normal mixture distribution*
The discrete-time returns $r_{t,h}$ *over* $[t-h,t]$*, for* $0 \leq h \leq t \leq T$*, from the continuous sample path diffusion* $dp_t = \mu_t dt + \sigma_t dW_t$*,* $0 \leq t \leq T$*, are distributed as a normal mixture,*

$$r_{t,h} | \sigma\{\mu_{t,h}, RV_{t,h}\} \sim N\left(\mu_{t,h}, RV_{t,h}\right), \tag{2.15}$$

provided that the Brownian Motion, W_t*, is independent of* μ_{p_t,σ_t} *and* σ_t*.*

The consistency of the realized variance and the normal mixture distribution in Proposition 5 imply simple alternative empirical return-variation measurement. But there are two issues which complicate the practical use of the nice convergence results. As the realized variance is a consistent estimator with increasing sampling frequency, $n \to \infty$, a continuum of instantaneous return observations must be used in order for the realized volatility estimate to converge to the realized volatility. In practice, we can observe only discrete prices, and thus an inevitable discretization error is present. On the other hand, market microstructure effects such as price discreteness, bid-ask spread and bid-ask bounce contaminate the return observations. Thus in practice, the return process should not be sampled too often, regardless of the number of observations available, to avoid large bias from the market microstructure. The literature has extensively studied the noise-to-signal ratio and has constructed optimal sampling schemes, which range from 5 to 30 minutes, for instance. The main literature is nicely surveyed by Hansen and Lunde (2006), Bandi and Russell (2006b) and McAleer M (2008); Andersen and Benzoni (2007). Recent notable contributions to this literature include Zhang et al. (2005), Bandi and Russell (2006a) and Barndorff-Nielsen et al. (2008).

To make the definitions complete, we also define the realized volatility, which will be mainly used in the empirical part of this work. The realized volatility is defined as the square root of the realized variance.

Definition 3 *Realized volatility*
The realized volatility over $[t-h, t]$*, for* $0 \leq h \leq t \leq T$*, is defined as*

$$\widehat{RV}_{t,h}^{1/2} \tag{2.16}$$

where $\widehat{RV}_{t,h}$ *is the realized variance defined in the Equation 2.12.*

From now on, when referring to the realized variance, we will be referring to the Equation 2.12 while when referring to the realized volatility, we will be referring to the square root of the realized variance (Equation 2.16).

This approach to handling data still results in the discarding of large amount of available information. Consider the case of original observed data sampled once every second. Trying to avoid microstructure noise and sampling the data at 5-minute frequency, we would discard 299 out of every 300 data points. In practice, with data available for more liquid stocks, we would throw away even larger amounts of data. It is very difficult to accept that throwing away data, especially in such quantities, is the optimal solution. Zhang et al. (2005) propose a solution that we describe in detail in the next section.

2.3 The effects of microstructure noise

While estimator 2.12 is not efficient and becomes inconsistent in the presence of noise, we follow Zhang et al. (2005)'s approach to generalize the setting from the previous section with the assumption of noise.

Proposition 6 *Let $(y_t)_{t\in[0,T]}$ be the observed log prices, which will be equal to the latent, so-called "true log-price process", $dp_t = \mu_t dt + \sigma_t dW_t$, $0 \leq t \leq T$, and will contain microstructure noise ϵ_t*

$$y_t = p_t + \epsilon_t, \tag{2.17}$$

where ϵ_t is zero mean i.i.d. noise with variance η^2

The thoughtful reader will have noticed that we have dropped the jumps to zero for now, just to introduce the notion of realized measures in the presence of noise. The main object of interest is the estimated integrated variance of the observed process 2.17. From now on, we will denote $\langle p,p\rangle_t = \int_{t-h}^{t} \sigma_t^2 dt$ over the time period $[t-h,t]$, the *true* integrated variance of the process $p_{t,h}$. The reason for referring to $\langle p,p\rangle_t$ as *true* is straightforward; it is actually the variation of the $p_{t,h}$ process which is of interest, while $y_{t,h}$ contains noise. The estimate of the integrated variance $\langle p,p\rangle_t$ will always be denoted as $\widehat{RV}_{t,h}^{(estimator)}$ in the further text, where $(estimator)$ will be replaced by the abbreviation of the estimator used; for example, the RV estimator defined in the previous section by Equation 2.12, will be denoted as $\widehat{RV}_{t,h}^{(sparse)}$. Note that we in fact refer to $[p,p]_t$ using this notation.

If one uses all the log-returns data available in good faith so as not to throw away the data and maintain the consistency of the estimator with $n \to \infty$ as in Proposition 4, one will get huge bias:

$$\widehat{RV}_{t,h}^{(all)} \overset{\mathcal{L}}{\approx} \underbrace{\langle p,p\rangle_t}_{\text{true IV}} + \underbrace{2nE[\epsilon^2]}_{\text{bias due to noise}} + \left[\underbrace{\underbrace{4nE[\epsilon^4]}_{\text{bias due to noise}} + \underbrace{\frac{2T}{n}\int_{t-h}^{t}\sigma_t^4 dt}_{\text{bias due to discretization}}}_{\text{total variance}} \right]^{1/2} Z_{total} \tag{2.18}$$

conditional on the $p_{t,h}$ process, where $\overset{\mathcal{L}}{\approx}$ denotes stable convergence in law and Z_{total} is a standard normal variable. For large n, the realized variance diverges to infinity linearly in n as the bias is of order $O(n)$ and the *true* quadratic variation, $\langle p,p\rangle_t$, is of order $O(1)$.

In practice, we will not use all the data for the estimation. Instead, the estimator $\widehat{RV}_{t,h}^{(sparse)}$ would be constructed by summing the squared log-returns at lower frequencies. For example, a frequency of 5, 10, 15 or 30 minutes, is typically used. Reducing n will also dramatically reduce the bias $2nE[\epsilon^2]$. Still, one of the basic lessons in statistics is the warning that we should not be doing this.

The solution to the problem of throwing away data is the *Two-Scale Realized Volatility* estimator (TSRV henceforth) of Zhang et al. (2005). In computing the TSRV, we have to first partition the original grid of observation times, $G = \{t_0,\ldots,t_n\}$, into subsamples

$G^{(k)}$, $k = 1, \ldots, K$, where $n/K \to \infty$ as $n \to \infty$. For example, $G^{(1)}$ will start at the first observation and take an observation every 5 minutes, $G^{(2)}$ will start at the second observation and take an observation every 5 minutes, etc. Finally, we average these estimators through the subsamples, so we average the variation of the estimator also. Thus the average estimator:

$$\widehat{RV}_{t,h}^{(average)} = \frac{1}{K} \sum_{k=1}^{K} \widehat{RV}_{t,h}^{(k)} \tag{2.19}$$

can be constructed by averaging the estimators $\widehat{RV}_{t,h}^{(k)}$ obtained on K grids of average size $\bar{n} = n/K$. The properties of this estimator are

$$\begin{aligned} \widehat{RV}_{t,h}^{(average)} \overset{\mathcal{L}}{\approx} & \underbrace{\langle p, p \rangle_t}_{\text{true IV}} + \underbrace{2\bar{n}E[\epsilon^2]}_{\text{bias due to noise}} \\ & + \left[\underbrace{\underbrace{4\frac{\bar{n}}{K}E[\epsilon^4]}_{\text{bias due to noise}} + \underbrace{\frac{4T}{3\bar{n}} \int_{t-h}^{t} \sigma_t^4 dt}_{\text{bias due to discretization}}}_{\text{total variance}} \right]^{1/2} Z_{total}. \end{aligned} \tag{2.20}$$

While $\widehat{RV}_{t,h}^{(averge)}$ is a better estimator than $\widehat{RV}_{t,h}^{(all)}$ as $\bar{n} < n$, bias still is present. The idea of Zhang et al. (2005) was that $E[\epsilon^2]$ can actually be consistently approximated using the realized variance computed using all the observations, $\widehat{E[\epsilon^2]} = \frac{1}{2n}\widehat{RV}_{t,h}^{(all)}$. Hence, the bias of $\widehat{RV}_{t,h}^{(average)}$ can be consistently estimated, and the TSRV is proposed as the bias-adjusted estimator of $\langle p, p \rangle_t$.

Definition 4 *Two-scale realized variance estimator*
The two-scale realized variation over $[t-h, t]$, for $0 \leq h \leq t \leq T$, is defined by

$$\widehat{RV}_{t,h}^{(TSRV)} = \underbrace{\widehat{RV}_{t,h}^{(average)}}_{\textit{slow time scale}} - \frac{\bar{n}}{n} \underbrace{\widehat{RV}_{t,h}^{(all)}}_{\textit{fast time scale}} . \tag{2.21}$$

For the optimal $K^* = cn^{2/3}$, the TSRV has the following properties:

$$\widehat{RV}_{t,h}^{(TSRV)} \overset{\mathcal{L}}{\approx} \underbrace{\langle p,p\rangle_t}_{\text{true IV}} + \\ + \frac{1}{n^{1/6}} \Bigg[\underbrace{\underbrace{\frac{8}{c^2} E[\epsilon^4]}_{\text{bias due to noise}} + \underbrace{c\frac{4T}{3}\int_{t-h}^{t} \sigma_t^4 dt}_{\text{bias due to discretization}}}_{\text{total variance}} \Bigg]^{1/2} Z_{total}. \quad (2.22)$$

$\widehat{RV}_{t,h}^{(TSRV)}$ provides the first consistent and asymptotic estimator of the quadratic variation of $\langle p,p\rangle_t$ with rate of convergence $n^{-1/6}$. Zhang (2006) shows a possible generalization to multiple scales converging at a slightly faster rate of convergence $n^{-1/4}$, but with a higher computational burden; the result will be very close to the TSRV estimator. The constant c can be set to minimize the total asymptotic variance from 11A.48. In small samples, a small sample refinement can be constructed:

$$\widehat{RV}_{t,h}^{(TSRV,adj)} = \left(1 - \frac{\bar{n}}{n}\right)^{(-1)} \widehat{RV}_{t,h}^{(TSRV)} \quad (2.23)$$

Again for simplicity in the notation, when using $\widehat{RV}_{t,h}^{(TSRV)}$ we always refer to the estimator adjusted for small sample bias, $\widehat{RV}_{t,h}^{(TSRV,adj)}$.

Another estimator, which is able to deal with the noise and which we use for the comparison in our study is the realized kernels (RK) estimator introduced by Barndorff-Nielsen et al. (2008). Their non-negative estimator takes on the following form.

Definition 5 *Relized Kernel estimator*
The realized kernel variance estimator over $[t-h,t]$, *for* $0 \le h \le t \le T$ *is defined by*

$$\widehat{RV}_{t,h}^{(RK)} = \gamma_{t,h,0} + \sum_{\eta=1}^{H} k\left(\frac{\eta-1}{H}\right)(\gamma_{t,h,\eta} + \gamma_{t,h,-\eta}), \quad (2.24)$$

with $\gamma_{t,h,\eta} = \sum_{i=1}^{n} r_{t-h+\left(\frac{i}{n}\right)h} r_{t-h+\left(\frac{i-\eta}{n}\right)h}$ *denoting the* η*-th realized autocovariance with* $\eta = -H,\ldots,-1,0,1,\ldots,H$ *and* $k(.)$ *denotes the kernel function.*

Please note that for $\eta = 0$, $\gamma_{t,h,\eta} = \gamma_{t,h,0} = \widehat{RV}_{t,h}$ is estimate of the realized variance from 2.12. For the estimator to work, we need to choose the kernel function $k(.)$. In our study, we will focus on the Parzen kernel because it satisfies the smoothness conditions, $k'(0) = k'(1) = 0$ and is guaranteed to produce a non-negative estimate. The Parzen

kernel function is given by

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1. \\ 0 & x > 1 \end{cases} \tag{2.25}$$

We should note that the realized kernel estimator is computed without accounting for end effects, i.e. replacing the first and the last observation by local averages to eliminate the corresponding noise components (so-called "jittering"). Barndorff-Nielsen et al. (2008) argue that these effects are important theoretically, but are negligible practically.

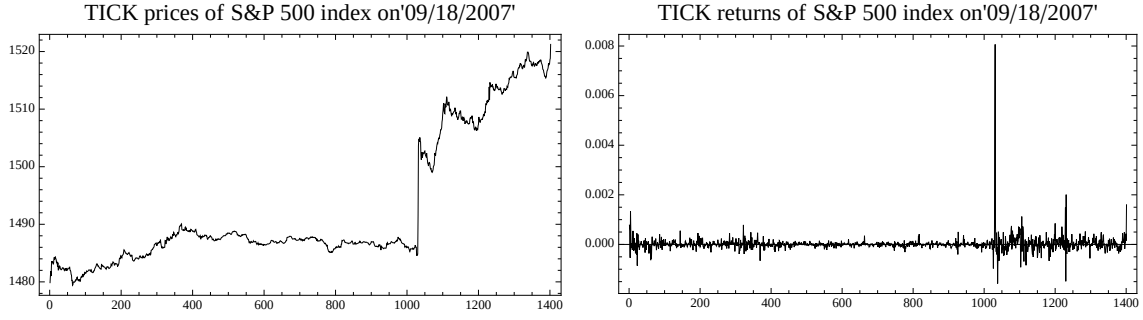
By introducing the TSRV and the RK estimators, we will be able to consistently estimate the realized variation from noisy observations. The last step we need is to add jumps to the analysis. To illustrate the importance of jumps in the estimation of realized measures, let us break the theoretical introductions with the following example.

2.4 The effects of jumps

While Zhang (2006) provides an estimator which enables us to estimate the integrated variance in the presence of noise, it is only solution to one part of the problem, as we consider the quadratic variation to be composed of integrated variance as well as jumps (Eq. 2.11). Naturally, it is much more appropriate to a use jump-diffusion model for financial data, as they contain a number of jumps which influence the distribution significantly.

Jump diffusion models have a rich history in financial econometrics, started probably by Merton (1976b). While it is natural to assume the price process to follow a jump diffusion process, researchers face lots of difficulties in estimating it on real data, as discontinuous sample paths create discontinuities in the econometric objective function. Here again, high-frequency data seems to be very helpful. Let's interrupt the text with an example of the impact of jumps on the realized variance measurement.

Example 1 *Consider the example of y_t realizations of the S&P 500 over one day, September 18th, 2007, on the left figure. Its returns are plotted on the right. When*



we compute the realized variation using all the data, it comes to 0.0119. Clearly, the

quadratic variation of the S&P 500 index from this day also contains the variation of the large jump. When we separate this jump and compute the realized variation without it, we arrive at an estimate of 0.0088. Thus we can see that ignoring the jump on this day would bias the true realized variance by 35.6%.

So, assuming the presence of jumps in the process, and considering the price process p_t to be driven, for example, by the jump-diffusion model represented by 2.10, seems to be much more appropriate. But then the $\langle p,p\rangle_t$ estimator, even if it is able to deal with the noise, will still contain the jump variation. Thus it is natural to separate the two components of the quadratic variation:

$$QV_{t,h} = \underbrace{\int_{t-h}^{t} \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h\leq s\leq t} J_s^2}_{\text{Jump Variation}} \quad . \tag{2.26}$$

Once we have the jumps separated from the process, we are able to determine $\langle p,p\rangle_t$ as well.

Barndorff-Nielsen and Shephard (2004b, 2006) develop a very powerful and complete way of detecting the presence of jumps in high-frequency data. The basic idea is to compare two measures of the integrated variance, one containing the jump variation and the other being robust to jumps and hence containing only the integrated variation part. In our work, we use the Andersen et al. (2011) adjustment of the original Barndorff-Nielsen and Shephard (2004b) estimator, which helps render it robust to certain types of microstructure noise.

Definition 6 *Bipower variation estimator*
The bipower variation over $[t-h,t]$*, for* $0\leq h\leq t\leq T$*, is defined by*

$$\widehat{RV}_{t,h}^{(BV)} = \mu_1^{-2}\frac{n}{n-2}\sum_{i=3}^{n} |r_{t-h+\left(\frac{i-2}{n}\right)h}|.|r_{t-h+\left(\frac{i}{n}\right)h}|, \tag{2.27}$$

where $\mu_a = \pi/2 = E(|Z|^a)$*, and* $Z\sim N(0,1)$*,* $a\geq 0$ *and* $\widehat{RV}_{t,h}^{(BV)} \to \int_{t-h}^{t}\sigma_s^2 ds$*.*

Thus $\widehat{RV}_{t,h}^{(BV)}$ provides a consistent estimator of the integrated variance and $\widehat{RV}_{t,h}^{(sparse)}$ provides a consistent estimator of the integrated variance plus the jump variation. Then, the jump variation can be estimated consistently as the difference between the realized variance and the realized bipower variation:

$$\text{plim}_{n\to\infty}(\widehat{RV}_{t,h}^{(sparse)} - \widehat{RV}_{t,h}^{(BV)}) = \sum_{l=1}^{N_t} J_{t,h,l}^2. \tag{2.28}$$

Under the assumption of no jump and some other regularity conditions, Barndorff-Nielsen and Shephard (2006) provided the joint asymptotic distribution of the jump variation.

Definition 7 *Jump detection test*
Under the null hypothesis of no within-day jumps,

$$Z_{t,h} = \frac{\frac{\widehat{RV}_{t,h}^{(sparse)} - \widehat{RV}_{t,h}^{(BV)}}{\widehat{RV}_{t,h}^{(sparse)}}}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\frac{1}{n}\max\left(1, \frac{\widehat{TQ}_{t,h}}{\left(\widehat{RV}_{t,h}^{(BV)}\right)^2}\right)}}, \tag{2.29}$$

where $\widehat{TQ}_{t,h} = n\mu_{4/3}^{-3}(\frac{n}{n-4})\sum_{j=5}^{n}|r_{t-h+\left(\frac{i-4}{n}\right)h}|^{4/3}|r_{t-h+\left(\frac{i-2}{n}\right)h}|^{4/3}|r_{t-h+\left(\frac{i-2}{n}\right)h}|^{4/3}$ *is asymptotically standard normally distributed.*

Using this theory, the contribution of the jump variation to the quadratic variation of the logarithmic price process 2.26 is measured by

$$J_{t,h} = I_{Z_{t,h}>\Phi_\alpha}\left(\widehat{RV}_{t,h}^{(sparse)} - \widehat{RV}_{t,h}^{(BV)}\right), \tag{2.30}$$

where $I_{Z_{t,h}>\Phi_\alpha}$ denotes the indicator function and Φ_α refers to the chosen critical value from the standard normal distribution. The measure of integrated variance is defined as

$$C_{t,h} = I_{Z_{t,h}\leq\Phi_\alpha}\widehat{RV}_{t,h}^{(sparse)} + I_{Z_{t,h}>\Phi_\alpha}\widehat{RV}_{t,h}^{(BV)}, \tag{2.31}$$

ensuring that the jump measure and the continuous part add up to the estimated variance without jumps.

Finally, Corsi et al. (2010) generalize this approach to threshold bipower variation, but we do not use this estimator, as we use the described jump detection methodology as the reference and we focus on wavelet methods for detecting jumps in the data, as described in the following section, where we present the complete theory using wavelets to generate an estimator of integrated variance which is unbiased and consistent in the presence of jumps and noise in the underlying process.

Chapter 3

Wavelet-based realized variation theory

The previous chapter considered the theory of quadratic return variation commonly used in literature to estimate the realized variance of returns from high-frequency data. In this chapter, we would like to generalize this theory and estimate the realized variation using wavelets. This chapter thus presents a complete framework for the estimation of quadratic variation using wavelet methods. The novelty of our approach lies in decomposition of the realized variance into the time-frequency space, which reveals new knowledge about the variation process. Another very appealing feature of wavelets is that they can be embedded into stochastic processes, as shown by Antoniou and Gustafson (1999). Thus we can conveniently use them to extend the theory of realized measures. One of the issues with the interpretation of wavelets in economic applications is that they behave like a filter. Thus wavelets can hardly be used for forecasting exercises most of the time. But in the realized measures, we only use wavelets to decompose the daily variation of the returns using intraday information, so they can be used very well for forecasting.

There have been several attempts to use wavelets in realized variation in the past few years. Hø g and Lunde (2003) were the first to suggest a wavelet estimator of realized variance. Capobianco (2004), for example, proposes to use a wavelet transform as a comparable estimator of quadratic variation. Subbotin (2008) uses wavelets to decompose variance into a multi-horizon scale. Next, Nielsen and Frederiksen (2008) compare the finite sample properties of three integrated variance estimators, i.e., realized variance, Fourier and wavelet estimators. They consider several processes generating time series with a long memory, jump processes as well as bid-ask bounce. Gençay et al. (2010) mention the possible use of wavelet multiresolution analysis to decompose realized variance in their paper, while they concentrate on developing much more complicated structures of variance modeling in different regimes through wavelet-domain hidden Markov models.

One remarkable exception in the theory is the work of Fan and Wang (2007), who were the first to use the wavelet-based realized variance estimator and also the methodology for the estimation of jumps from the data. In our work, we generalize the results of Fan

and Wang (2007) in several ways. Instead of using the Discrete Wavelet Transform we use the Maximum Overlap Discrete Wavelet Transform, which is a more efficient estimator and is not restricted to sample sizes that are powers of two. We also use the Daubechies family of wavelets instead of the Haar type in our work. Our contribution is in providing a complete theoretical framework for the estimation of wavelet realized variation, as this cannot be found in the current literature. Moreover, in the next chapters of this thesis, we will introduce a generalization of this approach to covariation estimation. In the next section, we will define the wavelet-based framework for the estimation of realized variance.

3.1 Wavelet decomposition of integrated variance

Wavelets and certain stochastic processes have a common structure. In fact, wavelet theory may be embedded in stochastic processes, as shown by Antoniou and Gustafson (1999), who compare wavelets with martingales and stochastic processes. In our analysis, we use Daubechies compactly supported wavelets. The original construction was first published in Daubechies (1988), while a detailed discussion about the Daubechies type of wavelets can be found in Daubechies (1992). The advantage of using Daubechies family filters is that they improve the frequency-domain characteristics of the Haar wavelet, but it can still be interpreted as generalized differences of adjacent averages. For more details see Daubechies (1988), Daubechies (1992) and Gençay et al. (2002). We also provide a brief introduction to Daubechies family wavelets in Appendix 11A.1.

For our analysis, we need to define the continuous wavelet transform (Daubechies, 1988):

Definition 8 *Continuous wavelet transform*
If $\psi \in L^2(\mathbb{R})$ satisfies the admissibility condition

$$C_\psi := \int_{\mathbb{R}} \left|\hat{\psi}(s)\right|^2 \frac{1}{|s|}\,\mathrm{d}s < +\infty, \tag{3.1}$$

where $\hat{}$ denotes the Fourier transform, then ψ is called a basic wavelet. Relative to every basic wavelet ψ, the continuous (integral) transform on $L^2(\mathbb{R})$ is defined by

$$(W_\psi f)(j,k) = \langle \psi_{j,k}, f\rangle = \mid j \mid^{-1/2} \int_{\mathbb{R}} \overline{\psi\left(\frac{s-k}{j}\right)} f(s)\mathrm{d}s \quad f \in L^2(\mathbb{R}), \tag{3.2}$$

where $\langle .,. \rangle$ defines the L^2-inner product and $j,k \in \mathbb{R}$ with $j \neq 0$.

Now we introduce the Calderón reconstruction formula (Chui, 1992)

Proposition 7 *Calderón reconstruction formula*
Let $\psi \in L^2(\mathbb{R})$ be a basic wavelet which defines a continuous wavelet transform $(W_\psi f)(j,k)$. Then for any $f \in L^2(\mathbb{R})$ and $s \in \mathbb{R}$ at which f is continuous,

$$f(s) = \frac{1}{C_\psi} \int_{\mathbb{R}} \int_{\mathbb{R}} (W_\psi f)(j,k)\psi_{j,k}(s)\frac{1}{j^2}\mathrm{d}k\mathrm{d}j. \tag{3.3}$$

Furthermore, let ψ satisfy the extra conditions

$$\int_0^{+\infty} \left|\hat{\psi}(s)\right|^2 \frac{1}{s}\,\mathrm{d}s = \int_0^{+\infty} \left|\hat{\psi}(-s)\right|^2 \frac{1}{s}\,\mathrm{d}s = \frac{1}{2}C_\psi. \tag{3.4}$$

Then

$$f(s) = \frac{2}{C_\psi}\int_0^{+\infty}\left[\int_{\mathbb{R}} (W_\psi f)(j,k)\psi_{j,k}(s)\,\mathrm{d}k\right]\frac{1}{j^2}\,\mathrm{d}j \tag{3.5}$$

for any $f \in L^2(\mathbb{R})$ and $s \in \mathbb{R}$ at which f is continuous.

For the proof, see Chui (1992).

The admissibility condition ensures that the Fourier transform of the wavelet $\hat{\psi}(s)$ has sufficient decay as $s \to 0$ (Daubechies, 1988). The finiteness of C_ψ is guaranteed if $\hat{\psi}(0) = 0$, which is equivalent to zero mean of the wavelet $\psi(.)$ (Mallat, 1998),

$$\hat{\psi}(0) = \int_{-\infty}^{\infty} \psi(s)ds = 0. \tag{3.6}$$

Further, we impose the unit energy condition on the wavelet $\psi(.)$

$$\int_{-\infty}^{\infty} |\psi(s)|^2 ds = 1 \tag{3.7}$$

Conditions 3.6 and 3.7 ensure that the wavelet has some non-zero terms, but all excursions away from zero must cancel out.

Theorem 1 *Let ψ be a Daubechies wavelet function D4. Then the extra conditions in Proposition 7*

$$\int_0^{+\infty} \left|\hat{\psi}(s)\right|^2 \frac{1}{s}\,\mathrm{d}s = \int_0^{+\infty} \left|\hat{\psi}(-s)\right|^2 \frac{1}{s}\,\mathrm{d}s = \frac{1}{2}C_\psi \tag{3.8}$$

are satisfied.

The **proof** is provided in Appendix 11A.3.

For more details about the continuous wavelet transform and the Calderón reconstruction formula see Mallat (1998), Calderón (1964), Mallat (1998), Daubechies (1988) and Najzar (2004)

.

Based on the proposed theory, we are able to extend the martingale representation theorem 2.5 using the continuous wavelet transform. Following the theorem will allow us to use the wavelet theory for the integrated variance decomposition.

Proposition 8 *Wavelet representation theorem*
For any univariate, square-integrable, continuous sample path, logarithmic price process $(p_t)_{t\in[0,T]}$ *which is not locally riskless, there exists a representation which can be decomposed using wavelets such that for all* $0 \leq h \leq t \leq T$

$$\begin{aligned} r_{t,h} &= \mu_{t.h} + M_{t,h} = \int_{t-h}^{t} \mu_s ds + \int_{t-h}^{t} \sigma_s dW_s \qquad (3.9) \\ &= \frac{2}{C_\psi} \int_{t-h}^{t} \int_{0}^{\infty} \int_{\mathbb{R}} \psi_{j,k}(s) \langle \psi_{j,k}, \mu_s \rangle dk \frac{1}{j^2} dj ds + \frac{2}{C_\psi} \int_{t-h}^{t} \int_{0}^{\infty} \int_{\mathbb{R}} \psi_{j,k}(s) \langle \psi_{j,k}, \sigma_s \rangle dk \frac{1}{j^2} dj dW_s, \end{aligned}$$

where μ_s *is an integrable, predictable and finite-variation stochastic process and* σ_s *is a strictly positive càdlàg stochastic process.* $\psi_{j,k} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ *represents the Daubechies (D4) wavelet function with a compact support.*

The **proof** is provided in the Appendix 11A.4.

Thus if we consider the example of stochastic volatility models represented by jump-diffusion model 2.10, its quadratic return variation over the $[t-h, t]$ time interval, $0 \leq h \leq t \leq T$, is:

$$QV_{t,h} = \underbrace{\int_{t-h}^{t} \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \leq s \leq t} J_s^2}_{\text{Jump Var.}} \qquad (3.10)$$

The quadratic return variation 3.10 can be decomposed using a Daubechies $D(4)$ wavelet as:

$$QV_{t,h} = \underbrace{\frac{2}{C_\psi} \int_{t-h}^{t} \int_{0}^{\infty} \int_{\mathbb{R}} \psi_{j,k}(s) \langle \psi_{j,k}, \sigma_s^2 \rangle dk \frac{1}{j^2} dj ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \leq s \leq t} J_s^2}_{\text{Jump Var.}}, \qquad (3.11)$$

where

$$\langle \psi_{j,k}, \sigma_s^2 \rangle = |j|^{-1/2} \int_{\mathbb{R}} \overline{\psi\left(\frac{s-k}{j}\right)} \sigma_s^2(s) ds \qquad (3.12)$$

Based on Proposition 8, a model-free measure of the integrated variation part may be proposed in analogy to the simple realized variance estimator. In order to be able to define the estimator, we need to define the tools for the discrete wavelet transform first.

3.2 Estimation of the realized variance using wavelets

The continuous wavelet transform is a very important concept which helps us with the derivation of theoretical behavior on the time-scale space. Since we work with real data, we need some form of sampling to compute the estimators, i.e., we have to use a suitable form of discretization. We will drop time for the explanation of the wavelet decomposition

of variance. As intraday returns will be decomposed, we define $X_j = p_{t-h+\frac{j}{n}} - p_{t-h+\frac{j+1}{n}}$ for all $j = 1, \ldots, n-1$ as in the intraday returns time series to be decomposed. The next section discusses a special form of discrete wavelet transformation, so from this point on we restrict the scale j and the translation k parameters to integers only.

3.2.1 The maximal overlap discrete wavelet transform

The maximal overlap discrete wavelet transformation (MODWT) is a translation-invariant type of discrete wavelet transformation, i.e., it is not sensitive to the choice of starting point of the examined process. Furthermore, the MODWT does not use a downsampling procedure as in the case of the discrete wavelet transform[1] (DWT), so the wavelet and scaling coefficient vectors at all levels (scales) have equal length. As a consequence, the MODWT is not restricted to sample sizes that are powers of two. This feature is very important for the analysis of real market data, since this limitation is usually too restrictive.

Both the DWT and the MODWT wavelet and scaling coefficients can be used for energy decomposition and analysis of variance. Conversely, the MODWT details and smooths cannot be used for such analysis. In the literature the MODWT is also called the algorithme à trous, the stationary wavelet transform, the translation invariant transform and the undecimated wavelet transform. It was first introduced by Holschneider et al. (1989). For more details about the MODWT see Mallat (1998), Percival and Walden (2000) and Gençay et al. (2002).

The MODWT is a very convenient tool for variance and energy analysis of a time series in the time-frequency domain. Percival (1995) demonstrates the advantages of the MODWT estimator of variance over the DWT estimator. Serroukh et al. (2000) analyze the statistical properties of the MODWT variance estimator for (locally) non-stationary and non-Gaussian processes.

3.2.2 Definition of MODWT filters

First, let us introduce the notion of MODWT scaling and wavelet filters on the j-th level, $\tilde{g}_{j,l}$ and $\tilde{h}_{j,l}$, as rescaled scaling and wavelet filters used in a simple DWT, i.e., $\tilde{g}_{j,l} \equiv g_{j,l}/2^{j/2}$ and $\tilde{h}_{j,l} \equiv h_{j,l}/2^{j/2}$. Both the scaling and wavelet filters have the same width, that is similar for the DWT filters as well:

$$L_j \equiv (2^j - 1)(L - 1) + 1, \tag{3.13}$$

where L denotes the length of a basic wavelet. For example, the Daubechies D(4) wavelet filter has length $L = 4$. There are three basic properties that both the MODWT filters must fulfill. Let us show these properties on the first level, $j = 1$:

$$\sum_{l=0}^{L-1} \tilde{h}_l = 0, \;\; \sum_{l=0}^{L-1} \tilde{h}_l^2 = 1/2, \;\; \sum_{l=-\infty}^{\infty} \tilde{h}_l \tilde{h}_{l+2n} = 0, \;\; n \in \mathbb{Z}_N, \tag{3.14}$$

[1] For a definition and detailed discussion of the discrete wavelet transform see Mallat (1998), Percival and Walden (2000) and Gençay et al. (2002)

and for the MODWT scaling filter:

$$\sum_{l=0}^{L-1} \tilde{g}_l = 1, \ \sum_{l=0}^{L-1} \tilde{g}_l^2 = 1/2, \ \sum_{l=-\infty}^{\infty} \tilde{g}_l \tilde{g}_{l+2n} = 0, \ \ n \in \mathbb{Z}_N. \tag{3.15}$$

Let's consider a time series $(X_i)_{i\in[0,N-1]}$ of N intraday returns. We obtain the MODWT wavelet and scaling coefficients on $i = 0, ..., N-1$ via circular filtering with the j-th level MODWT wavelet and scaling filters:

$$\widetilde{W}_{j,i} \equiv \sum_{l=0}^{L_i-1} \tilde{h}_{j,l} X_{i-l mod N}, \tag{3.16}$$

$$\widetilde{V}_{j,i} \equiv \sum_{l=0}^{L_i-1} \tilde{g}_{j,l} X_{i-l mod N}. \tag{3.17}$$

With periodization of the MODWT wavelet and scaling filters to length N we can write (Percival and Walden, 2000)):

$$\widetilde{W}_{j,i} = \sum_{l=0}^{N-1} \tilde{h}^{\circ}_{j,l} X_{i-l mod N}, \tag{3.18}$$

$$\widetilde{V}_{j,i} = \sum_{l=0}^{N-1} \tilde{g}^{\circ}_{j,l} X_{i-l mod N}. \tag{3.19}$$

The transfer function of a DWT filter $\{h_l\}$ at frequency f is defined via the Fourier transform as:

$$H(f) \equiv \sum_{l=-\infty}^{\infty} h_l e^{-i2\pi fl} = \sum_{l=0}^{L-1} h_l e^{-i2\pi fl}, \tag{3.20}$$

with the squared gain function

$$\mathcal{H}(f) \equiv |H(f)|^2. \tag{3.21}$$

The transfer functions of the j-th level MODWT wavelet and scaling filter are given as follows:

$$\widetilde{H}_j(f) \equiv \widetilde{H}_1(2^{j-1}f) \prod_{l=0}^{j-2} \widetilde{G_1}(2^l f), \tag{3.22}$$

$$\widetilde{G}_j(f) \equiv \prod_{l=0}^{j-1} \widetilde{G_1}(2^l f). \tag{3.23}$$

3.2.3 Energy decomposition of a stochastic process

For our analysis, it is important to show that we are able to decompose the energy of a stochastic process on a scale-by-scale basis, i.e., we can get the energy contribution of every level j, with the maximum level of decomposition $J \leq \log_2 N$.

Proposition 9 *Energy decomposition in discrete time*
The energy of the time series X_i, $i = 1, \ldots, N-1$ can be decomposed on a scale-by-scale basis $J \leq \log_2 N$ so that

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_J\|^2 \tag{3.24}$$

where $\|\mathbf{X}\|^2 = \sum_{i=0}^{N-1} X_i^2$, $\|\widetilde{\mathbf{W}}_j\|^2 = \sum_{i=0}^{N-1} W_{j,i}^2$, $\|\widetilde{\mathbf{V}}_J\|^2 = \sum_{i=0}^{N-1} V_{J,i}^2$ and $\widetilde{\mathbf{W}}_j$ and $\widetilde{\mathbf{V}}_j$ are N dimensional vectors of the j-th level MODWT wavelet and scaling coefficients.

The proof of the energy decomposition 3.24 using the MODWT can be found in Appendix 11A.5.

It is worth noting that the squared norm $\|.\|$ is similar to the realized measure discussed in the preceding sections. For example, in the case of the realized variance estimator (RV) the energy decomposition can reveal the contributions of particular scales to the overall energy, hence we can see what form this realized measure takes.

For simplicity in notation let us define a vector $\widetilde{\mathcal{W}}$ that consists of $J_s = J + 1$ $N-$dimensional subvectors, where the first J subvectors are the MODWT wavelet coefficients at levels $j = 1, ..., N$ and the last subvector consists of the MODWT scaling coefficients at level J:

$$\widetilde{\mathcal{W}} = \begin{bmatrix} \widetilde{\mathbf{W}}_1 \\ \widetilde{\mathbf{W}}_2 \\ \ldots \\ \widetilde{\mathbf{W}}_J \\ \widetilde{\mathbf{V}}_J \end{bmatrix}, \tag{3.25}$$

i.e., for Equation 3.24 the following holds:

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_J\|^2 = \sum_{j=1}^{J_s+1} \|\widetilde{\mathcal{W}}_j\|^2 \tag{3.26}$$

3.2.4 Wavelet variance

For a real-valued covariance stationary stochastic process X_i with mean zero, the sequence of the MODWT wavelet coefficients $\widetilde{W}_{(X)j,i}$, for all $j, i > 0$ unaffected by the boundary conditions, obtained by the wavelet decomposition at scale j is also a stationary process

with mean zero. The variance of the wavelet coefficients at scale j is the wavelet variance, i.e.,

$$\nu^2_{(X)j} = var(\widetilde{W}_{(X)j,i}) \tag{3.27}$$

While the variance of a covariance stationary process X_i is equal to the integral of the spectral density function $S_X(.)$, the wavelet variance at a particular level j is the variance of the wavelet coefficients $(\widetilde{W}_{(X)j,i})$ with spectral density function $S_{(X)j}(.)$:

$$\nu^2_{(X)j} = \int_{-1/2}^{1/2} S_{(X)j}(f)df = \int_{-1/2}^{1/2} \widetilde{\mathcal{H}}_j(f) S_{(X)}(f)df, \tag{3.28}$$

where $\widetilde{\mathcal{H}}_j(f)$ is the squared gain function of the wavelet filter $\tilde{h}_j$ (Percival and Walden, 2000). Since the variance of a process X_i is the sum of the contributions of the variances at all scales we can write:

$$var(X_i) = \sum_{j=1}^{\infty} \nu^2_{(X)j} \tag{3.29}$$

However, for a finite number of scales we have:

$$var(X_i) = \int_{-1/2}^{1/2} S_{(X)}(f)df = \sum_{j=1}^{J} \nu^2_{(X)j} + var(\widetilde{V}_{(X)J,i}) \tag{3.30}$$

3.2.5 Wavelet-based realized variance estimator

We have introduced all the theory needed to return to the estimation of the realized variance, so we can proceed to defining our wavelet-based estimators in this section.

Definition 9 *Wavelet-based realized variance*
The wavelet-based realized variance over $[t-h, t]$, for $0 \leq h \leq t \leq T$, is defined by

$$\widehat{RV}^{(WRV)}_{t,h} = \sum_{j=1}^{J_s+1} \sum_{k=1}^{n} \widetilde{\mathcal{W}}^2_{j,t-h+\frac{k}{n}h}, \tag{3.31}$$

where n is the number of intraday observations in $[t-h, t]$ and J_s is the number of scales we consider. $\widetilde{\mathcal{W}}_{j,t-h+\frac{k}{n}h}$ are the MODWT coefficients defined in 3.26 on returns data $r_{t,h}$ on scales $j = 1, \ldots, J_s + 1$, where $J_s \leq \log_2 n$.

Proposition 10 *Wavelet-based realized variance as an unbiased variance estimator*
If the return process is square-integrable and $\mu_t \equiv 0$, then for any value of $n \geq 1$,

$$E\left[RV_{t,h}|\mathcal{F}_t\right] = E\left[\widehat{RV}^{(WRV)}_{t,h}|\mathcal{F}_t\right]. \tag{3.32}$$

Proposition 11 *Consistency of wavelet-based realized variance*
The wavelet realized variance provides a consistent nonparametric measure of the variance,

$$\text{plim}_{n\to\infty} \widehat{RV}_{t,h}^{(WRV)} = RV_{t,h}, \quad 0 \leq h \leq t \leq T, \tag{3.33}$$

where the convergence is uniform in probability.

The proof of Proposition 10 and Proposition 11 is provided in Appendix 11A.6.
Consequently, Proposition 5 can be rewritten using this result to:

$$r_{t,h}|\sigma \left\{ \mu_{t,h}, \widehat{RV}_{t,h}^{(WRV)} \right\} \sim N\left(\mu_{t,h}, \widehat{RV}_{t,h}^{(WRV)} \right). \tag{3.34}$$

The wavelet realized variance estimator decomposes the realized variance. Thus it is unbiased estimator and with increasing sampling frequency $n \to \infty$ is also a consistent estimator of the integrated variance part of the process described by Eq. 2.6. Still, under the assumption of the presence of noise (see Proposition 6), as well as jumps, in the data, both the wavelet realized volatility and the realized volatility are biased. Therefore, in the following section we will introduce the concept of treating jumps using wavelets. This will be the last step in proposing the final estimator, which will be robust to jumps as well as microstructure noise.

3.2.6 Estimation of the wavelet variance

Following Percival and Walden (2000) we define the unbiased MODWT wavelet variance estimator for a covariance stationary Gaussian process X_i (or a covariance stationary process after $d-$th backward differences), $L \geq 2d$, at level j as:

$$\hat{\nu}^2_{(X)j} \equiv \frac{1}{M_j} \sum_{i=L_j-1}^{N-1} \widetilde{W}^2_{(X)j,i} \tag{3.35}$$

where $M_j = N - L_j + 1 > 0$ is the number of j-th level MODWT coefficients unaffected by the boundary conditions. If we take all the MODWT wavelet coefficients N we obtain a biased MODWT variance estimator. However, as $N \to \infty$ the ratio $\frac{N}{M_j}$ goes to unity, so consequently the estimators gives more or less the same results.

The estimator $\hat{\nu}^2_{(X)j}$ is also a random variable, so it is of interest to know how close it is to the real value of the wavelet variance $\nu^2_{(X)j}$. We assume that the sequences of wavelet coefficients $\widetilde{W}_{(X)j,i}$ are normally distributed random variables with zero mean and spectral distribution function $S_{(X)j}(f)$. Following the result of Percival (1995), if $S_{(X)j}(f) > 0$ almost everywhere and if

$$\int_{-1/2}^{1/2} S^2_{(X)j}(f) df < \infty, \tag{3.36}$$

then the MODWT variance estimator $\hat{\nu}^2_{(X)j}$ is an unbiased and asymptotically normally distributed estimator with large sample variance $2\int_{-1/2}^{1/2} S^2_{(X)j}(f)df/M_j$ – see also Percival and Walden (2000) and Serroukh et al. (2000). Thus, for large enough M_j, we can write

$$\frac{\sqrt{M_j}\left(\hat{\nu}^2_{(X)j} - \nu^2_{(X)j}\right)}{\sqrt{2\int_{-1/2}^{1/2} S^2_{(X)j}(f)df}} =^d N(0,1). \tag{3.37}$$

Note also that Serroukh et al. (2000) derived the asymptotic distribution of the MODWT wavelet variance estimator for other classes of processes, not only Gaussian and linear ones.

An interesting comparison of the DWT and MODWT wavelet variance estimators is discussed in Percival (1995). Asymptotic relative efficiency is used to compare the DWT with the MODWT wavelet variance estimator at the first level ($j = 1$). Further, it is assumed that the first-level wavelet coefficient sequence $(\widetilde{W}_{(X)1,i})$ has a square summable spectral density function $S_{(X)1}(f)$, for which $S_{(X)1}(f) > 0$ holds almost everywhere. The DWT and MODWT variance estimators at the first level, denoted as $\hat{\nu}^2_{(D)1}$ and $\hat{\nu}^2_{(M)1}$, are then asymptotically normally distributed with mean ν^2_1. The asymptotic relative efficiency of the DWT estimator with respect to the MODWT estimator is defined as follows:

$$\begin{aligned} e(\hat{\nu}_{(D)1}, \hat{\nu}_{(M)1}) &\equiv \lim_{N\to\infty} \frac{var(\hat{\nu}_{(M)1})}{var(\hat{\nu}_{(D)1})} \\ &= \frac{\int_{-1/2}^{1/2} S^2_{(X)1}(f)df}{\int_{-1/2}^{1/2} S^2_{(X)1}(f)df + \int_{-1/2}^{1/2} S_{(X)1}\left(\frac{f}{2}+\frac{1}{2}\right)df} \end{aligned} \tag{3.38}$$

Identity 3.38 clearly shows that the DWT variance estimator cannot be more efficient than the MODWT estimator since we assume that any spectral density function is nonnegative. In some cases, we can obtain a significant reduction in the large sample variance by using the MODWT estimator. Using large Monte Carlo simulations Percival (1995) show, for example, that for a white noise process the asymptotic relative efficiency of the DWT variance estimator with respect to the MODWT variance estimator using the Daubechies $D(4)$ is 0.82, i.e., the variance of the MODWT-based estimator is significantly lower.

3.3 Realized jump estimation using wavelets

As discussed earlier, the presence of jumps in the process generating stock prices is needed to describe the real-world data well. This is commonly done by considering the price process $(p_t)_{t\in[0,T]}$ to be driven, for example, by the jump-diffusion model represented by 2.10. But then the $\langle p,p\rangle_t$ estimator, although being able to deal with the noise, will

still contain the jump variation. Thus it is natural to separate the two components of the quadratic variation:

$$QV_{t,h} = \underbrace{\int_{t-h}^{t} \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h \le s \le t} J_s^2}_{\text{Jump Variation}} \quad . \tag{3.39}$$

Once we have the jumps separated from the process, we are able to determine the $\langle p, p \rangle_t$.

Wavelets can also be used for estimating jumps and separating integrated variance from jump variation. Fan and Wang (2007) were the first to show that by using the wavelet approach we can estimate jumps in high-frequency data consistently. We follow their approach and present the jump detection framework.

Definition 10 *Jump estimation using wavelets*
Let $\widetilde{\mathcal{W}}_{1,k}$ be the 1^{st} level wavelet coefficients of y_t over $[t-h, t]$ from 6. If for some $\widetilde{\mathcal{W}}_{1,k}$

$$|\widetilde{\mathcal{W}}_{1,k}| > \frac{median\{|\widetilde{\mathcal{W}}_{1,k}|, k = 1, \ldots, n\}}{0.6745} \sqrt{2 \log n}, \tag{3.40}$$

then $\hat{\tau}_l = \{k\}$ is the estimated jump location with size $\bar{y}_{\hat{\tau}_{l+}} - \bar{y}_{\hat{\tau}_{l-}}$ (averages over $[\hat{\tau}_l, \hat{\tau}_l + \delta_n]$ and $[\hat{\tau}_l, \hat{\tau}_l - \delta_n]$, respectively, with $\delta_n > 0$ being the small neighborhood of the estimated jump location $\hat{\tau}_l \pm \delta_n$; 0.6745 is a robust estimate of the standard deviation).

The jump variation is then estimated by the sum of the squares of all the estimated jump sizes:

$$\widehat{WJV} = \sum_{l=1}^{N_t} (\bar{y}_{\hat{\tau}_{l+}} - \bar{y}_{\hat{\tau}_{l-}})^2. \tag{3.41}$$

The sample path of p_t has a finite number of jumps ($a.s.$). Fan and Wang (2007) apply wavelet jump detection to the deterministic functions with $i.i.d.$ additive noise ϵ_t of Wang (1995) and Raimondo (1998). Following these references, we can make the following proposition.

Proposition 12 *Suppose that the sample path of p_t has $q = N_1 < \infty$ jumps at $\tau_1, \ldots, \tau_q$,*

$$\lim_{n \to \infty} P\left(\hat{q} = q, \sum_{l=1}^{q} |\hat{\tau}_l - \tau_l| \le n^{-1} log^2 n \;\middle|\; X \right) = 1, \tag{3.42}$$

where $\hat{q} = N_t$ is the number of estimated jumps with locations $\hat{\tau}_q$.

Proposition 13 *Consistency of the wavelet jump estimator*
With $n \to \infty$

$$\text{plim}_{n \to \infty} \widehat{WJV} = \sum_{l=1}^{N_t} J_l^2, \tag{3.43}$$

with the convergence rate $N^{-1/4}$.

Thus we are able to estimate the jump variation from the process consistently. In the following analysis, we will be able to separate the continuous part of the price process containing noise from the jump variation. This result can be found in Fan and Wang (2007) and it states that the jump-adjusted process $y^{(J)} = y_t - \widehat{WJV}$ converges in probability to the continuous part without jumps, thus its variation is exactly the first part of 3.39, the integrated variance. Thus, if we are able to deal with the noise in $y^{(J)}$, we will be able to estimate the *true* $\langle p, p \rangle_t$.

3.4 Wavelet-based estimator robust to jumps and noise

Finally, let us propose an estimator that will be able to estimate jumps from the process consistently. With $n \to \infty$, it will be able to recover the true integrated variance from noisy data. Moreover, it will decompose the integrated variance into J_s components. In the final estimator, we will utilize what we already know: the TSRV estimator of Zhang et al. (2005), the Wavelet Representation Theorem from Proposition 8 and the jump detection method proposed by Definition 10.

Definition 11 *Jump wavelet TSRV (JWTSRV) estimator*
Let $\widehat{RV}_{t,h}^{(estimator,J)}$ *denote an estimator of realized variance over* $[t-h, t]$*, for* $0 \leq h \leq t \leq T$*, on the jump-adjusted observed data,* $y_{t,h}^{(J)} = y_{t,h} - \sum_{l=1}^{N_t} J_l$*. The jump-adjusted wavelet two-scale realized variance estimator is defined as:*

$$\widehat{RV}_{t,h}^{(JWTSRV)} = \widehat{RV}_{t,h}^{(WRV,J)} - \frac{\bar{n}}{n}\widehat{RV}_{t,h}^{(all,J)}, \tag{3.44}$$

where $\widehat{RV}_{t,h}^{(WRV,J)} = \frac{1}{G}\sum_{g=1}^{G}\sum_{j=1}^{J_s+1}\sum_{k=1}^{n}\widetilde{\mathcal{W}}_{j,t-h+\frac{k}{n}h}^2$ *obtained from wavelet coefficient estimates on a grid of size* $\bar{n} = n/G$ *on the jump-adjusted observed data,* $y_{t,h}^{(J)} = y_{t,h} - \sum_{l=1}^{N_t} J_l$*.*

Proposition 14 *JWTSRV unbiased variance estimator*
If the return process is square-integrable and $\mu_t \equiv 0$*, then for any value of* $n \geq 1$*,*

$$E\left[RV_{t,h}|\mathcal{F}_t\right] = E\left[\widehat{RV}_{t,h}^{(JWTSRV)}|\mathcal{F}_t\right]. \tag{3.45}$$

Proposition 15 *Consistency of JWTSRV*
The wavelet realized variance provides a consistent nonparametric measure of the variance,

$$\text{plim}_{n\to\infty}\,\widehat{RV}_{t,h}^{(JWTSRV)} = RV_{t,h}, \quad 0 \leq h \leq t \leq T, \tag{3.46}$$

where the convergence is uniform in probability.

The **proof** of Proposition 14 and Proposition 15 is provided in Appendix 11A.7. The JWTSRV estimator decomposes the realized variance into an arbitrary chosen number

of investment horizons and jumps. Thus it is unbiased estimator and with increasing sampling frequency $n \to \infty$ is also a consistent estimator of the integrated variance part of the process described by Eq. 2.10 as it converges in probability to the *true* integrated variance $\langle p, p \rangle_t$ of the process p_t.

Thus we have defined a wavelet-based realized variation theory which is able to estimate realized variance consistently in the presence of noise and jumps. In the next sections, we will test the small sample performance of the estimators and perform an empirical study on real-world data.

In small samples, a small sample refinement can be constructed similarly as in 2.23:

$$\widehat{RV}_{t,h}^{(JWTSRV,adj)} = \left(1 - \frac{\bar{n}}{n}\right)^{(-1)} \widehat{RV}_{t,h}^{(JWTSRV)}. \tag{3.47}$$

As noted in the previous part, when referring to the realized volatility estimated using our JWTSRV estimator, we will refer to the $\sqrt{\widehat{RV}_{t,h}^{(JWTSRV,adj)}}$.

Chapter 4

Numerical study of the small sample performance of the estimators

In this section, we study the small sample performance of estimators using Monte Carlo simulations designed to capture the real nature of the data. We use several experiments using different volatility models, including a fractional stochastic volatility model capturing long memory in volatility. The main purpose of the study is to show that the proposed methodology is robust to noise and jumps under various settings. Our Monte Carlo study is quite large and brings interesting new results to the literature. Each experiment compares the performance of the realized variation estimator defined by 2.12, the bipower variation estimator defined by 2.27, the two-scale realized volatility defined by 2.21, the realized kernel defined by 2.24, and the jump wavelet two-scale realized variation defined by 3.44. All the estimators are adjusted for small sample bias, similarly to 2.23 and 3.47. For convenience, we refer to the estimators in the description of the results as RV, BV, TSRV, RK and JWTSRV, respectively. Moreover, we also compare the minimum variance estimators TSRV* and JWTSRV*, which minimize the total asymptotic variance from 2.22.

4.1 Jump-diffusion model with stochastic volatility

The first data generating model we assume in our study is a one-factor jump-diffusion model with stochastic volatility, described by the following equations:

$$\begin{aligned} dX_t &= (\mu - \sigma_t^2/2)dt + \sigma_t dW_{x,t} + c_t dN_t \\ d\sigma_t^2 &= \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dW_{y,t}, \end{aligned} \tag{4.1}$$

where W_x and W_y are standard Brownian motions with correlation ρ, and $c_t dN_t$ is a compound Poisson process with random jump size distributed as $N \sim (0, \sigma_J)$. We set

Table 4.1: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 10,000 simulations of jump-diffusion model with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	0.90 (0.65)	-4.13 (0.82)	-6.03 (0.43)	-0.28 (0.02)	-15.18 (2.51)	-6.08 (0.43)	-0.37 (0.02)
ϵ_2	100.10 (0.93)	97.36 (1.18)	-5.25 (0.45)	0.98 (0.51)	-4.40 (2.63)	-3.86 (0.45)	2.29 (0.52)
ϵ_3	394.14 (2.10)	412.43 (2.87)	-5.15 (0.45)	-1.31 (0.90)	19.66 (2.91)	0.19 (0.48)	3.95 (0.93)
ϵ_4	885.81 (5.40)	949.39 (8.00)	-4.52 (0.43)	-0.47 (1.34)	52.94 (3.13)	7.71 (0.58)	11.93 (1.48)
				One Jump			
ϵ_1	247.73 (19.31)	53.84 (1.85)	236.63 (18.64)	245.55 (18.09)	225.41 (23.19)	-5.64 (0.44)	-0.25 (0.02)
ϵ_2	354.79 (20.91)	164.24 (2.77)	246.24 (19.67)	253.69 (19.61)	241.88 (23.10)	-0.35 (0.48)	4.36 (0.52)
ϵ_3	648.69 (23.12)	495.58 (5.15)	241.06 (19.79)	251.24 (20.44)	260.10 (25.62)	18.12 (0.64)	23.94 (1.10)
ϵ_4	1139.00 (27.54)	1044.80 (10.79)	248.00 (20.30)	256.50 (21.02)	303.39 (25.25)	58.29 (1.41)	64.39 (2.29)
				Two Jumps			
ϵ_1	503.32 (41.12)	117.87 (3.84)	489.24 (39.47)	501.61 (38.99)	471.67 (47.36)	-5.27 (0.43)	-0.36 (0.02)
ϵ_2	616.80 (41.99)	237.65 (4.56)	500.37 (39.51)	513.15 (39.69)	489.82 (45.65)	3.43 (0.49)	7.41 (0.54)
ϵ_3	910.28 (44.71)	582.94 (7.67)	499.52 (39.83)	508.95 (39.52)	517.36 (48.27)	38.99 (0.81)	43.39 (1.25)
ϵ_4	1398.40 (47.55)	1160.20 (15.04)	496.34 (39.15)	505.27 (38.93)	551.50 (47.75)	108.73 (2.34)	113.95 (3.06)
				Three Jumps			
ϵ_1	772.53 (62.38)	191.00 (6.58)	753.28 (60.11)	766.80 (58.86)	730.70 (72.17)	-5.62 (0.46)	-0.37 (0.02)
ϵ_2	858.07 (61.60)	312.01 (7.34)	741.10 (58.62)	759.90 (58.56)	720.73 (68.89)	6.04 (0.51)	10.21 (0.53)
ϵ_3	1169.30 (68.71)	671.31 (10.71)	756.73 (61.89)	767.36 (60.72)	769.49 (74.86)	59.15 (0.95)	61.90 (1.37)
ϵ_4	1650.50 (69.52)	1257.80 (18.55)	742.31 (58.93)	757.31 (59.37)	787.06 (71.48)	160.10 (3.19)	167.24 (3.94)

the parameters to values which are reasonable for a stock price, as in Zhang et al. (2005), who used model 4.1 without jumps, $\mu = 0.05$, $\alpha = 0.04$, $\kappa = 5$, $\gamma = 0.5$, $\rho = -0.5$ and $\sigma_J = 0.025$. The volatility parameters satisfy Feller's condition $2\kappa\alpha \geq \gamma^2$, which keeps the volatility process away from the zero boundary. We generate $10,000$ independent sample paths[1] of the process using the Euler scheme at a time interval of $\delta = 1s$, each with $6.5 \times 60 \times 60$ steps $n = 23,400$, corresponding to a 6.5 trading hour day. On each simulated path, we estimate $\widehat{\langle p, p \rangle}_t$ over $t = 1$ day, as the parameter values are annualized (i.e., $t = 1/252$). The results are computed for sampling of 5 minutes (M=78) for RV, BV, TSRV, RK and JWTSRV, as well as for the optimal sampling frequency found by minimizing the total asymptotic variance from 2.22 for TSRV* and JWTSRV*.

We repeat the simulation with different levels of noise as well as different numbers of jumps. We assume that the market microstructure noise, ϵ_t, comes from a Gaussian distribution with different standard deviations: $(E[\epsilon^2])^{1/2} = \{0, 0.0005, 0.001, 0.0015\}$. Thus, the first simulated model, $(E[\epsilon^2])^{1/2} = 0$, has zero noise. The remaining three models have levels of microstructure noise corresponding to 0.05%, 0.1% and 0.15% of the value of the asset price.

Moreover, we add different amounts of jumps, controlled by intensity λ from the Poisson process $c_t dN_t$. We start with $\lambda = 0$, with model 4.1 reducing to a modification of the standard Heston volatility model without jumps, and continue with jump coefficients implying up to three jumps per day in the process. This number is realistic according to findings in the literature. The size of the jumps is controlled by parameter σ_J, which is set to 0.025, implying that a one standard deviation jump changes the price level by 2.5%. Finally, we have 16 models with different levels of noise and numbers of jumps,

[1]For comparison, we also include the results based on 1,000 generated independent sample paths in the Table 11.1 in the Appendix 11B.

Table 4.2: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 10,000 simulations of fractional stochastic volatility model with Hurst parameter $H = 0.5$ with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	7.65 (10.51)	-19.57 (13.32)	-26.55 (6.86)	-1.16 (0.23)	-66.40 (39.16)	-26.80 (6.86)	-1.48 (0.23)
ϵ_2	104.62 (11.14)	82.08 (14.41)	-26.79 (6.70)	-0.41 (0.86)	-59.74 (38.38)	-25.51 (6.71)	0.74 (0.86)
ϵ_3	407.48 (15.07)	383.91 (19.41)	-23.47 (6.71)	-0.98 (1.45)	-20.79 (41.69)	-18.27 (6.74)	4.32 (1.48)
ϵ_4	896.20 (22.25)	888.97 (29.63)	-25.28 (6.85)	-5.32 (2.23)	19.05 (44.65)	-13.75 (7.04)	6.14 (2.36)
				One Jump			
ϵ_1	254.70 (32.31)	97.88 (18.00)	219.71 (27.19)	249.51 (19.81)	167.85 (67.92)	-27.29 (6.69)	-1.42 (0.25)
ϵ_2	356.65 (32.73)	196.24 (19.36)	219.05 (26.04)	247.03 (18.80)	184.96 (67.27)	-20.04 (6.68)	4.05 (0.84)
ϵ_3	654.63 (37.40)	507.79 (24.60)	222.84 (27.44)	249.36 (20.25)	213.83 (70.42)	1.88 (7.19)	24.29 (1.64)
ϵ_4	1151.80 (45.69)	1026.90 (36.71)	226.66 (27.63)	251.67 (21.35)	266.50 (73.94)	39.10 (8.15)	60.25 (3.10)
				Two Jumps			
ϵ_1	510.21 (53.31)	217.50 (22.70)	470.75 (47.16)	505.47 (38.33)	411.80 (97.74)	-25.56 (6.75)	-0.42 (0.26)
ϵ_2	611.27 (57.48)	317.64 (24.09)	471.07 (49.70)	506.63 (40.45)	424.19 (101.42)	-20.62 (6.88)	5.74 (0.87)
ϵ_3	914.79 (60.52)	636.31 (30.92)	476.78 (49.28)	505.09 (40.70)	466.31 (103.95)	21.00 (7.38)	42.32 (1.79)
ϵ_4	1396.70 (67.41)	1155.20 (42.77)	474.58 (47.27)	504.16 (40.06)	506.18 (103.54)	93.30 (9.22)	117.05 (4.00)
				Three Jumps			
ϵ_1	765.95 (78.40)	346.13 (28.96)	719.80 (69.95)	750.18 (57.99)	670.56 (134.88)	-23.75 (6.88)	-1.96 (0.26)
ϵ_2	855.63 (76.82)	436.22 (29.67)	713.92 (66.84)	750.49 (58.47)	666.32 (127.53)	-15.91 (6.74)	9.47 (0.88)
ϵ_3	1161.90 (81.72)	762.38 (37.14)	721.15 (68.21)	758.76 (58.42)	705.35 (134.87)	35.08 (7.65)	61.87 (1.96)
ϵ_4	1662.10 (95.44)	1299.40 (52.60)	722.50 (69.09)	758.79 (59.70)	746.30 (136.93)	135.71 (10.19)	162.66 (4.72)

and we compare the bias of all the estimators for each simulated day.

Table 4.1 shows the results. The first model, without jumps, corresponds to the findings of Zhang et al. (2005) and Aït-Sahalia and Mancini (2008), although we add a higher level of noise to the simulations as suggested by the literature. The results show how robust the TSRV-based and RK estimators are to an increase in noise. Even a small increase in the magnitude of noise causes large bias in the other estimators, but the TSRV-based and RK estimators contain bias of order less than 10^{-4}. What we add to the original results of Zhang et al. (2005) and Aït-Sahalia and Mancini (2008) are jumps. While TSRV and RK are robust to an increase in noise, they are not robust to an increase in jumps at all. From the rest of the results, we can see nicely how the wavelets detect all of the jumps in the process and the JWTSRV stays unbiased. From the results we can also see that with a mixture of relatively high noise and a large number of jumps in the process even the JWTSRV estimator suffers from bias. This suggests that jumps are sometimes indistinguishable from noise and remain undetected under the large noise. We can also see that the BV is able to deal with jumps to some extent, but is hurt heavily by noise.

4.2 Fractional stochastic volatility model

Empirical evidence suggests that the volatility process may exhibit long memory. Previous models approximate this behavior, but a much more powerful class of models designed to capture long memory is known by the literature, namely, fractional Brownian motion. Instead of describing the solution and method of simulation of this class of models here, we rather point the interested reader to Comte and Renault (1999) and Marinucci and

Table 4.3: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 10,000 simulations of fractional stochastic volatility model with Hurst parameter $H = 0.7$ with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	9.47 (10.57)	-14.18 (13.44)	-25.81 (6.87)	-0.62 (0.24)	-61.83 (39.91)	-26.17 (6.86)	-0.94 (0.23)
ϵ_2	106.09 (11.24)	78.59 (14.57)	-22.93 (6.73)	-0.29 (0.84)	-49.16 (39.28)	-21.66 (6.75)	0.86 (0.84)
ϵ_3	404.06 (14.44)	380.66 (18.75)	-23.64 (6.79)	-1.01 (1.45)	-13.44 (43.10)	-17.93 (6.88)	4.50 (1.48)
ϵ_4	899.67 (22.67)	895.53 (29.96)	-21.95 (6.89)	-1.66 (2.19)	32.94 (45.65)	-9.40 (7.12)	10.72 (2.33)
				One Jump			
ϵ_1	260.24 (32.42)	99.77 (17.58)	226.07 (27.77)	252.00 (19.93)	175.24 (71.81)	-24.61 (6.75)	-0.66 (0.24)
ϵ_2	361.23 (33.51)	204.48 (19.56)	222.55 (26.96)	250.42 (19.87)	194.31 (70.22)	-20.36 (6.68)	3.47 (0.85)
ϵ_3	658.78 (36.67)	507.47 (24.81)	229.15 (26.77)	253.33 (20.29)	221.70 (71.31)	1.16 (7.28)	21.96 (1.62)
ϵ_4	1140.50 (47.95)	1014.50 (37.09)	221.27 (28.05)	248.39 (22.10)	260.55 (74.86)	35.43 (8.07)	61.22 (3.19)
				Two Jumps			
ϵ_1	514.66 (55.01)	219.27 (23.17)	473.71 (48.22)	503.64 (39.76)	430.23 (100.78)	-23.00 (6.69)	-1.45 (0.24)
ϵ_2	615.38 (57.57)	318.85 (24.74)	481.64 (49.01)	508.26 (39.87)	453.16 (102.60)	-14.61 (6.95)	5.80 (0.87)
ϵ_3	903.32 (59.69)	630.21 (30.51)	470.80 (47.55)	498.66 (39.14)	467.10 (102.37)	20.01 (7.23)	41.69 (1.78)
ϵ_4	1400.90 (66.50)	1164.00 (43.24)	467.00 (46.26)	505.48 (39.79)	500.94 (102.72)	86.73 (9.19)	115.27 (4.02)
				Three Jumps			
ϵ_1	765.72 (78.57)	340.76 (28.99)	720.49 (70.78)	754.51 (59.35)	676.34 (135.14)	-28.45 (6.80)	-1.85 (0.25)
ϵ_2	873.97 (79.58)	452.01 (30.08)	731.76 (70.61)	765.04 (59.59)	682.13 (134.89)	-12.12 (6.85)	12.12 (0.88)
ϵ_3	1164.00 (82.53)	767.45 (36.59)	718.24 (67.72)	752.01 (58.67)	704.64 (132.81)	38.63 (7.86)	63.43 (1.96)
ϵ_4	1663.50 (91.73)	1299.90 (48.60)	731.80 (69.58)	758.67 (59.03)	756.10 (138.55)	141.26 (10.54)	161.96 (4.84)

Robinson (1999) for more details.

In our simulations, we use the fractional jump-diffusion model:

$$\begin{aligned} dX_t &= (\mu - \sigma_t^2/2)dt + \sigma_t dW_{x,t} + c_t dN_t \\ d\sigma_{H,t}^2 &= \kappa(\alpha - \sigma_{H,t}^2)dt + \gamma dW_{H,t}, \end{aligned} \tag{4.2}$$

where W_x is a standard Brownian motion, $dW_{H,t}$ is a fractional Brownian motion (FBM) with Hurst parameter $H \in (0, 1]$ and $c_t dN_t$ is a compound Poisson process with random jump size distributed as $N \sim (0, \sigma_J)$. We set the parameters to values $\mu = 0.05$, $\alpha = 0.2$, $\kappa = 20$, $\gamma = 0.012$ and $\sigma_J = 0.025$ as in Aït-Sahalia and Mancini (2008), although these authors use a process without jumps.

We generate $10,000$ independent sample paths[2] of the process using the Euler scheme at a time interval of $\delta = 1s$, each with $6.5 \times 60 \times 60$ steps $n = 23,400$, corresponding to 6.5 trading hours. The results are computed for sampling of 5 minutes (M=78) for RV, BV, TSRV, RK and JWTSRV, as well as for the optimal sampling frequency found by minimizing the total asymptotic variance from 2.22 for TSRV* and JWTSRV*. We again repeat the simulation with different levels of noise as well as different numbers of jumps. We assume that the market microstructure noise, ϵ_t, comes from a Gaussian distribution with different standard deviations: $(E[\epsilon^2])^{1/2} = \{0, 0.0005, 0.001, 0.0015\}$, and we again start without jumps, and continue with jump coefficients implying up to three jumps per day in the process. Finally, we have 16 models with different levels of noise and numbers of jumps, and we compare the bias of all the estimators for each simulated day on three processes with different long memory parameters.

[2]For comparison, we also include the results based on 1,000 generated independent sample paths in

Table 4.4: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 10,000 simulations of fractional stochastic volatility model with Hurst parameter $H = 0.9$ with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	5.90 (10.50)	-19.58 (13.51)	-26.67 (6.78)	-0.50 (0.25)	-61.72 (40.10)	-27.11 (6.78)	-0.92 (0.25)
ϵ_2	110.47 (11.65)	84.15 (14.78)	-22.39 (7.05)	0.18 (0.83)	-47.77 (40.61)	-21.14 (7.07)	1.33 (0.84)
ϵ_3	399.57 (15.33)	372.67 (19.88)	-29.78 (6.77)	-1.92 (1.45)	-34.79 (42.18)	-25.00 (6.83)	3.31 (1.47)
ϵ_4	882.50 (22.98)	879.81 (30.30)	-28.32 (6.74)	-0.72 (2.18)	14.36 (44.32)	-17.21 (6.93)	10.63 (2.30)
				One Jump			
ϵ_1	269.61 (35.26)	100.30 (17.80)	233.23 (29.92)	258.49 (21.56)	184.42 (73.31)	-25.58 (6.86)	-2.19 (0.25)
ϵ_2	364.35 (34.40)	200.35 (19.23)	226.94 (28.28)	258.79 (21.54)	201.67 (71.72)	-21.96 (6.82)	4.05 (0.85)
ϵ_3	648.93 (38.20)	498.06 (24.94)	218.29 (27.72)	249.82 (20.78)	214.87 (74.11)	-5.09 (7.19)	23.16 (1.66)
ϵ_4	1143.10 (44.73)	1017.00 (35.55)	221.50 (27.14)	250.37 (21.52)	255.73 (71.84)	36.01 (8.13)	60.87 (3.15)
				Two Jumps			
ϵ_1	507.64 (54.73)	217.73 (23.69)	468.53 (48.86)	499.98 (37.75)	422.53 (106.44)	-27.23 (7.05)	-1.50 (0.25)
ϵ_2	618.08 (57.83)	323.80 (24.72)	475.53 (49.28)	505.72 (39.53)	446.02 (102.66)	-13.93 (6.99)	6.57 (0.88)
ϵ_3	902.48 (63.40)	620.44 (30.54)	470.85 (50.44)	502.56 (40.29)	462.64 (106.41)	15.21 (7.49)	43.52 (1.81)
ϵ_4	1399.10 (70.64)	1156.50 (43.00)	470.35 (49.76)	498.97 (40.92)	504.29 (109.07)	87.73 (9.16)	114.08 (3.94)
				Three Jumps			
ϵ_1	767.20 (77.56)	337.59 (28.64)	721.80 (68.54)	755.62 (56.93)	674.80 (130.51)	-25.42 (6.82)	-2.66 (0.25)
ϵ_2	866.12 (78.84)	443.31 (30.34)	720.71 (69.21)	754.90 (58.38)	689.69 (134.96)	-13.72 (6.83)	11.64 (0.91)
ϵ_3	1164.80 (83.67)	759.78 (36.66)	730.27 (69.86)	758.64 (59.37)	713.23 (135.13)	41.37 (7.85)	60.48 (1.93)
ϵ_4	1661.80 (93.00)	1303.50 (50.63)	724.24 (69.10)	752.55 (59.56)	762.91 (145.29)	142.24 (10.54)	163.84 (4.82)

Increments of the volatility process with $H \in (0.5, 1]$ exhibit the desired long memory. Thus we will study this model for a Hurst exponent equal to $H = \{0.5, 0.7, 0.9\}$. While the first case has independent increments, the second and third cases exhibit quite strong long memory processes in volatility.

Tables 4.2, 4.3 and 4.4 summarize the results for the different $H = \{0.5, 0.7, 0.9\}$, respectively. The results confirm exactly the same behavior for all the estimators as in the previous case without long memory. Thus we can conclude that our JWTSRV estimator is robust to jumps and noise on small samples even if we consider the volatility process with long memory, and it proved to be the best estimator of $\langle p, p \rangle_t$ even on small samples. While we studied only the in-sample performance of the estimator, we present the out-of-sample, or forecasting, performance in the next section.

4.3 One-day-ahead forecasts of IV using JWTSRV

One of the many potential useful applications of the proposed framework is volatility forecasting. In particular, the one-day-ahead return variation forecast, $var(p_{t+1}|\mathcal{F}_t)$, is of huge interest for practitioners. Thus we would like to study the forecasting ability of the proposed methodology in this section. While we showed that the in-sample performance of the estimators is the same for different models and that the JWTSRV estimator tends to consistently estimate $\langle p, p \rangle_t$ regardless of the level of noise and number of jumps in the process, we will reduce our simulation scheme to model 4.1 with a fixed level of noise and number of jumps. This setting will allow us to study the impact of noise and jumps on

Tables 11.2, 11.3 and 11.4 in Appendix 11B.

the forecasting performance of the estimators and to see if the JWTSRV holds its power and is able to forecast $var(p_{t+1}|\mathcal{F}_t)$.

Denoting the annualized one day time interval $T_1 - T_0 = T_2 - T_1$,

$$E\left[\sigma^2_{T_1}|\mathcal{F}_{T_0}\right] = e^{-\kappa(T_1-T_0)}\sigma^2_{T_0} + \alpha(1 - e^{-\kappa(T_1-T_0)}), \tag{4.3}$$

where σ^2_t follows model 4.1 and $\mathcal{F}_T = \sigma\{\sigma^2_t; t \leq T\}$ is the information set generated by the instantaneous variance process up to time T. If we use integration operators, we have

$$E\left[\int_{T_0}^{T_1} \sigma^2_t dt|\mathcal{F}_{T_0}\right] = \frac{1}{\kappa}(1 - e^{-\kappa(T_1-T_0)})\sigma^2_{T_0} + \alpha(T_1 - T_0) - \frac{\alpha}{\kappa}(1 - e^{-\kappa(T_1-T_0)}). \tag{4.4}$$

If we want to express the one-day-ahead forecast, we simply use equations 4.3 and 4.4 and we get:

$$E\left[\int_{T_m}^{T_{m+1}} \sigma^2_t dt|\mathcal{F}_{T_{m-1}}\right] = e^{-\kappa D}E\left[\int_{T_{m-1}}^{T_m} \sigma^2_t dt|\mathcal{F}_{T_{m-1}}\right] + \alpha(1 - e^{-\kappa D})D, \tag{4.5}$$

where $D = T_{m+1} - T_m = T_m - T_{m-1}$. Equation 4.5 is the exact conditional forecast of $\int_{T_m}^{T_{m+1}} \sigma^2_t dt$, but it is not feasible, as $E\left[\int_{T_{m-1}}^{T_m} \sigma^2_t dt|\mathcal{F}_{T_{m-1}}\right]$ is not observed in practice. But if we replace this term by the estimate of the integrated variance on day m we arrive at a simple method for forecasting the integrated variance on day $m + 1$. In empirical applications the true underlying model parameters are unknown and the properties of the observed data differ from the simulated ones, even though the simulations are based on estimated parameters on real-world data. Hence, the estimation is required to be realistic, and the AR(1) process seems to serve well in this case.

We use the simulation scheme for model 4.1 from the previous section. This time, we simulate 101 "continuous" sample paths over days $[0, T_1], \ldots, [T_{99}, T_{100}]$, $[T_{100}, T_{101}]$, that is, $101 \times 23,400$ log returns. We split each simulated path into two parts. The first part, of $100 \times 23,400$, is used to estimate the time series of 100 daily integrated variations using the tested estimators. Then, the AR(1) model is used to estimate the coefficients of forecast equation 4.5, where the conditional expectation in the right-hand side is replaced by the estimated integrated variation. The second part, the last (101th) day, is saved for out-of-sample comparison purposes as the true integrated variance of the day, which is compared with the AR(1) forecast of the integrated variance for the $m + 1$th day. This procedure is repeated for each simulated sample path of $101 \times 23,400$ log returns and all the estimators tested in the previous exercise.

We employ the traditional Mincer and Zarnowitz (1969) approach to assess the forecasting performance of the individual estimators. We compare alternative variance forecasts by projecting the true realized integrated variance on day $m + 1$, $\int_{T_m}^{T_{m+1}} \sigma^2_t dt$, on a constant and various estimator forecasts. For example, we evaluate the JWTSRV forecasting performance by running the following regression:

$$\langle p, p\rangle_{T_m+1} = \alpha + \beta V^{JWTSRV}_{T_{m+1|T_m}} + \epsilon, \tag{4.6}$$

Table 4.5: Out-of-sample Mincer-Zarnowitz regressions (Eq. 4.7) on model with no jumps. Results significant at 95% are in bold; OLS standard errors in parenthesis.

Joint Mincer-Zarnowitz regression

	const.	RV	BV	TSRV	RK	JWTSRV	R^2
	-0.055 (0.003)	**1.568** (0.099)	**-0.440** (0.103)				0.895
	0.009 (0.003)	0.072 (0.093)	**-0.172 (0.078)**	**1.070** (0.039)			0.941
	0.009 (0.003)	0.082 (0.090)	-0.108 (0.077)	**1.184** (0.0407)	**-0.199** (0.027)		0.944
	0.009 (0.003)	0.082 (0.090)	-0.109 (0.077)	**1.054** (0.293)	**-0.199** (0.027)	0.131 (0.294)	0.944

Individual Mincer-Zarnowitz regression

	const.	RV	BV	TSRV	RK	JWTSRV	R^2
RV	**-0.059** (0.003)	**1.145** (0.013)					0.893
BV	**-0.063** (0.003)		**1.167** (0.014)				0.869
TSRV	-0.002 (0.002)			**0.995** (0.008)			0.940
RK	-0.002 (0.003)				**1.017** (0.016)		0.807
JWTSRV	0.001 (0.002)					**0.997** (0.008)	0.939

Mincer-Zarnowitz regression for minimum variance TSRV estimators

	const.	TSRV*	JWTSRV*	R^2
TSRV*	-.002 (0.001)	**0.993** (0.006)		0.959
JWTSRV*	0.001 (0.001)		**0.996** (0.006)	0.959

where $V^{JWTSRV}_{T_{m+1|T_m}}$ is the one-day-ahead forecast of integrated variance from day m to day $m+1$ using the AR(1) prediction. Thus, equation 4.6 regresses the true realized variance $\langle p,p\rangle_{T_{m+1}}$ from day $m+1$ on a constant and the variance forecast using the JWTSRV estimator. If the JWTSRV estimator performs well, the forecast should be unbiased and the forecast error is small. In other words, $\alpha = 0$ and $\beta = 1$, and the R^2 of the regression is close to 1. Thus we will test the null hypothesis of $H_0 : \alpha = 0$ and $H_0 : \beta = 1$ against the alternatives $H_A : \alpha \neq 0$ and $H_A : \beta \neq 1$.

In our simulations, we study a Mincer-Zarnowitz style regression combining several estimators:

$$\begin{aligned}\left\{\langle p,p\rangle_{T_{m+1}}\right\}_j &= \alpha + \beta_1\left\{V^{RV}_{T_{m+1|T_m}}\right\}_j + \beta_2\left\{V^{BV}_{T_{m+1|T_m}}\right\}_j + \beta_3\left\{V^{TSRV}_{T_{m+1|T_m}}\right\}_j \\ &\quad +\beta_4\left\{V^{RK}_{T_{m+1|T_m}}\right\}_j + \beta_5\left\{V^{JWTSRV}_{T_{m+1|T_m}}\right\}_j + \epsilon_j \end{aligned} \tag{4.7}$$

for $j = 1,\ldots,10,000$ simulated sample paths. $V^{\mathcal{M}}_{T_{m+1|T_m}}$ is the one-day-ahead forecast of integrated variance from day m to day $m+1$ given by the AR(1) model for the time series of daily variance estimated by the $\mathcal{M}$ estimator of realized variance. Regression 4.7 can be naturally interpreted as a variance forecast encompassing regression, as a coefficient significantly different from zero implies that the information in that particular forecast is not included in the forecasts of other models. To test the robustness of the results, we also run individual regressions where we consider only a constant and a single forecasting model. Thus we run four separate regressions to supplement the joint regression from 4.7.

4.3.1 Forecasting without jumps

We run the simulations for two model settings using model 4.1 with one jump and with no jumps. Let us start with the model without jumps first. The OLS estimates of all the

Table 4.6: Out-of-sample Mincer-Zarnowitz regressions (Eq. 4.7) on model with 1 jump. Results significant at 95% are in bold; OLS standard errors in parenthesis.

Joint Mincer-Zarnowitz regression

	const.	RV	BV	TSRV	RK	JWTSRV	R^2
	-0.032 (0.006)	**-0.500** (0.045)	**1.538** (0.037)				0.811
	0.032 (0.010)	**-1.857** (0.185)	**1.512** (0.036)	**1.251** (0.166)			0.822
	0.032 (0.010)	**-1.873** (0.186)	**1.514** (0.036)	**1.182** (0.183)	0.088 (0.098)		0.822
	0.000 (0.006)	0.129 (0.122)	-0.045 (0.043)	-0.042 (0.113)	**-0.204** (0.059)	**1.078** (0.026)	0.936

Individual Mincer-Zarnowitz regression

	const.	RV	BV	TSRV	RK	JWTSRV	R^2
RV	**-0.100** (0.009)	**1.123** (0.037)					0.480
BV	**-0.079** (0.005)		**1.181** (0.019)				0.788
TSRV	**-0.050** (0.007)			**1.028** (0.032)			0.500
RK	**-0.051** (0.008)				**1.041** (0.035)		0.476
JWTSRV	-0.003 (0.002)					**1.024** (0.009)	0.935

Mincer-Zarnowitz regression for minimum variance TSRV estimators

	const.	TSRV*	JWTSRV*	R^2
TSRV*	**-0.049** (0.007)	**1.021** (0.032)		0.506
JWTSRV*	-0.003 (0.002)		**1.014** (0.007)	0.957

forecast evaluation regressions for the model without jumps are reported in Table 4.5. The results suggests that the TSRV performs as the best forecasting vehicle. Comparing the individual regressions, the TSRV has the highest R^2 and the coefficient closest to 1 with an insignificant coefficient, which suggests that the forecasts of the TSRV are biased only very slightly (as the coefficient is significantly different from 1). When looking at the joint regressions, we can see that the addition of all the other estimators does not improve this result. Moreover, when the TSRV is included in the regression, it is the only significant estimator, meaning that none of the other estimators has additional information not included in the TSRV forecast. In other words, adding the other estimators' forecasts to the TSRV brings no additional explanatory power to the regression. The JWTSRV forecast has the same performance as the simple TSRV, as there are no jumps in the simulated process, thus the asymptotic behavior of these two estimators should be the same. The JWTSRV is expected to have much better performance in the simulations where we include jumps. All the estimators are estimated with a 5-minute sampling frequency.

In addition, we provide results for the optimal sampling minimizing variance of the estimator in the last part of the table. The TSRV* with optimally chosen sampling outperforms the 5 min. TSRV. The JWTSRV* again has the same performance as expected.

4.3.2 Forecasting with jumps

Let's see how the results change when we add a single jump to the simulated model. The OLS estimates of all the forecast evaluation regressions for the model with jumps are reported in Table 4.6. Looking at the results of the individual regressions, one can see that the JWTSRV largely outperforms all the other estimators, with R^2 close to the results from the model without jumps from the previous section. This suggests that the

JWTSRV is robust to jumps even when we consider forecasting. The joint regression confirms this result. The regression including all the forecasts using the four considered estimators has the largest explanatory power. Moreover, the coefficient of the JWTSRV is significant, while the other coefficients are not significant, suggesting that the other estimators carry no additional information. Taking the JWTSRV forecasts away from the regression results in much lower R^2. It is interesting to note that in this case all the other coefficients are significant, suggesting multicollinearity caused by jumps in the process. The reader can also note how the addition of the BV improves the result. In fact, the BV rules the TSRV, with much higher R^2. In fact, the BV is used for jump detection, so this finding confirms the results from the literature.

In addition, we again include results for optimal sampling, which minimizes the variance of the TSRV-based estimators. In this case again, we can see that the result improves and the JWTSRV* yields the best result.

To conclude this section, the results suggest that when the JWTSRV estimator is used for variance forecasting in the presence of jumps and noise, the forecasts will be unbiased even on small samples. This makes the JWTSRV estimator a very powerful tool for forecasting the variance of stock market returns. With the theoretical results in hand, we can move to empirical examples and use the JWTSRV to forecast the volatility of real-world data.

Chapter 5

Decomposition of empirical volatility

Without (real) data, econometrics is merely *"l'art pour l'art"*[1]. In this section, therefore, we turn our focus to real-world data estimation of the proposed theory. We will test several integrated volatility estimators in comparison to our JWTSRV estimator and study their distributional properties. The JWTSRV proved to have lowest bias in the Monte Carlo simulations, thus we also expect it to have the best performance on the real data set. We will also use the wavelet-based estimator to decompose realized volatility into several investment horizons, which will allow us to construct a new forecasting model.

5.1 Data description

Foreign exchange future contracts are traded on the Chicago Mercantile Exchange (CME) on a 24-hour basis. As these markets are among the most liquid, they are suitable for analysis of high-frequency data. We will estimate the realized volatility of British pound (GBP), Swiss franc (CHF) and euro (EUR) futures. All contracts are quoted in the unit value of the foreign currency in US dollars. It is advantageous to use currency futures data for the analysis instead of spot currency prices, as they embed interest rate differentials and do not suffer from additional microstructure noise coming from over-the-counter trading. The cleaned data are available from Tick Data, Inc.[2]

It is very important to look first at the changes in the trading system before we proceed with the estimation on the data. In August 2003, for example, the CME launched the Globex trading platform, and for the first time ever in a single month, the trading volume on the electronic trading platform exceeded 1 million contracts every day. On Monday, December 18, 2006, the CME Globex(R) electronic trading platform started

[1] A French slogan that translates as *"art for art's sake"*.

[2] http://www.tickdata.com/

Table 5.1: The table summarizes the daily log-return distributions of GBP, CHF and EUR futures. The sample period extends from January 5, 2007 through November 17, 2010, accounting for a total of 944 observations.

	Mean	St.dev.	Skew.	Kurt.
GBP	0.0001	0.0119	-0.3852	4.4356
CHF	0.0002	0.0068	0.2440	5.4662
EUR	0.0002	0.0099	0.1536	4.4951

offering nearly continuous trading. More precisely, the trading cycle became 23 hours a day (from 5:00 pm on the previous day until 4:00 pm on current day, with a one-hour break in continuous trading), from 5:00 pm on Sunday until 4:00 pm on Friday. These changes certainly had a dramatic impact on trading activity and the amount of information available, resulting in difficulties in comparing the estimators on the pre-2003 data, the 2003–2006 data and the post–2006 data. For this reason, we restrict our analysis to a sample period extending from January 5, 2007 through November 17, 2010, which contains the most recent financial crisis. The futures contracts we use are automatically rolled over to provide continuous price records, so we do not have to deal with different maturities.

The tick-by-tick transactions are recorded in Chicago Time, referred to as Central Standard Time (CST). Therefore, in a given day, trading activity starts at 5:00 pm CST in Asia, continues in Europe followed by North America, and finally closes at 4:00 pm in Australia. To exclude potential jumps due to the one-hour gap in trading, we redefine the day in accordance with the electronic trading system. Moreover, we eliminate transactions executed on Saturdays and Sundays, US federal holidays, December 24 to 26, and December 31 to January 2, because of the low activity on these days, which could lead to estimation bias. Finally, we are left with 944 days in the sample. Looking more deeply at higher frequencies, we find a large amount of multiple transactions happening exactly at the same time stamp. We use the arithmetic average for all observations with the same time stamp.

5.2 Statistical properties of unconditional return and integrated volatility

Having prepared the data, we can estimate the integrated volatilities and study their statistical properties as well as the properties of the daily unconditional returns. For each futures contract, the daily integrated volatility is estimated using the square root of realized variance estimator defined by 2.12, the bipower variation estimator defined by 2.27, the two-scale realized volatility defined by 2.21, the realized kernel defined by 2.24 and the jump wavelet two-scale realized variance defined by 3.44. All the estimators are adjusted for small sample bias. For convenience, we refer to the estimators in the description of the results as RV, BV, TSRV, RK and JWTSRV, respectively. The RV and BV estimates are estimated on 5-min log-returns. The TSRV and the JWTSRV are

Table 5.2: The table summarizes the daily standardized daily log-return distributions for GBP, CHF and EUR futures using $r_t/IV_t^{1/2}$ and daily distributions of integrated volatility $IV_t^{1/2}$. Integrated volatility $IV_t^{1/2}$ is estimated using the RV, the BV on 5-min. log-returns, and the TSRV and JWTSRV on 5 minutes for a slow time scale and the RK. The sample period extends from January 5, 2007 through November 17, 2010, accounting for a total of 944 observations.

	Distributions of $r_t/IV_t^{1/2}$					**Distributions of $IV_t^{1/2}$**			
	GBP futures					**GBP futures**			
	Mean	St.dev.	Skew.	Kurt.		Mean	St.dev.	Skew.	Kurt.
RV	0.0419	0.8834	-0.0880	2.6029	RV	0.0075	0.0038	1.8394	7.5736
BV	0.0448	0.9266	-0.0669	2.6941	BV	0.0073	0.0037	1.7336	6.7996
TSRV	0.0451	0.9026	-0.0710	2.5744	TSRV	0.0073	0.0037	1.7611	7.0767
RK	0.0458	0.9406	-0.0757	2.5162	RK	0.0070	0.0037	1.8201	7.6473
JWTSRV	0.0489	0.9035	-0.0710	2.7512	JWTSRV	0.0071	0.0037	1.7629	7.0112
	CHF futures					**CHF futures**			
RV	0.0238	0.8959	0.0380	2.6272	RV	0.0076	0.0029	1.6875	8.2794
BV	0.0272	0.9424	0.0727	2.7020	BV	0.0073	0.0028	1.5696	7.5983
TSRV	0.0278	0.9180	0.0568	2.6161	TSRV	0.0073	0.0028	1.5572	7.3379
RK	0.0281	0.9530	0.0425	2.5371	RK	0.0070	0.0028	1.8179	9.9149
JWTSRV	0.0389	0.9253	0.0611	2.7170	JWTSRV	0.0070	0.0026	1.4359	6.5452
	EUR futures					**EUR futures**			
RV	0.0379	0.9550	-0.0215	2.5728	RV	0.0068	0.0031	1.4785	5.8493
BV	0.0410	0.9970	-0.0271	2.6219	BV	0.0066	0.0031	1.5001	5.9803
TSRV	0.0397	0.9638	-0.0133	2.5502	TSRV	0.0068	0.0031	1.4263	5.4871
RK	0.0415	0.9898	-0.0069	2.4497	RK	0.0065	0.0031	1.5351	6.2713
JWTSRV	0.0452	0.9587	0.0014	2.8144	JWTSRV	0.0064	0.0030	1.4345	5.4716

estimated using a slow time scale of 5 minutes.

Table 5.1 presents the summary statistics for the daily log-returns of GBP, CHF and EUR futures over the sample period, $t = 1, \ldots, 944$, i.e., January 5, 2007 to November 17, 2010. The summary statistics display an average return very close to zero, skewness, and excess kurtosis which is consistent with the large empirical literature started probably by Fama (1965) and Mandelbrot (1963). As observed by Andersen et al. (2001), when the log-returns are standardized by the integrated volatility, $r_t/IV_t^{1/2}$, the unconditional returns are very close to a Gaussian distribution.

Table 5.2 summarizes the unconditional distribution of the daily log-returns standardized by the integrated volatility, $r_t/IV_t^{1/2}$, and confirms this result. However, quite significant differences can be found among the estimators. While the high kurtosis (above 4) for the raw returns is reduced to the range of 2.51–2.81 for the log-returns standardized using the integrated volatility estimator, there is a notable difference between the estimators. The RV is expected to perform the worst, as it should be biased by microstructure noise and jumps, which is confirmed. The TSRV as well as the RK are not biased by noise, but it still contains a jump component of integrated variance. The BV should consistently estimate the jump components; the statistical distribution of

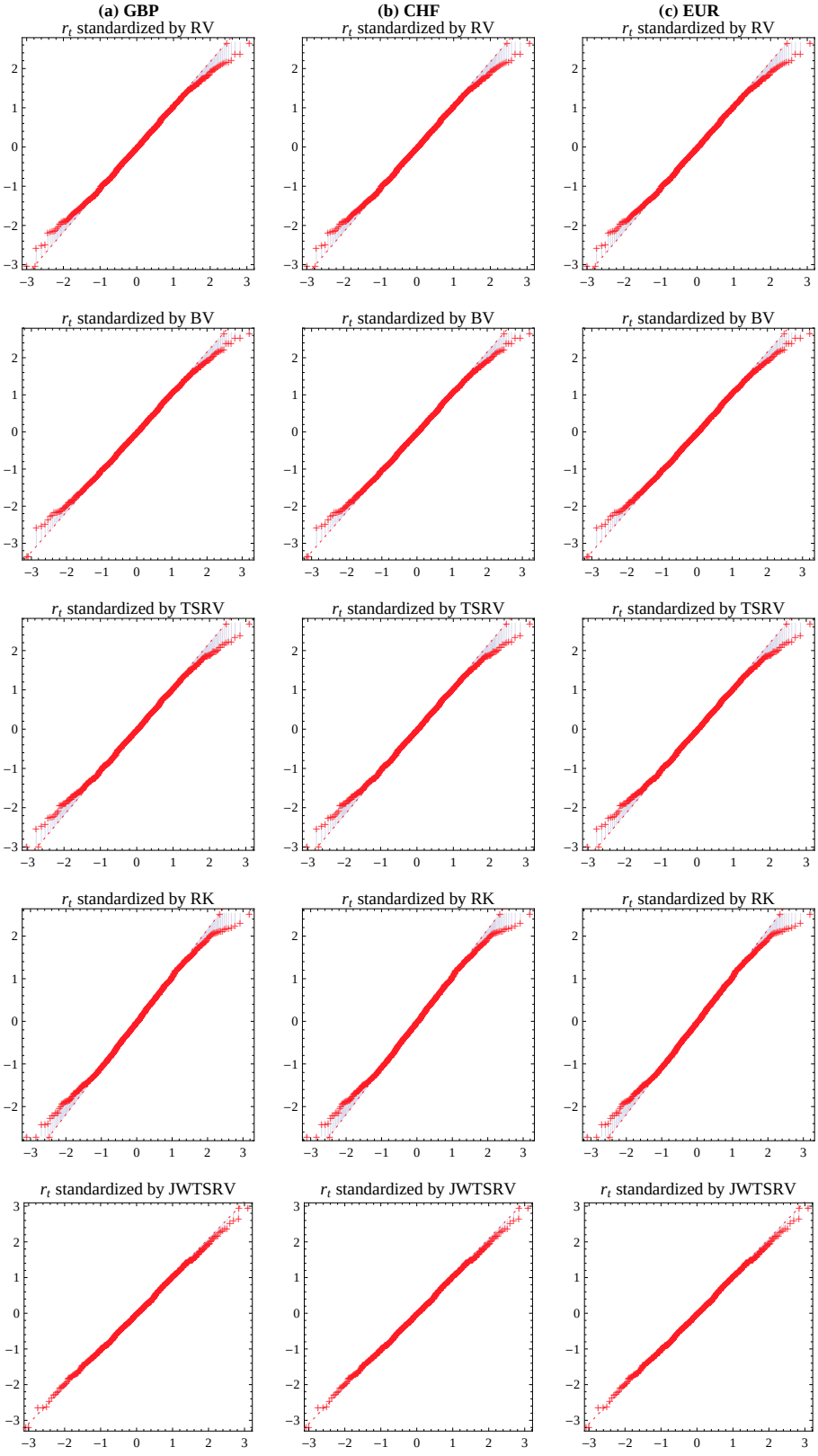


Figure 5.1: QQ plots of normalized daily log-returns r_t by RV, BV, TSRV, RK and JWTSRV estimators. (a) GBP futures, (b) CHF futures and (c) EUR futures

$r_t/IV_t^{1/2}$, where IV_t is estimated by the BV, should be closer to Gaussian. Finally, we expect JWTSRV estimator to perform the best, as it proved to be robust to noise and jumps in the Monte Carlo simulations. We also borrow the QQ plots plotted in Figure 5.1 for help. Similarly as Fleming and Paye (2011) and Andersen et al. (2011), we ask whether the jumps account for the non-normality of the unconditional log-returns standardized by the integrated volatility estimators found in the literature. We add the TSRV, RK and JWTSRV estimators for comparison. Figure 5.1 shows that returns standardized by integrated volatility using the JWTSRV provide the best approximation of the standard normal distribution. This result is in line with what we expected, as the JWTSRV proved to be robust to noise and jumps in our large Monte Carlo study. The result from the BV leaves us puzzled. While it is expected to be robust to jumps, it should be able to perform better. The returns standardized by the BV have higher kurtosis than those standardized by the RV, TSRV or RK, thus the BV outperforms these estimators to some extent. However, the JWTSRV confirms the theory presented in the previous sections. Figure 11.1, which also shows histograms of the standardized series, can be found in Appendix 11B

Moving from the distributional properties of the standardized daily log-returns, Table 5.2 also shows the distributional properties of the $IV_t^{1/2}$ estimators. Again, the JWTSRV provides lower estimates of $IV_t^{1/2}$ and is also less volatile than the RV. This finding is consistent with the fact that the RV can be affected by microstructure noise, and, as demonstrated in the Monte Carlo simulations, the JWTSRV is able to estimate the true integrated variance with the lowest bias in the presence of noise and jumps in the data. It is surprising, though, that the average estimate of $IV_t^{1/2}$ using the JWTSRV is 6.34% lower than the average estimate from the RV (computed as arithmetic averages on the estimators on GBP futures, CHF futures and EUR futures) with kurtosis 12.32% lower than the RV. The average estimate of $IV_t^{1/2}$ using the JWTSRV is 3.76% lower than the average estimate using GBP, with kurtosis 6.34% lower. Finally, the average estimate of $IV_t^{1/2}$ using the JWTSRV is 4.52% lower than the average estimate using the TSRV, with kurtosis 4.39% lower. Finally, the average estimate of $IV_t^{1/2}$ using the JWTSRV is the same as the average estimate using the RK with kurtosis 25.39% lower. It is thus interesting that while the TSRV accounts for noise but not jumps and the BV accounts for jumps but is not able to deal with noise, they have same deviations from the JWTSRV, which seems to estimate the integrated volatility without jumps and noise. Most interesting is that the average estimate of the RK is exactly the same as the average estimate of the JWTSRV. However, the RK estimates has much higher kurtosis. This result shows that the RK is powerful estimator of the realized variance. Finally let us note that these differences are economically significant, as they result in different asset pricing.

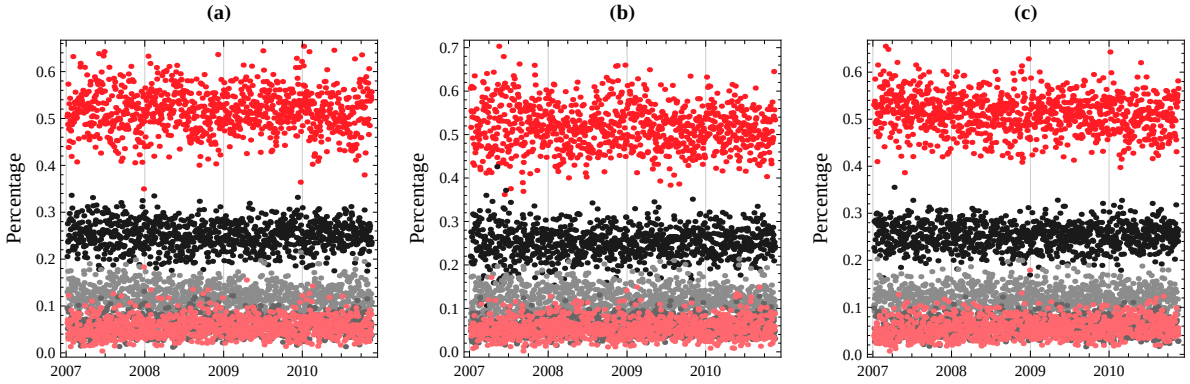


Figure 5.2: $JWTRSV_j$, $j = 1, \ldots, 5$, contributions of components of integrated volatility IV_t corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day. (a) GBP futures, (b) CHF futures and (c) EUR futures.

5.3 IV_t decomposition using wavelets

From the previous analysis, we could see that the JWTSRV provides the best estimator not only theoretically, but also on empirical data sets. Although this is the most important property of the JWTSRV, it is not the only one we can take advantage of. Another advantage is that by using the JWTSRV, we are able to decompose the integrated variance into several investment horizons, or components. In our analysis, we limit ourselves to decomposition into four scales corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes and 40–80 minutes, and the rest (80 minutes up to 1 day). As shown in the theoretical part of this work, we can comfortably decompose the integrated variance into these components, as their sum will always give the integrated variance estimator.

More precisely, the components of the JWTSRV from 3.44 correspond to various investment horizons. Thus, we will refer to these as $JWTSRV_j$:

$$JWTSRV = \sum_{j=1}^{5} JWTSRV_j, \qquad j = 1, \ldots, 5 \tag{5.1}$$

where $j = 1, \ldots, 4$ are scales corresponding to 5–10 minutes, 10–20 minutes, 20–40 minutes and 40–80 minutes, and $j = 5$ will contain the 80 minutes up to 1 day investment horizon. Similarly to the previous section, we look at the distributional properties of the integrated variance decomposed into investment horizons. Table 11.5 in Appendix 11B provides the descriptive statistics for the $JWTSRV_j$ estimates for all the currency futures studied.

Much more interesting are the autocorrelation functions of integrated variance on the different investment horizons shown by Figure 11.3 in Appendix 11B, showing similar long memory and temporal dependence in all investment horizons. These types of behavior are well documented empirical features of financial asset returns. Table 11.5 in Appendix

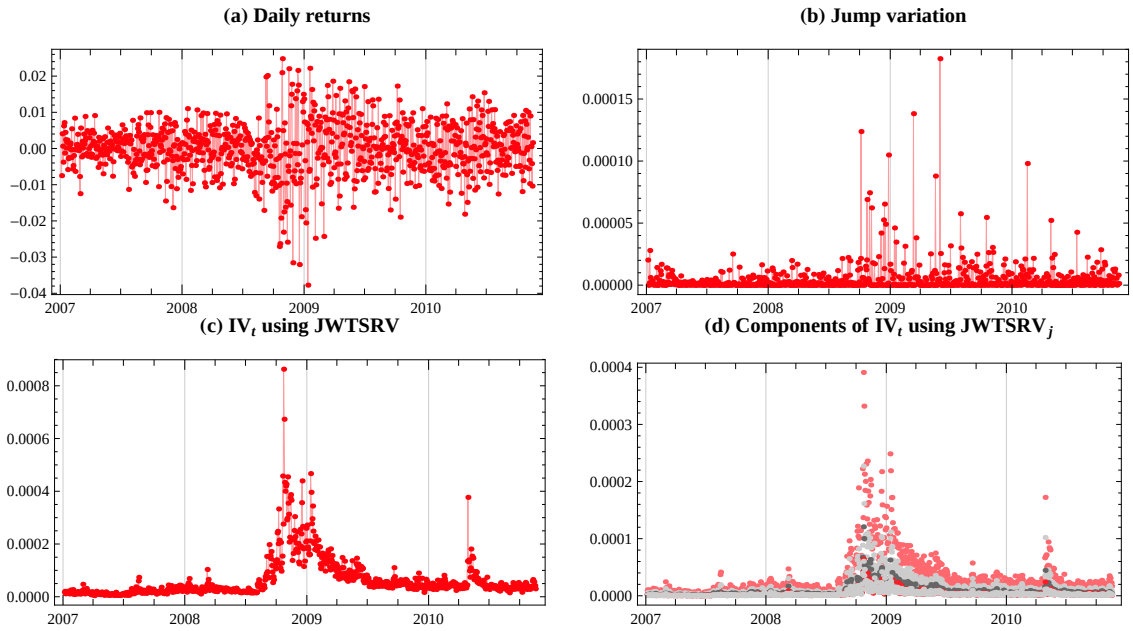


Figure 5.3: GBP futures: (a) daily returns, (b) JWTSRV estimated jump variation, (c) IV_t estimated by JWTSRV (d) decomposition of IV_t using $JWTSRV_j$ for $j = 1, \ldots, 5$ corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day.

11B also reports the Ljung-Box portmanteau test, which strongly rejects the joint null hypothesis of zero autocorrelations up to lag 20, i.e., about one month of trading. All the values are far from the critical values at the 1% confidence level, documenting the well-known volatility clustering feature of the data even on the decomposed integrated variances.

The final decomposition of the integrated variance into jumps and several investment horizons can be best seen from Figures 5.3, 5.4 and 5.5, which provide the returns, estimated jumps and finally integrated volatilities on investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day. We provide $JWTSRV_j$, which is the decomposed IV_t. It is interesting to observe that most of the information of the integrated variance is carried by the fastest scale, the 5–10 minute investment horizon. This is about 50% of the total variation. The longer the horizon, the lower the contribution of the variance to the total. For this purpose, we compute the weighted contributions of various investment horizon volatilities to the total. More precisely, we compute the contributions as:

$$\frac{JWTSRV_j}{\sum_{j=1}^{5} JWTSRV_j}, \tag{5.2}$$

for each $j = 1, \ldots, 5$. The results are shown in Figure 5.2 for all scales.[3] Ratio 5.2 is intuitive. If it equals zero, the investment horizon j has zero contribution to the overall

[3] A more detailed view is provided by Figure 11.2 in Appendix 11B.

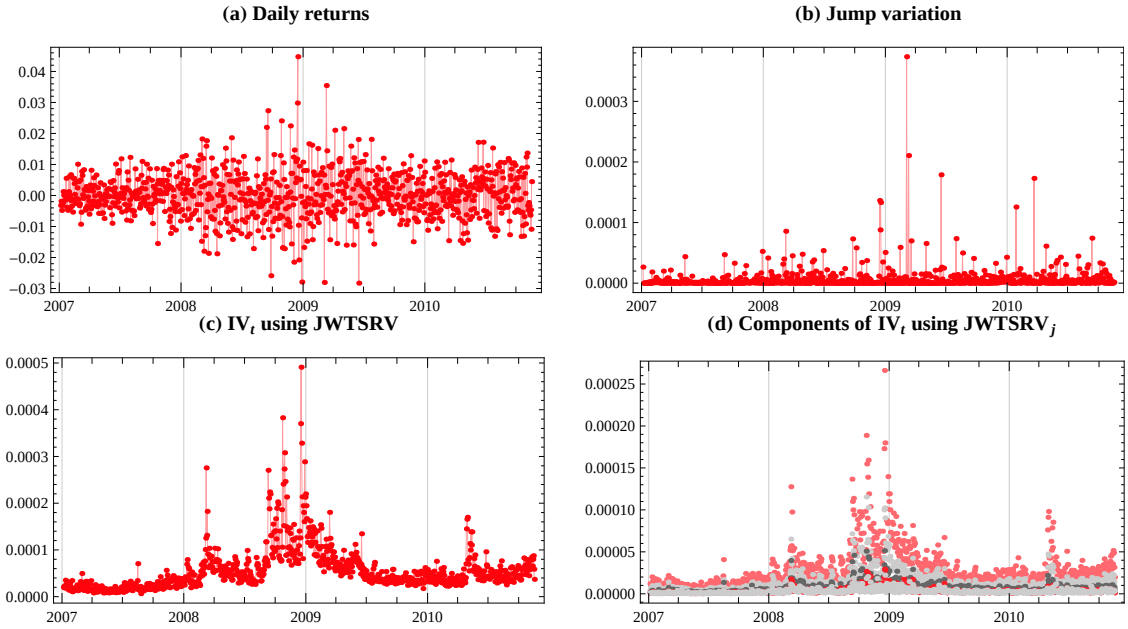


Figure 5.4: CHF futures: (a) daily returns, (b) JWTSRV estimated jump variation, (c) IV_t estimated by JWTSRV (d) decomposition of IV_t using $JWTSRV_j$ for $j = 1, \ldots, 5$ corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day.

variance. If it equals one, the corresponding investment horizon j explains all of the total variance. From Figure 5.2 we can see that the ratios are the same through all the currencies tested. They change quite considerably over the sample period. While the contribution of the first investment horizon, $j = 1$, corresponding to 5–10 minutes, to the total IV_t is around 51.5%, it is also the one with the largest dispersion. Over time, it changes from 40% to 60%. The second investment horizon (10-20 minutes), corresponding to $j = 2$, accounts for approximately 25% of the variance, followed by the third and fourth horizons (20–40 and 40–80 minutes, corresponding to $j = 3$ and $j = 4$), which account for only 12% and 6% approximately. The remaining 5%–6% are in the last $j = 5$.

For better illustration, we annualize the square root of the integrated variance in order to get the annualized volatility and we compute the components of the volatility on our investment horizons. Figure 5.6 shows this decomposition. The first plot, 5.6 (a), shows total volatility estimate, while 5.6 (b) to 5.6 (f) show the investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day, respectively. It is very interesting that most of the volatility (around 50%) comes from the fast, 10-minute investment horizon. This is a new insight we bring to volatility modeling. In fact, it is a logical finding, as it shows that volatility is created on fast scales of up to 10 minutes rather than on slower scales.

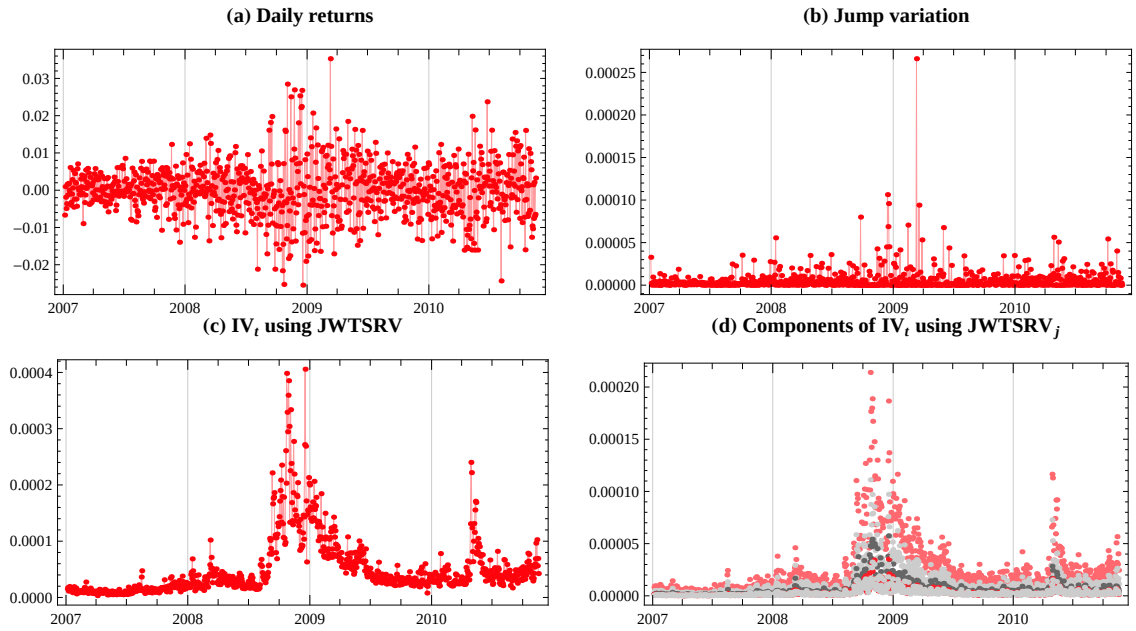


Figure 5.5: EUR futures: (a) daily returns, (b) JWTSRV estimated jump variation, (c) IV_t estimated by JWTSRV (d) decomposition of IV_t using $JWTSRV_j$ for $j = 1, \ldots, 5$ corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day.

5.4 Forecasting model based on decomposed integrated volatilities

Similarly to Lanne (2007) and Andersen et al. (2011), we use the decomposition of the quadratic variation with the intention of building a more accurate forecasting model. Our approach is very different though, as we use wavelets to decompose the integrated volatility into several investment horizons. More precisely, we build a long memory forecasting model to forecast each investment horizon separately and then put the forecasts together. We hope that this will result in better in-sample fits of the data as well as out-of-sample forecasts. This framework is motivated by the statistical properties of the decomposed volatility series found in the previous sections, which suggest that each scale might have somewhat different behavior. Thus, hopefully, there is some information to be extracted from it.

Empirical evidence on the strong temporal dependence of realized volatility has been well documented since Andersen et al. (2001). This evidence suggests that realized variance should be modeled by models allowing for a slowly decaying autocorrelation function and possibly long memory. For example, Müller et al. (1997), Arneodo et al. (1998) and Lynch and Zumbach. (2003) show that volatility over long time intervals has a strong influence on volatility at shorter time intervals, but volatility over short time intervals does not have an effect on longer intervals. A possible economic interpretation is that long term volatility matters for short-term traders, while short-term volatility does not

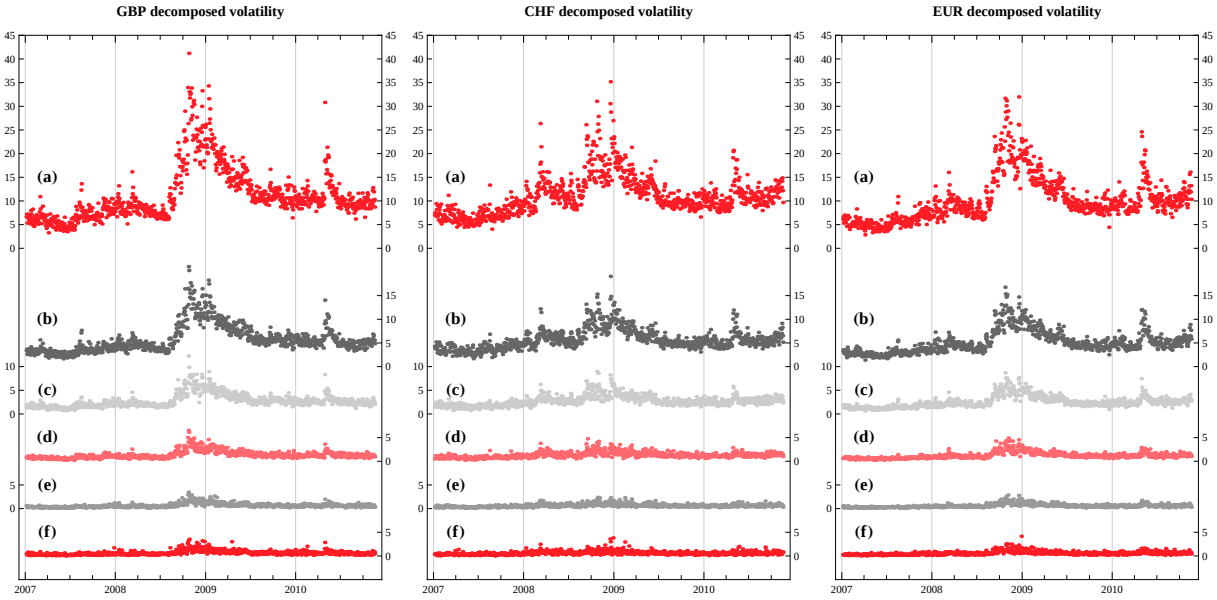


Figure 5.6: Decomposed annualized volatility (by 252 days). (a) total $IV_t^{1/2}$ estimate on GBP, CHF and EUR futures using JWTSRV, (b) volatility on investment horizon of 5–10 minutes, (c) volatility on investment horizon of 10–20 minutes, (d) volatility on investment horizon of 20—40 minutes, (e) volatility on investment horizon of 40–80 minutes, (f) volatility on investment horizon of 80 minutes up to 1 day. Note that sum of components (b), (c), (d), (e) and (f) give total volatility plotted in (a).

affect long-term trading strategies. While standard, ARCH-type volatility models (Engle, 1982), generalizations thereof (Bollerslev, 1986, 1987) and one-factor stochastic volatility models are not able to capture long memory and do not reproduce scaling and volatility cascades, fractionally integrated models Granger (1980) improve on these features, but still are not able to reproduce multiscaling and cascades of volatility. LeBaron (2001) shows that a combination of only three AR(1) processes can result in apparent long memory. These additive processes are called models with heterogeneous components, and they are able to generate all of these stylized facts. Further, the heterogeneous market hypothesis introduced by Müller et al. (1993) brings heterogeneity in different time horizons. Agents perceive, react and cause different volatility components.

While the volatility cascade model composed of hierarchical processes from low to high frequencies was needed to capture all of the desired properties of the empirical data, the realized volatility measures made the volatility observable. Thus Corsi (2009) proposed a cascade of a few heterogeneous realized volatility components called the Heterogeneous Autoregressive Model (HAR), which is able to reproduce the stylized facts of the data, including the apparent multiscaling. A problem with the HAR model is that it is only an approximate long-memory model, and as a result it might not be able to capture the dynamics of long memory properties in volatility well. Thus in our forecasting exercise, we follow Andersen et al. (2003) and adopt the autoregressive fractionally integrated moving average (ARFIMA) class of models.

If we assume that the volatility series belong to the class of ARFIMA processes introduced into econometrics by Granger and Joyeux (1980), then the dth difference of each series is a stationary and invertible ARMA process where parameter d can be any real number such that $-1/2 < d < 1/2$ to ensure stationarity and invertibility. More precisely, σ_t is an ARFIMA(p, d, q) process if it follows:

$$\alpha(L)(1-L)^d(\sigma_t - \mu) = \beta(L)v_t, \tag{5.3}$$

where $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ and $\beta(z) = 1 + \beta_1 z + \cdots + \beta_q z^q$ are polynomials of order p and q, respectively, in the lag operator $L(L\sigma_t = \sigma_{t-1})$, which roots strictly outside the unit circle, v_t is iid with zero mean and σ_v^2 variance, and $(1-L)^d$ is defined by its binomial expansion

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j \tag{5.4}$$

using gamma function, $\Gamma(.)$.

The parameter d determines the memory of the process. For $d > 0$, the process is said to have long memory, since its autocorrelations die out at a hyperbolic rate and are no longer absolutely summable, in contrast to the much faster exponential rate in the weak dependence case of $d = 0$, when the process captures the behavior of the short-memory ARMA model.

Once we have estimated the ARFIMA(p, d, q) model with the Haslett and Raftery (1989) maximum likelihood estimator, forecasting is carried out by extrapolating the estimated model. Deo et al. (2006), Andersen et al. (2003) and Martens and Zein (2004) show that forecasting log realized volatility based on a simple ARFIMA$(1, d, 0)$ specification is a very good competitor to other time-series methods of forecasting realized volatility. We estimate a simple ARFIMA$(1, d, 1)$. Log volatilities are used in the literature as they allow for a linear forecasting model. Thus, many researchers compare the realized volatility and its log-transformation in the forecasting exercise. In our work, we will follow this convention and use both realized volatility and its log-transformation.

5.4.1 Forecast evaluation

To analyze the forecast efficiency and information content of different volatility estimators, we employ the popular approach of Mincer and Zarnowitz (1969) regressions on both the realized volatility and its logarithmic transformation. The regression takes the form:

$$V_{t+1}^{(m)} = \alpha + \beta_1 V_t^{(k)ARFIMA} + \epsilon_t, \tag{5.5}$$

with $V_{t+1}^{(m)}$ being the integrated volatility (or its logarithmic transformation) estimated using the square root of the mth estimator, namely, realized variance defined by 2.12, the bipower variation estimator defined by 2.27, the two-scale realized volatility defined by 2.21, the realized kernel defined by 2.24 and the jump wavelet two-scale realized variance defined by 3.44. $V_t^{(k)ARFIMA}$ denotes the 1-day ahead forecast of $V_{t+1}^{(m)}$ using the kth

estimator based on ARFIMA$(1, d, 1)$, while we consider the same estimators. We report in-sample as well as rolling out-of-sample results.

After testing the forecasting efficiency of the different volatility estimators we would also like to test the information content of the wavelet decomposition of the realized volatility. For this purpose, we separately estimate ARFIMA$(1, d, 1)$ for all components $JWTSRV_j$ for $j = 1, \ldots, 5$ of the realized volatility as well as the estimated jumps. We should note that in the case of logarithmic transformation of the realized volatility, we also take logarithms of the decomposed levels $JWTSRV_j$. After obtaining the forecast for each level, we transform the forecasts back to be able to compare the results. For convenience, we refer to the estimators in the description of the results as RV, CBPV, TSRV, RK and JW, where CBPV and JW are referring to the continuous part of the realized variance estimated by the BV and JWTSRV estimator respectively, and finally $\sum JW$ referring to the sum of the individual forecasts of the decomposed realized volatilities.

Finally, we test the information content of the separate decomposed realized volatilities by estimating the following regressions:

$$JW_{t+1} = \alpha + \beta_1 W_{t,j}^{ARFIMA} + \epsilon_t, \tag{5.6}$$

where $W_{t,j}^{ARFIMA}$ denotes the forecasts of the individual components $JWTSRV_j$ for $j = 1, \ldots, 5$, corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day, respectively, and

$$JW_{t+1} = \alpha + \beta_1 J_t^{ARFIMA} + \epsilon_t, \tag{5.7}$$

where J_t^{ARFIMA} denotes the forecasts of the jumps. Thus we test the information content of the long memory forecasts of the realized volatility estimators using the coefficient of determination, R^2, of the regression.

5.4.2 Does decomposition bring any improvement in volatility forecasting?

We use the period from January 5, 2007 to December 31, 2009 to perform the estimations of all the models. We refer to this period as the in-sample period and it contains GBP futures, CHF futures and EUR futures. The year 2010 is saved for comparison of the out-of-sample forecasts, which are done on a rolling basis.

Tables 5.4 and 5.3 contain the results of the realized volatility and the logarithmic transform of the realized volatility, respectively. A striking result is that JW is the easiest to forecast in terms of having the highest R^2 for the out-of-sample forecasts for all the currencies. It suggests that the information content of our JW estimator is the best in comparison with the other estimators. The continuous part of the realized volatility thus seems to have the highest information content. This result is also proved by the finding that CBPV turns out to be the second easiest to forecast. The other estimators containing jump components are more difficult to forecast.

JW also seems to have the highest information content, as on average it forecasts all the other realized volatility estimates the best. An even more striking result is that the

Table 5.3: Results for $\log \widehat{RV}_t^{1/2}$*: R^2 for the Minzer-Zarnowitz regressions regressing ARFIMA forecasts of RV, CBPV, TSRV, RK, JW and $\sum JW$ on its estimates , W_j denotes $JWTSRV_j$, $j = 1, \ldots, 5$ components of realized volatility and Jump estimated jumps.*

GBP

	in-sample							out-of-sample						
	RV	CBPV	TSRV	RK	JW	∑JW	Avg	RV	CBPV	TSRV	RK	JW	∑JW	Avg
RV	0.871	0.873	0.872	0.862	0.877	0.877	0.872	0.320	0.330	0.303	0.313	0.372	0.378	0.336
CBPV	0.884	0.885	0.884	0.874	0.890	0.890	0.885	0.388	0.394	0.370	0.368	0.437	0.441	0.400
TSRV	0.871	0.873	0.873	0.862	0.878	0.877	0.872	0.295	0.305	0.283	0.299	0.346	0.346	0.313
RK	0.788	0.791	0.787	0.783	0.795	0.795	0.790	0.089	0.106	0.081	0.096	0.120	0.127	0.103
JW	0.891	0.893	0.892	0.882	0.899	0.899	**0.893**	0.415	0.431	0.392	0.396	0.458	0.462	**0.426**
Avg	0.861	0.863	0.862	0.853	**0.868**	0.867		0.302	0.313	0.286	0.294	0.347	**0.351**	
	W1	W2	W3	W4	W5	Jump		W1	W2	W3	W4	W5	Jump	
JW	0.896	0.891	0.886	0.879	0.878	0.039		0.456	0.418	0.403	0.401	0.406	0.075	

CHF

	in-sample							out-of-sample						
	RV	CBPV	TSRV	RK	JW	∑JW	Avg	RV	CBPV	TSRV	RK	JW	∑JW	Avg
RV	0.736	0.742	0.737	0.715	0.750	0.751	0.739	0.047	0.081	0.052	0.025	0.079	0.092	0.063
CBPV	0.760	0.766	0.761	0.735	0.774	0.776	0.762	0.083	0.121	0.088	0.037	0.123	0.140	0.098
TSRV	0.735	0.741	0.738	0.719	0.750	0.751	0.739	0.059	0.096	0.067	0.020	0.106	0.111	0.076
RK	0.591	0.593	0.593	0.593	0.594	0.599	0.594	0.001	0.005	0.003	0.001	0.009	0.011	0.005
JW	0.804	0.810	0.806	0.778	0.819	0.819	**0.806**	0.135	0.174	0.147	0.051	0.202	0.213	**0.154**
Avg	0.725	0.730	0.727	0.708	0.737	**0.739**		0.065	0.095	0.071	0.027	0.104	**0.113**	
	W1	W2	W3	W4	W5	Jump		W1	W2	W3	W4	W5	Jump	
JW	0.813	0.812	0.796	0.782	0.771	0.057		0.204	0.199	0.166	0.155	0.058	0.005	

EUR

	in-sample							out-of-sample						
	RV	CBPV	TSRV	RK	JW	∑JW	Avg	RV	CBPV	TSRV	RK	JW	∑JW	Avg
RV	0.836	0.843	0.838	0.831	0.846	0.846	0.840	0.327	0.340	0.315	0.317	0.329	0.332	0.327
CBPV	0.849	0.856	0.851	0.841	0.860	0.860	0.853	0.398	0.409	0.388	0.370	0.406	0.405	0.396
TSRV	0.834	0.841	0.837	0.831	0.845	0.845	0.839	0.311	0.316	0.300	0.298	0.315	0.314	0.309
RK	0.734	0.743	0.736	0.737	0.742	0.745	0.740	0.143	0.148	0.117	0.133	0.125	0.129	0.133
JW	0.882	0.888	0.885	0.872	0.892	0.892	**0.885**	0.426	0.426	0.423	0.390	0.427	0.421	**0.419**
Avg	0.827	0.834	0.829	0.822	0.837	**0.838**		0.321	**0.328**	0.309	0.301	0.320	0.320	
	W1	W2	W3	W4	W5	Jump		W1	W2	W3	W4	W5	Jump	
JW	0.889	0.887	0.882	0.875	0.873	0.021		0.425	0.391	0.382	0.398	0.293	0.002	

decomposed model, i.e., all five forecasts of the decomposed realized volatilities separately and when combined to provide the final forecast, is the best in most cases. This points to the result that decomposed volatility processes seem to contain information which helps us in forecasting. When looking at the information content of the separate investment horizons of the realized volatilities for the total volatility, the first level, i.e. 5–10 minutes, seems to carry the most significant forecasting information, while the impact slowly decays with longer horizon levels.

Finally, when looking at the results from the jumps we can see that they carry some information and could be useful in forecasting volatility. We should note that all the in-sample fits have estimated parameters significantly different from zero and they fit the

Table 5.4: Results for $\widehat{RV}_t^{1/2}$: R^2 for the Minzer-Zarnowitz regressions regressing ARFIMA forecasts of RV, CBPV, TSRV, RK, JW and $\sum JW$ on its estimates , W_j denotes $JWTSRV_j$, $j = 1, \ldots, 5$ components of realized volatility and Jump estimated jumps.

GBP

	in-sample							out-of-sample						
	RV	CBPV	TSRV	RK	JW	∑JW	Avg	RV	CBPV	TSRV	RK	JW	∑JW	Avg
RV	0.851	0.854	0.851	0.842	0.855	0.855	0.851	0.375	0.375	0.348	0.352	0.422	0.425	0.383
CBPV	0.871	0.874	0.872	0.861	0.876	0.875	0.872	0.452	0.450	0.422	0.412	0.495	0.499	0.455
TSRV	0.853	0.856	0.854	0.844	0.857	0.857	0.853	0.338	0.343	0.320	0.330	0.386	0.385	0.350
RK	0.785	0.789	0.785	0.776	0.788	0.788	0.785	0.126	0.136	0.112	0.118	0.157	0.163	0.136
JW	0.881	0.883	0.882	0.870	0.885	0.885	**0.881**	0.476	0.489	0.444	0.439	0.520	0.522	**0.482**
Avg	0.848	0.851	0.849	0.839	**0.852**	**0.852**		0.353	0.359	0.329	0.330	0.396	**0.399**	
	W1	W2	W3	W4	W5	Jump		W1	W2	W3	W4	W5	Jump	
JW	0.881	0.880	0.872	0.859	0.869	0.337		0.521	0.481	0.466	0.470	0.468	0.093	

CHF

	in-sample							out-of-sample						
	RV	CBPV	TSRV	RK	JW	∑JW	Avg	RV	CBPV	TSRV	RK	JW	∑JW	Avg
RV	0.675	0.685	0.677	0.643	0.692	0.689	0.677	0.043	0.083	0.048	0.004	0.086	0.094	0.060
CBPV	0.710	0.721	0.714	0.675	0.730	0.728	0.713	0.073	0.117	0.078	0.008	0.126	0.135	0.090
TSRV	0.674	0.684	0.679	0.649	0.693	0.690	0.678	0.054	0.094	0.062	0.004	0.111	0.111	0.073
RK	0.510	0.515	0.516	0.503	0.520	0.520	0.514	0.001	0.005	0.002	0.000	0.010	0.012	0.005
JW	0.770	0.782	0.773	0.733	0.790	0.787	**0.773**	0.128	0.177	0.141	0.020	0.215	0.213	**0.149**
Avg	0.668	0.678	0.672	0.641	**0.685**	0.683		0.060	0.095	0.066	0.007	0.110	**0.113**	
	W1	W2	W3	W4	W5	Jump		W1	W2	W3	W4	W5	Jump	
JW	0.782	0.782	0.758	0.742	0.738	0.134		0.210	0.209	0.162	0.160	0.040	0.032	

EUR

	in-sample							out-of-sample						
	RV	CBPV	TSRV	RK	JW	∑JW	Avg	RV	CBPV	TSRV	RK	JW	∑JW	Avg
RV	0.830	0.837	0.833	0.821	0.836	0.836	0.832	0.358	0.361	0.343	0.337	0.358	0.354	0.352
CBPV	0.844	0.850	0.847	0.833	0.851	0.851	0.846	0.425	0.430	0.414	0.389	0.430	0.424	0.418
TSRV	0.826	0.833	0.830	0.819	0.832	0.832	0.829	0.340	0.334	0.325	0.327	0.341	0.337	0.334
RK	0.732	0.737	0.734	0.724	0.734	0.734	0.732	0.181	0.176	0.150	0.170	0.155	0.160	0.165
JW	0.876	0.881	0.879	0.866	0.883	0.883	**0.878**	0.459	0.454	0.452	0.428	0.458	0.450	**0.450**
Avg	0.822	**0.827**	0.824	0.813	**0.827**	**0.827**		**0.353**	0.351	0.337	0.330	0.348	0.345	
	W1	W2	W3	W4	W5	Jump		W1	W2	W3	W4	W5	Jump	
JW	0.880	0.878	0.875	0.856	0.862	0.087		0.447	0.425	0.398	0.438	0.354	0.153	

data well. For reasons of space, we do not provide all the in-sample results.

To conclude, we have shown that decomposition of the realized volatility into several investment horizons and jumps brings an improvement in forecasting.

Chapter 6

Conclusion

In the first part of the thesis, we present the complete theoretical framework of wavelet-based realized variation generalizing the current realized variation theory to the time-frequency domain.

After the necessary introductions to the theory of quadratic variation and realized variation measurement, the first part of the dissertation defines the wavelet-based realized variation theory. Standing on our theoretical results proposing the Wavelet Representation Theorem, which extends the well-known Martingale Representation Theorem, the estimator of wavelet-based realized variation is defined together with its theoretical properties. Using wavelets, the estimator is able to consistently estimate jumps from the price process. It is robust to noise and it generates an unbiased consistent estimator of the *true* underlying variation. The theoretical part also contains an important discussion of the similarities between wavelet theory and stochastic processes.

To support the theory, a numerical study of the small sample performance of the estimators is carried out. In this study, we compare our estimators to several of the most popular estimators, namely, realized variance, bipower variation, two-scale realized volatility and realized kernels. The wavelet-based estimator proves to have lowest bias of all the estimators in the jump-diffusion model with stochastic volatility as well as the fractional stochastic volatility model simulated with different levels of noise and numbers of jumps. While all the other estimators suffer from substantial bias caused either by jumps or by noise, our theory proves to hold its properties under both noise and jumps. As predictability of volatility is of interest to researchers as well as practitioners, a numerical study of the behavior of the forecasts is also carried out. Again, our theory proves to be the most powerful in forecasting volatility under the different simulation settings.

While the first chapters of Part I derive the theory and show its power on the small sample study, the last chapter uses the theory to decompose the empirical volatility. By studying the statistical properties of unconditional daily log-return distributions standardized by volatility estimated using the different estimators we find that standardization by our wavelet-based estimator brings the returns close to the Gaussian normal distribution. All the other estimators are affected by the presence of jumps in the data.

The differences are economically significant, as we find that the average volatility estimated using our wavelet-based theory is 6.34% lower than the volatility estimated with the standard estimator.

Furthermore, we decompose the realized volatility into several intraday horizons. Here we note that the theory is able to decompose the realized measures into any arbitrary investment horizon, i.e., from 1-minute up to 1-month, when estimating monthly measures. In our analysis performed on forex data, we limit ourselves to illustrating the theory on the decomposition of daily realized measures. Specifically, we decompose the realized volatility into investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes and 40–80 minutes, and the rest (80 minutes up to 1 day). The analysis uncovers interesting dynamics. Most of the action in the stock markets comes from higher frequencies. We find that, on average, about 52% of the volatility of the forex markets examined is created on the 5–10 minute investment horizon, approximately 25% comes from the 10–20 minute investment horizon, and only 12%, 6% and 5% correspond to the horizons of 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day), respectively. Note that by adding the contributions of the different investment horizons we always get 100%.

The last part of the univariate empirical analysis is devoted to the forecasting of realized volatility. One of the issues with the interpretation of wavelets in economic applications is that they behave like a filter. Thus wavelets can hardly be used for forecasting of economic time series most of the time. But in the realized measures, we only use wavelets to decompose the daily volatility of the returns using intraday information, while forecasting daily volatility. We build a new forecasting model based on a long memory ARFIMA which uses the decomposition provided by our theory. In-sample as well as more important out-of-sample forecasts show that our theory is able to forecast volatility with the lowest error.

Concluding the empirical findings, we show that our wavelet-based theory brings a significant improvement to volatility estimation and forecasting. It also offers a new method of time-frequency modeling of realized volatility which helps us to better understand the dynamics of stock market behavior. Specifically, our theory uncovers that most of the volatility is created on higher frequencies.

Part II

Wavelet-based realized co-variation theory

Chapter 7

Introduction

One of the most fundamental issues in finance is research of the covariance generating process between asset returns. Demand for accurate covariance estimation is becoming more important for risk measurement and portfolio optimization than ever before. The increasing availability of high-frequency data for a wide range of securities has allowed a shift from parametric conditional covariance estimation based on daily data toward the model-free measurement of so-called "realized quantities" on intraday data. Using a seminal result in semi-martingale process theory, Andersen et al. (2003) show that realized variance becomes a consistent estimator of integrated variance with increasing sampling frequency under the assumption of zero microstructure noise. Barndorff-Nielsen and Shephard (2004a) generalize the idea to a multivariate setting of so-called "realized covariation" and provide an asymptotic distribution theory for covariance (and correlation) analysis – again with the assumption of zero microstructure noise.

Although the theory is very appealing and intuitive, it assumes that the observed high-frequency data are the true underlying process. But real-world data are contaminated with microstructure noise and jumps, which makes statistical inference difficult. Realized measures suffer from large bias and inconsistency with the presence of noise and jumps in the observed data. The first approach to dealing with noise actually throws away a large amount of data. While this may not seem to be a logical step, the reason can be found quickly when one looks at the data at various sampling frequencies. The higher the frequency of the data we use (i.e., 1 second, 1 tick), the more microstructure noise they contain and the more biased the estimator is. Thus, a lot of researchers use lower frequencies (i.e., 5 minutes), which results in the throwing away of a very large amount of data directly. This is not an appropriate solution for a statistician to use. In the recent literature, a number of ways have been proposed to restore consistency through subsampling, for example Zhang et al. (2005)'s two-scale realized volatility estimator described in the previous part of this thesis. Zhang (2011) generalizes these ideas to a multivariate setting and defines a two-scale covariance estimator. Barndorff-Nielsen et al. (2011) achieve positive semi-definiteness of the variance-covariance matrix using multivariate kernel-based estimation.

While inference under noise and jumps in realized variation theory has been widely studied in recent contributions, its generalization to covariation theory is only now emerging in the literature. Together with important contributions by Zhang (2011) and Barndorff-Nielsen et al. (2011), Griffin and Oomen (2011) and Aït-Sahalia et al. (2010) deal with microstructure noise and non-synchronous trading and propose a consistent and efficient estimator of realized covariance. Audrino and Corsi (2010) propose a forecasting model for realized correlations. This research is becoming very active and stands at the frontier of current research in financial econometrics.

In the second part of this dissertation, we contribute to the current literature and provide a generalization of our wavelet-based realized variation theory contained in the first part. The theoretical results for the univariate setting motivate multivariate volatility modeling and forecasting based on realized covariation measures. This part is organized as follows. The first chapter briefly introduces the general continuous multivariate stochastic volatility semi-martingale processes and theory for the estimation of its variance-covariance matrix. It also discusses the important synchronization of the data and the effects of jumps and co-jumps. The second chapter introduces our new wavelet-based covariation theory together with a methodology for detecting multivariate co-jumps using wavelets. Wavelet decomposition is also used to define wavelet-based realized correlation and beta measures. The estimators are tested in a numerical study in the third chapter, and they are used for the decomposition of empirical volatility in the last chapter. Interesting results on multivariate unconditional volatility distributions as well as the dynamics of the decomposed dependencies and their forecasts are the empirical contributions of this part of the dissertation. Finally, we build a portfolio and study the decomposition of the realized beta.

Chapter 8

General multivariate framework

8.1 Continuous multivariate stochastic volatility semimartingales

Demand for accurate covariance estimation is becoming more important for risk measurement and portfolio optimization than ever before. Interest in extending the univariate results to a multivariate framework is therefore growing. The pioneering contribution in this context has been made by Barndorff-Nielsen and Shephard (2004a), who construct a unified framework for modeling multivariate high-frequency financial data using realized covariation. The authors provide an asymptotic distribution theory for standard methods such as regression, correlation analysis and covariance. Following Barndorff-Nielsen and Shephard (2004a), and the theory introduced in the first part of this thesis, we introduce a multivariate setting.

Consider an m-dimensional logarithmic asset price process defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The return process, evolving in continuous time over the time interval $[t-h, t]$, $0 \leq h \leq t \leq T$, is $\mathbf{r}_{t,h} = \mathbf{p}_t - \mathbf{p}_{t-h}$, where $\mathbf{p}_t = (p_{(1)t}, \ldots, p_{(m)t})'$ denotes an $m \times 1$ vector of log prices at time t and is used as notation for a multivariate price process. We further consider natural information filtration, an increasing family of σ-fields $(\mathcal{F}_t)_{t\in[0,T]} \subseteq \mathcal{F}$, which satisfies the usual conditions. Information set $\mathcal{F}_t$ contains the full history up to time t of the realized values of the asset price and other relevant state variables. A fundamental result of stochastic integration theory states that such a process can be uniquely decomposed according to Proposition 1. We recall the martingale representation theorem 2.5 to use it for the $\mathbf{p}_t$ process.

Definition 12 *Martingale representation theorem*
For any m-dimensional, square-integrable, continuous sample path, logarithmic price process $(\mathbf{p}_t)_{t\in[0,T]}$*, with a continuous sample path and a full rank of the associated* $m \times m$

quadratic variation process, $[r, r]_t$*, there exists a representation such that for all* $0 \leq t \leq T$

$$\mathbf{r_{t,h}} = \int_{t-h}^{t} \boldsymbol{\mu_s} ds + \int_{t-h}^{t} \boldsymbol{\Theta_s} d\boldsymbol{W_s}, \tag{8.1}$$

where $\boldsymbol{\mu}_s$ *is an integrable, predictable and finite-variation* $m \times 1$ *vector,* $\boldsymbol{\Theta}$ *represents a multivariate stochastic volatility process with càdlàg elements, and vector* $\boldsymbol{W_t}$ *is* $m \times 1$ *standard Brownian motion.*

The l-th row of matrix $\boldsymbol{\Theta_t}$ is $(\sigma_{(l,1)t}, \sigma_{(l,2)t}, \ldots, \sigma_{(l,m)t})$. Then the spot (or instantaneous) covariance is defined as:

$$\boldsymbol{\Sigma_t} = \boldsymbol{\Theta_t} \boldsymbol{\Theta_t}', \tag{8.2}$$

satisfying for all $t < \infty$:

$$\int_{t-h}^{t} \Sigma_{(l,l)u} du < \infty, \qquad l = 1, \ldots, m, \tag{8.3}$$

where $\Sigma_{(l,q)t}$ is the (l, q) element of the $\boldsymbol{\Sigma_t}$ process.

The requirement that the $m \times m$ matrix $[r, r]_t$ be of full rank for all t implies that no asset is redundant at any time.[1] The quadratic covariation between the l-th and q-th price processes over $[t-h, t]$, for $0 \leq h \leq t \leq T$, can be expressed as

$$CV_{(l,q)t,h} = \int_{t-h}^{t} \Sigma_{(l,q)s} ds = \int_{t-h}^{t} \sum_{i=1}^{m} \sigma_{(l)i,s} \sigma_{(q)i,s} ds. \tag{8.4}$$

$CV_{(l,q)t,h}$ consistently estimates the integral of the conditional covariance of increments of the local martingale component of $\mathbf{r}_{t,h}$ over the assumed interval. In the special case of $\mathbf{r}_{t,h}$ being univariate ($m = 1$), $CV_{(l,q)t,h}$ represents the quadratic variation of $\boldsymbol{p_{t,h}}$ (Eq. 2.3).

As in the first part of the thesis, we interrupt the theory with an example of the general bivariate jump-diffusion process we will use in this work.

$$dp_{(q)t} = \mu_{(q)t} dt + \sigma_{(q)t} dW_{(q)t} + \xi_{(q)t} dz_{(q)t}, \qquad q = \{1, 2\} \tag{8.5}$$

where z is a constant-intensity Poisson process with jump magnitude controlled by $\xi_t \sim N(\bar{\xi}, \sigma_\xi^2)$, $W_{1,t}$ is a standard Brownian motion, and for all $t \in [0, T]$, $dW_{(2)t} = \rho_t dW_{(1)t} + \sqrt{1 - q_t^2} dW_{(3)t}$, where $W_{(3)t}$ is an independent standard Brownian motion and ρ is a stochastic process with càdlàg paths. Finally, $z_{(1)t}$ and $z_{(2)t}$ are possibly correlated pure jump processes. The quadratic covariation of process 8.5 over the time interval $[t-h, t]$, $0 \leq h \leq t \leq T$, is then

$$CV_{(1,2)t,h} = \underbrace{\int_{t-h}^{t} \rho \sigma_{(1)s} \sigma_{(2)s} ds}_{IC_{t,h}} + \underbrace{\sum_{t-h \leq s \leq t} J_{(1)s} J_{(2)s}}_{\text{Jump Covariation}}, \tag{8.6}$$

where $J_{(q)t} = \xi_{(q)t} dz_{(q)t}$ and is non-zero only if we have co-jumps. Thus the quadratic covariation will be composed of the *Integrated Covariance* and the covariance of common jumps.

[1] This condition is not redundant, see Andersen et al. (2003).

8.2 Estimation of realized covariation

Andersen et al. (2003) suggest estimating the quadratic covariation matrix analogously to the realized variation defined in Part I of this thesis by taking the outer product of the observed high-frequency return over the period.

Definition 13 *Realized covariance*
The realized covariance over $[t-h,t]$*, for* $0 \leq h \leq t \leq T$*, is defined by*

$$\widehat{RC}_{t,h} = \sum_{i=1}^{n} \mathbf{r}_{t-h+\left(\frac{i}{n}\right)h} \mathbf{r}'_{t-h+\left(\frac{i}{n}\right)h}, \tag{8.7}$$

where n *is the number of observations in* $[t-h,t]$*.*

Proposition 16 *Realized covariance as an unbiased covariance estimator*
If the return process is square-integrable and $\mu_t \equiv 0$*, then for any value of* $n \geq 1$ *and* $h > 0$*,*

$$E\left[RC_{t,h} \middle| \mathcal{F}_t\right] = E\left[\widehat{RC}_{t,h} | \mathcal{F}_t\right]. \tag{8.8}$$

Proposition 17 *Consistency of realized covariance*
The realized covariance estimator provides a consistent nonparametric measure of the covolatility,

$$\text{plim}_{n\to\infty} \widehat{RC}_{t,h} = RC_{t,h}, \quad 0 \leq h \leq t \leq T, \tag{8.9}$$

where the convergence is uniform in probability.

Details of these results can be found in Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004a) who show that the *ex-post* realized covariance $\widehat{RC}_{t,h}$ is an unbiased estimator of the *ex-ante* expected covariation $RC_{t,h}$. With increasing sampling frequency, the realized covariance is, moreover, a consistent estimator of the covariation over any fixed time interval $h > 0$, as $n \to \infty$. Conditional on sample path realization of $\mu_{t,h}$ and $\Sigma_{t,h}$, the distribution of the continuously compounded h-period returns follows similarly to Proposition 5.

Definition 14 *Any* m*-dimensional returns process* $\mathbf{r}_{t,h}$ *over* $[t-h,t]$*,* $0 \leq h \leq t \leq T$*, for the continuous sample path 8.1 is distributed as a normal mixture,*

$$\mathbf{r}_{t,h} | \sigma\{\boldsymbol{\mu_{t,h}}, \boldsymbol{\Sigma_{t,h}}\} \sim N\left(\int_{t-h}^{t} \boldsymbol{\mu_s} ds, \int_{t-h}^{t} \boldsymbol{\Sigma_s} ds\right), \tag{8.10}$$

provided that the Brownian motion, $\boldsymbol{W_t}$*, is independent of* $\boldsymbol{\mu_{p_t,\sigma_t}}$ *and* σ_t*.*

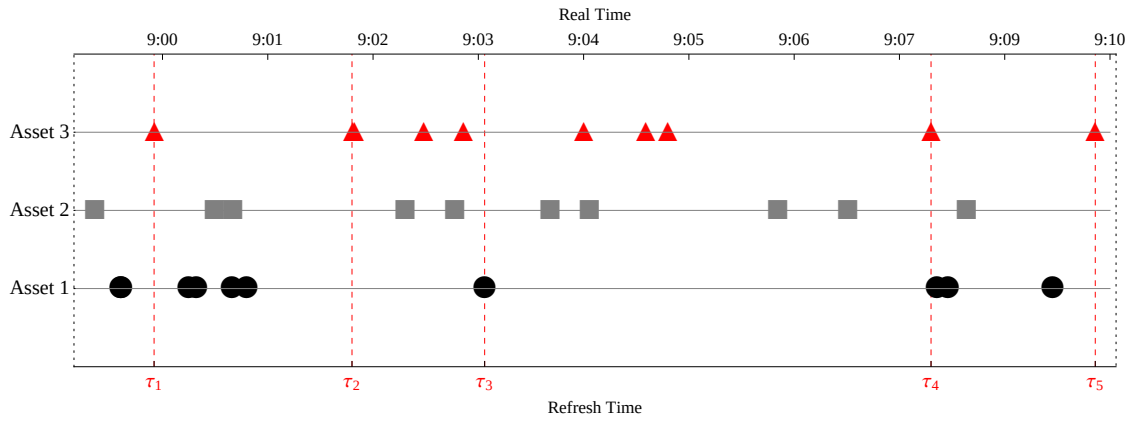
Similarly to its counterpart, realized covariance theory suffers from the issues discussed in depth in Part I. In practice, we observe only discrete prices, thus bias from discretization is unavoidable. Much more damage is caused by market microstructure effects such as price discreteness, bid-ask spread and bid-ask bounce. Thus, when using this estimator

in practice, one is left with advice not to sample too often. While the optimal sampling frequency resulting from the vast research on the noise-to-signal ratio, nicely surveyed by Hansen and Lunde (2006), Bandi and Russell (2006b), McAleer M (2008) and Andersen and Benzoni (2007) and described in Part I, can be used, this approach still causes a large amount of available data to be discarded. As in the univariate case of Zhang et al. (2005)'s two-scale realized volatility estimator, multivariate generalization addresses the problem. Let us present it in the following section.

8.3 Data synchronization: refresh time

One last important assumption about the theory we did not mention is that the data are assumed to be synchronized, meaning that the prices of the m assets were collected at the same time stamp. In practice, trading is non-synchronous, delivering fresh prices at irregularly spaced times which differ across stocks. Research of non-synchronous trading has been an active field of financial econometrics in past years – see, for example, Hayashi and Yoshida (2005) and Voev and Lunde (2007). This practical issue induces bias in the estimators and may be partially responsible for the Epps effect (Epps, 1979), a phenomenon of decreasing empirical correlation between the returns of two different stocks with increasing data sampling frequency.

Example 2 *In the following example, we will illustrate how the Refresh Time scheme works for $d = 3$ assets. Black circles represent the times of the first asset $t_{(1)j}$, grey rectangles represent the times of the second asset $t_{(2)j}$, and red triangles represent the*



times of the third asset $t_{(3)j}$. Then τ_j are the refresh times, while $n_1 = 9$, $n_2 = 10$, $n_3 = 9$. The times were generated randomly just for illustration.

As noted by Barndorff-Nielsen et al. (2011), stale prices are a key feature of estimating covariances, as they induce cross-autocorrelation among asset returns. Thus, the refresh

time scheme, which may be used as a measure of staleness,[2] is appropriate to use in the case of low asset liquidity. Aït-Sahalia et al. (2010) compare various synchronization schemes and find that the estimates do not differ significantly from the estimates from the Refresh Time scheme on the same type of the data we use in our work. Thus, we can restrict ourselves to this synchronization scheme. However, we should note that this scheme is not appropriate in case when the securities are very heterogeneous, in particular if one of then is highly illiquid, when compared to others.

Let $N_{(q)t}$ be the counting process governing the number of observations in the k-th asset up to time t, with times of trades $t_{(q)1}, t_{(q)2}, \ldots$. Following Barndorff-Nielsen et al. (2011), we define the refresh time, which we use later with our estimator. This time scale was first used in a cointegration study of price discovery by Harris et al. (1995).

Definition 15 *Refresh Time for $t \in [0,1]$*
The first refresh time is defined as

$$\tau_1 = \max(t_{(1)1}, \ldots, t_{(d)1}), \tag{8.11}$$

for $d = 1, \ldots, m$ assets, and all subsequent refresh times are defined as

$$\tau_{j+1} = \max(t_{(1)N_{(1)\tau_j}+1}, \ldots, t_{(d)N_{(d)\tau_j}+1}). \tag{8.12}$$

with the resulting Refresh Time sample being of length N, while $n_{(q)} = N_{(q)}$.

τ_1 is thus the first time that all assets record prices, while τ_2 is the first time that all asset prices are refreshed. In the following analysis, we will always set our clock time to τ_j when using the estimators. This approach converts the problem into one where the Refreshed Times' sample size N is determined by the degree of non-synchronicity and $n_{(1)}, \ldots, n_{(d)}$.

8.4 Effects of microstructure noise

While the estimator defined by Definition 8.7 becomes inconsistent in the presence of noise, we follow the approach of Zhang (2011) and generalize the setting from the previous section assuming the presence of noise.

Proposition 18 *Let $(\mathbf{y}_t)_{t\in[0,T]}$ be an $m \times 1$ vector of observed log prices, which will equal the latent, so-called "true log-price process", $dp_{(q)t} = \mu_{(q)t}dt + \sigma_{(q)t}dW_{(q)t}$, $0 \le t \le T$, for $q = 1, \ldots, m$, and will contain microstructure noise represented by the $m \times 1$ vector, $\epsilon_{(q)t}$*

$$\mathbf{y_t} = \mathbf{p_t} + \boldsymbol{\epsilon_t}, \tag{8.13}$$

where $\boldsymbol{\epsilon_t}$ is a zero mean i.i.d. noise vector with variance η^2

[2] More precisely, the average time difference between the refresh time and the real time is the measure of staleness.

The realized covariance estimator suffers from bias when using all the available data. As in the univariate case discussed in detail in section 2.3, Zhang (2011) provides a two time scale realized covariance (TSCV) estimator, which simultaneously cancels out the Epps effect and the effect of microstructure noise. We are not going to discuss concept one again; we only define the final estimator and refer the reader to Section 2.3 for a detailed explanation of the grids idea and to Zhang (2011) for proofs. Moreover, we limit ourselves to a bivariate price process for simplicity of notation. Thus, $\mathbf{p}_{t,h}$ will include the $p_{(1)\tau}$ and $p_{(2)\tau}$ price processes.

First, the average estimator of the realized covariance is defined by:

$$\widehat{RC}_{t,h}^{(S)} = \frac{1}{S}\sum_{s=1}^{N}\left(y_{(1)\tau} - y_{(1)\tau-s}\right)\left(y_{(2)\tau} - y_{(2)\tau-s}\right). \tag{8.14}$$

Analogously to Zhang et al. (2005), the two-scale realized covariance (TSCV) is defined by

$$\widehat{RC}_{t,h}^{(TSCV,adj)} = c_N\left(\widehat{RC}_{t,h}^{(K)} - \frac{\hat{n}_K}{n_J}\widehat{RC}_{t,h}^{(J)}\right), \tag{8.15}$$

where c_N is a constant that can be tuned for small sample performance, and the two scales are chosen such that $1 \leq J << K = O(n^{2/3})$. In the classical two-scale setting, $J = 1$.

Another popular estimator of covariation has been developed recently by Barndorff-Nielsen et al. (2011). The class of their positive semi-definite multivariate realized kernels (MRK) is defined as:

Definition 16 *Multivariate Relized Kernel estimator*
The multivariate realized kernel estimator of variance-covariance matrix over $[t-h, t]$, *for* $0 \leq h \leq t \leq T$ *is defined by*

$$\widehat{RC}_{t,h}^{(MRK)} = \sum_{\eta=-H}^{H} k\left(\frac{\eta}{H}\right)\Gamma_{t,h,\eta}, \qquad \eta \geq 0, \tag{8.16}$$

and $\Gamma_{t,h,\eta} = \Gamma'_{t,h,-\eta}$ *for* $\eta < 0$, *where* $\Gamma_{t,h,\eta} = \sum_{i=\eta+1}^{N}\mathbf{y}_{t,h,i}\mathbf{y}'_{t,h,i}$ *denoting the* η*-th realized autocovariance and* $k(.)$ *denotes the kernel function.*

As in the univariate case, we use the Parzen kernel function defined by:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1. \\ 0 & x > 1 \end{cases} \tag{8.17}$$

By introducing the TSCV and MRK estimators, we will be able to consistently estimate the realized covariation from noisy observations. The last important step in the analysis is the measurement of jumps and co-jumps in the process. Before we introduce our approach to dealing with jumps, we follow the structure of the text in Part I and discuss the importance of co-jump detection in a multivariate setting.

8.5 Multivariate jumps

The recent empirical interest in jumps is motivated by the significant consequences of a possibly misspecified price process. Barndorff-Nielsen and Shephard (2006) introduce a test based on the difference between the bipower variation and the quadratic variation, but the work is currently unfinished. Andersen et al. (2007) and Huang and Tauchen (2005) present a study of multipower variations in order to assess the proportion of the quadratic variation attributable to jumps. Andersen et al. (2007) and Lee and Mykland (2008) introduce two very similar procedures which compare intraday returns to a local volatility measure. Fan and Wang (2007) develop the wavelet methods introduced in the first part of this dissertation. Jiang and Oomen (2008) construct a test based on the hedging error of a variance swap replication strategy. Aït-Sahalia and Jacod (2009) propose an estimator of truncated power variations computed at different sampling frequencies. Finally, Andersen et al. (2009) introduce a test for jumps constructed using the MedRV and MinRV measures. Other tests include Mancini (2009) and Lee and Hannig (2010). The harm imposed by ignoring jumps and co-jumps in assumed price processes can be large, especially with regard to forecasting, option pricing, portfolio risk management and credit risk management.

Jumps in the price process are assumed to have finite activity. The resulting process is a Brownian semi-martingale plus a finite activity jump process and takes the form:

$$\mathbf{r_{t,h}} = \int_{t-h}^{t} \boldsymbol{\mu}_s ds + \int_{t-h}^{t} \boldsymbol{\Theta_s} d\boldsymbol{W_s} + \sum_{j=1}^{N_t} J_j, \tag{8.18}$$

where the counting process N is for all $t < \infty$: $N_t < \infty$, and also $\sum_{j=1}^{N_t} J_{(k)j}^2 < \infty$, $k = 1, \ldots, m$.

The quadratic variation of this process will be:

$$\boldsymbol{QV_{t,h}} = \underbrace{\int_{t-h}^{t} \boldsymbol{\Sigma_s} ds}_{IC_{t,h}} + \underbrace{\sum_{t-h \le s \le t} \boldsymbol{J_s J_s'}}_{\text{Jump Covariation}} \quad , \tag{8.19}$$

where the l,q-th element of $\boldsymbol{QV_{t,h}}$ (we drop time t, h here for ease of notation) is the l, q-th quadratic covariation $[r]_{(l,q)} = [r_{(l)}, r_{(q)}]$ and the quadratic variation of $r_{(l)}$ is $[r_{(l)}, r_{(l)}] = [r_{(l)}]$:

$$\boldsymbol{QV} = \begin{pmatrix} [r_{(1)}] & [r_{(1)}, r_{(2)}] & \cdots & [r_{(1)}, r_{(m)}] \\ [r_{(2)}, r_{(1)}] & [r_{(2)}] & \cdots & [r_{(2)}, r_{(m)}] \\ \vdots & \vdots & \ddots & \vdots \\ [r_{(m)}, r_{(1)}] & [r_{(m)}, r_{(2)}] & \cdots & [r_{(m)}] \end{pmatrix} \tag{8.20}$$

While Barndorff-Nielsen and Shephard (2004b) extend the concepts of bipower variation and realized variance to multivariate settings, the corresponding multivariate test is currently incomplete. When using this concept, we will limit ourselves to generalization of $\widehat{RV}_T^{(BV)}$ (2.27):

Definition 17 *Bipower covariation estimator*
The bipower covariation over $[t-h,t]$*, for* $0 \leq h \leq t \leq T$*, is defined by*

$$\begin{aligned}\widehat{RC}_{t,h}^{(BC)} &= \frac{\pi}{8}\sum_{i=2}^{n} |r_{(1)t-h+\left(\frac{i}{n}\right)h} + r_{(2)t-h+\left(\frac{i}{n}\right)h}|.|r_{(1)t-h+\left(\frac{i-1}{n}\right)h} + r_{(2)t-h+\left(\frac{i-1}{n}\right)h}| - \\ &\quad |r_{(1)t-h+\left(\frac{i}{n}\right)h} - r_{(2)t-h+\left(\frac{i}{n}\right)h}|.|r_{(1)t-h+\left(\frac{i-1}{n}\right)h} - r_{(2)t-h+\left(\frac{i-1}{n}\right)h}| \qquad (8.21)\end{aligned}$$

In our work, we focus on generalization of Fan and Wang (2007)'s approach used in Part I of this thesis to a multivariate setting.

Chapter 9

Wavelet-based covariation theory

In this part, we will introduce a generalization of our findings from Chapter 3 to a multivariate setting. We will draw on the theory introduced in Chapter 3 and use it to determine its multivariate counterpart.

The chapter presents a complete wavelet-based covariation theory extending our results from the univariate part and the results of Fan and Wang (2007), who first bring the wavelet-based realized variation estimator to the literature. In our work, we generalize the results of Fan and Wang (2007) in several ways. Instead of using the Discrete Wavelet Transform we use the Maximum Overlap Discrete Wavelet Transform, which is a more efficient estimator and is not restricted to sample sizes that are powers of two. We also use the Daubechies family of wavelets instead of the Haar type in our work. Our biggest contribution is in providing a complete theoretical framework for the estimation of multivariate wavelet realized covariation, as this cannot be found in the current literature. The theory improves the efficiency of the estimated covariances. We build on the results and also define new measures of correlation and realized beta based on our wavelet-based estimators.

9.1 Wavelet covariance

Let X_i and Y_i be covariance stationary processes from two intraday returns, with the square integrable spectral density functions $S_{(X)}(.)$, $S_{(Y)}(.)$ and cross spectra $S_{(XY)}(.)$. Furthermore, let us define $S_{j,(XY)}(0)$ as the spectral density function of the product of the j-th level MODWT wavelet coefficients, unaffected by the boundary conditions, $(\widetilde{W}_{(X)j,i}, \widetilde{W}_{(Y)j,i})$, at zero frequency. Since we use the Daubechies family of wavelets, the processes can be stationary after the d_Xth and d_Yth difference, $d \leq L$. Let $\widetilde{W}_{(X)j,i}, \widetilde{W}_{(Y)j,i},\ j = 1, ..., J$ and $i \geq L_j - 1$[1], be the MODWT wavelet coefficients of X_i and Y_i, which are unaffected by the boundary conditions. The wavelet covariance of X_i and Y_i at level j is then defined as:

[1] $L_j - 1$ denotes the length of a wavelet filter at level j.

$$\gamma_{(XY)j} = Cov(\widetilde{W}_{(X)j,i},\, \widetilde{W}_{(Y)j,i}). \tag{9.1}$$

The covariance $\gamma_{(XY)j}$ can be decomposed with the wavelet covariance. For a particular level of decomposition $J \leq log_2(N)$, the covariance $\gamma_{(XY)j}$ is the sum of the covariances of the MODWT wavelet and the scaling coefficients:

$$Cov(X_i, Y_i) = Cov(\widetilde{V}_{(X)J,i},\, \widetilde{V}_{(Y)J,i}) + \sum_{j=1}^{J} \gamma_{(XY)j} \tag{9.2}$$

In the case where $J \to \infty$, the covariance of the scaling coefficients $(\widetilde{V}_{(X)J,i},\, \widetilde{V}_{(Y)J,i})$ goes to zero (Whitcher et al., 1999), hence we can rewrite 9.2 as:

Proposition 19 *Wavelet covariance*
Let $J \to \infty$. Then, the covariance between processes X_i and Y_i is

$$Cov(X_i, Y_i) = \sum_{j=1}^{\infty} \gamma_{(XY)j}. \tag{9.3}$$

The **proof** is provided in Appendix 11A.8

9.1.1 Estimator of wavelet covariance

For processes X_i and Y_i defined above, the estimator of wavelet covariance at level j is defined as

$$\hat{\gamma}_{(XY)j} = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{(X)j,i} \widetilde{W}_{(Y)j,i}, \tag{9.4}$$

where $M_j = N - L_j + 1 > 0$ is the number of j-th level MODWT coefficients for both processes that are unaffected by the boundary conditions. Whitcher et al. (1999) prove that for the Gaussian processes X_i and Y_i, the MODWT-based estimator of wavelet covariance is unbiased and asymptotically normally distributed, i.e.,

$$\hat{\gamma}_{(XY)j} \sim N\left(\gamma_{(XY)j}, M_j^{-1} S_{(XY),j}(0)\right), \tag{9.5}$$

where $S_{(XY),j}(0)$ denotes the spectral density function for the product of the j-th level MODWT wavelet coefficients $(\widetilde{W}_{(X)j,i} \widetilde{W}_{(Y)j,i})$ at zero frequency, i.e.,

$$S_{(XY)j}(0) = \int_{-1/2}^{1/2} S_{X(j)}(f) S_{(Y)j}(f) df + \int_{-1/2}^{1/2} S^2_{(XY)j}(f) df. \tag{9.6}$$

For more details see Lindsay et al. (1996), Whitcher et al. (1999) and Whitcher et al. (2000).

The complete proof of 9.5 can be found in Whitcher et al. (1999). The idea of the proof is to show that using the Daubechies wavelet filter $L \geq 2d$, the covariance of processes $(\widetilde{W}_{(X)j,i}, \widetilde{W}_{(Y)j,i})$ is absolutely summable. Since the estimator of the MODWT wavelet variance is in fact the sample mean of $\overline{W}_{(XY)j,i} \equiv \overline{W}_{(X)j,i}\overline{W}_{(Y)j,i}$, this process is an absolutely summable sequence, (Brillinger, 1981) (Theorem 2.9.1.). Following again Brillinger (1981) (Theorem 4.4.1) gives the result 9.5.

If the assumption of Gaussianity of X_i and Y_i is relaxed, result 9.5 still holds for a large class of non-Gaussian processes, but for estimation of the spectral density function $S_{(XY)j}(0)$ it is preferable to use the multitaper spectrum instead of the periodogram (Serroukh et al., 2000).

9.2 Wavelet-based realized covariance estimation

In the previous section we used wavelets to decompose the realized variance to obtain information about the behavior (energy contribution) on every scale. We can use a similar approach to decompose the realized covariance to a get deeper knowledge about the dependence between the two processes examined. The realized wavelet covariance (using the MODWT) is a scale by scale decomposition of the realized covariance defined by Definition 8.7.

Definition 18 *Wavelet-based realized covariance estimator*
The realized wavelet covariation of the l-th and q-th asset return from the m-dimensional vector $\mathbf{r}_{t,h}$ *over* $[t-h,t]$*, for* $0 \leq h \leq t \leq T$*, can be defined as*

$$\widehat{RC}^{(WRC)}_{(l,q)t,h} = \sum_{j=1}^{J_s+1} \sum_{k=1}^{n} \widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h} \widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}, \tag{9.7}$$

where n is the number of intraday observations over $[t-h,t]$ *and* J_s *is the number of scales considered.* $\widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}$ *are the MODWT coefficients defined in 3.26 on scales* $j = 1, ..., J_s + 1$*, where* $J_s \leq \log_2 n$*.*

Proposition 20 *WRC as an unbiased covariance estimator*
If the return process is square-integrable and $\mu_t \equiv 0$*, then for any value of* $n \geq 1$*,*

$$E\left[RC_{(l,q)t,h}\middle|\,\mathcal{F}_t\right] = E\left[\widehat{RC}^{(WRC)}_{(l,q)t,h}\middle|\,\mathcal{F}_t\right]. \tag{9.8}$$

Proposition 21 *Consistency of WRC*
The realized covariance estimator provides a consistent nonparametric measure of the covolatility,

$$\text{plim}_{n\to\infty}\, \widehat{RC}^{(WRC)}_{(l,q)t,h} = RC_{(l,q)t,h}, \quad 0 \leq h \leq t \leq T. \tag{9.9}$$

The proof is provided in Appendix 11A.9.

In fact, this extension directly mirrors the univariate setting, while $\widehat{RC}^{(WRC)}_{(l,l)t,h}$ is in fact the variance of the l-th process, thus $\widehat{RV}^{(WRV)}_{(l,l)t,h}$. Having defined the basic tool for our analysis – a wavelet-based realized covariation estimator which is able to consistently estimate the integrated covariation of process 8.1, where we have no jumps– we generalize the jump detection test presented in the univariate setting to the multivariate setting in the following section.

9.3 Disentangling jumps from co-jumps

Fan and Wang (2007) first proposed the use of wavelets to estimate jumps in high-frequency data. Their work is described in detail in Section 3.3, and we use it in our univariate estimator. In this part, we generalize this concept to a multivariate concept. We detect all jumps in the m assets separately using wavelet decomposition, and then we estimate the co-jumps. Let us define the procedure.

Definition 19 *Multivariate jump estimation using wavelets*
Let $\widetilde{\mathcal{W}}_{(q)1,k}$ be the 1^{st} level wavelet coefficients of $(y_{(q)t})_{t\in[0,T]}$ from Proposition 18. If for some $\widetilde{\mathcal{W}}_{(q)1,k}$

$$|\widetilde{\mathcal{W}}_{(q)1,k}| > \frac{median\{|\widetilde{\mathcal{W}}_{(q)1,k}|, k = 1, \dots, n\}}{0.6745}\sqrt{2\log n}, \tag{9.10}$$

for $q = 1, \dots, m$ assets, then $\hat{\tau}_{(q)l} = \{k\}$ is the estimated jump location with size $\bar{y}_{(q)\hat{\tau}_{l+}} - \bar{y}_{(q)\hat{\tau}_{l-}}$ (averages over $[\hat{\tau}_{(q)l}, \hat{\tau}_{(q)l}+\delta_n]$ and $[\hat{\tau}_{(q)l}, \hat{\tau}_{(q)l}-\delta_n]$, respectively, with $\delta_n > 0$ being the small neighborhood of the estimated jump location $\hat{\tau}_{(q)l} \pm \delta_n$; 0.6745 is a robust estimate of the standard deviation).

The jump variation of the q-th asset is then estimated by the sum of the squares of all its estimated jump sizes:

$$\widehat{MWJC}_{(q)} = \sum_{l=1}^{N_t} (\bar{y}_{(q)\hat{\tau}_{l+}} - \bar{y}_{(q)\hat{\tau}_{l-}})^2. \tag{9.11}$$

Following the theory in Fan and Wang (2007), we can say that $\widehat{MWJC}_{(j)}$ will be a consistent estimator of the jumps for all q assets in $\mathbf{p}_t$.

Proposition 22 *Consistency of multivariate wavelet jump estimator*
With $n \to \infty$

$$\text{plim}_{n\to\infty}\, \widehat{MWJC}_{(q)} = \sum_{l=1}^{N_t} J^2_{(q),l}, \tag{9.12}$$

with the convergence rate $N^{-1/4}$.

Once we have estimated all independent jumps in the studied $\mathbf{p}_t$ vector, we can propose an analysis of co-jumping in the series. The idea is to compare all the jump locations, and those which are the same across all $q = 1, \ldots, m$ assets in some small neighborhood will be co-jumps.

Definition 20 *Wavelet co-jump estimation*
Let $\hat{\tau}_{(q)l}$ be the estimated jump locations of $(y_{(q)t})_{t\in[0,T]}$ for all $q = 1, \ldots, m$ using Definition 19. Then co-jump location $\hat{\tau}_l^* = \{k\}$ can be estimated as:

$$\hat{\tau}_{(q)l} - \delta_n < \hat{\tau}_l^* < \hat{\tau}_{(q)l} + \delta_n, \quad \textit{for all} \quad q = 1, \ldots, m. \tag{9.13}$$

Co-jumps are particularly important in portfolio theory. For a well diversified large portfolio in the sense of the Arbitrage Pricing Theory, idiosyncratic jumps are diversified away, but common jumps, or co-jumps, remain a problem. Thus in the following subsection, we illustrate our technique on a portfolio multivariate extension.

9.4 Wavelet-based realized covariance estimator robust to jumps and noise

Finally, using all the presented theory, we propose an estimator of covariance which is robust to noise and also is able to deal with jumps in the data. Moreover, we will be able to decompose the integrated covariance into J_s components using our estimator. Similarly to the univariate version, we build the estimator on the Zhang (2011) generalization of TSRV to TSCV, connecting our wavelet-based decomposition of realized quantities as well as jump detection.

Definition 21 *Jump wavelet TSCV (JWTSCV) estimator*
Let $\widehat{RC}_{(l,q)t,h}^{(estimator,J)}$ denote an estimator of the realized covariance between the l-th and q-th asset return on the jump-adjusted observed data, $\mathbf{y_{t,h}}^{(J)} = \mathbf{y}_{t,h} - \widehat{\mathbf{MWJC}}$. The jump-adjusted wavelet two-scale realized covariance estimator is defined as:

$$\widehat{RC}_{(l,q)t,h}^{(JWTSCV)} = c_N \left(\widehat{RC}_{(l,q)t,h}^{(WRC,J)} - \frac{\bar{n_G}}{n_S} \widehat{RC}_{(l,q)t,h}^{(S,J)} \right), \tag{9.14}$$

where $\widehat{RC}_{(l,q)t,h}^{(WRC,J)} = \frac{1}{G}\sum_{g=1}^{G}\sum_{j=1}^{J_s+1}\sum_{k=1}^{n} \widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h} \widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}$ obtained from wavelet coefficient estimates using the MODWT on a grid of size $\bar{n} = n/G$ on the jump-adjusted observed data, $\mathbf{y}_{t,h}^{(J)} = \mathbf{y}_{t,h} - \widehat{\mathbf{MWJC}}$, and c_N is a constant that can be tuned for small sample performance.

Proposition 23 *JWTSCV unbiased covariance estimator*
If the return process is square-integrable and $\mu_t \equiv 0$, then for any value of $n \geq 1$,

$$E\left[RC_{(l,q)t,h}|\mathcal{F}_t\right] = E\left[\widehat{RC}_{(l,q)t,h}^{(JWTSCV)}|\mathcal{F}_t\right]. \tag{9.15}$$

Proposition 24 *Consistency of JWTSCV*
The wavelet realized covariance provides a consistent nonparametric measure of the covariation,

$$\text{plim}_{n\to\infty} \widehat{RC}^{(JWTSCV)}_{(l,q)t,h} = RC_{(l,q)t,h}, \quad 0 \le h \le t \le T, \tag{9.16}$$

where the convergence is uniform in probability.

The **proof** of Proposition 23 and Proposition 24 is provided in Appendix 11A.10.

Estimator 9.14 converges in probability to the *true* integrated covariance of 8.18, which is of primary interest in this analysis. Thus we have defined a new wavelet-based covariation theory which is able to estimate realized covariation consistently in the presence of noise and jumps. In the next section, we use this theory to propose estimators of covariance and realized beta, which are important for financial practitioners .

9.5 Wavelet-based realized beta and correlation

In the final section, we would like to introduce an extension of some transformations of the realized covariation matrix using the wavelet theory. Specifically, we focus on the realized regression, or realized beta and realized correlation. The theory from the previous text allows us to study the observed quantities of beta and correlation, which is very convenient and helpful for financial practitioners. Basic results are introduced by Andersen et al. (2001), while Barndorff-Nielsen and Shephard (2004a) also provide an asymptotic theory for the estimators. We will briefly introduce the results and extend them using our wavelet-based estimators.

9.5.1 Realized beta

Portfolio beta has played a central role in theoretical and empirical financial econometrics for nearly half a century. It is used to determine systematic risk in the one-factor capital asset pricing model (CAPM) – see, for example, Sharpe (1963) or Lintner (1965). Although CAPM has been heavily criticized in the literature (e.g. Fama and French, 1993), we should not be blinded, as it remains at the frontier of the research and is also used in industry applications.

Measures of realized variation simply provide us with a very convenient tool, as they can be used to observe the time-varying beta directly from the data. Let's introduce the concept.

Definition 22 *Estimator of realized beta*
The realized beta of the l-th and q-th asset over $[t-h, t]$*, for* $0 \le h \le t \le T$*, can be estimated as*

$$\hat{\beta}_{(l,q)t,h} = \frac{\sum_{i=1}^{n} r_{(l)t-h+\left(\frac{i}{n}\right)h} r_{(q)t-h+\left(\frac{i}{n}\right)h}}{\sum_{i=1}^{n} r^2_{(q)t-h+\left(\frac{i}{n}\right)h}}, \tag{9.17}$$

where n is the number of observations.

The probability limit follows from the theory of quadratic variation.

Proposition 25 *Consistency of realized beta estimator*

$$\text{plim}_{n\to\infty}\,\hat{\beta}_{(l,q)t,h} = \beta_{(l,q)t,h} = \frac{\int_{t-h}^{t}\Sigma_{(l,q)u}du}{\int_{t-h}^{t}\Sigma_{(q,q)u}du} \tag{9.18}$$

Under appropriate regularity conditions that allow for non-stationarity in the series, Barndorff-Nielsen and Shephard (2004a) derive the limiting distribution of the realized beta.

Proposition 26 *With sampling frequency $n \to \infty$,*

$$\frac{\hat{\beta}_{(l,q)t,h} - \beta_{(l,q)t,h}}{\sqrt{\left(\sum_{i=1}^{n} r^{2}_{(l)t-h+\left(\frac{i}{n}\right)h}\right)^{-2}\hat{g}_{(l,q)t,h}}} \sim N(0,1), \tag{9.19}$$

where $\hat{g}_{(l,q)t,h} = \sum_{i=1}^{n} x^{2}_{t-h+\left(\frac{i}{n}\right)h} - \sum_{i=1}^{n-1} x_{t-h+\left(\frac{i}{n}\right)h}x_{t-h+\left(\frac{i+1}{n}\right)h}$,
where $x_{t-h+\left(\frac{i}{n}\right)h} = r_{(l)t-h+\left(\frac{i}{n}\right)h}r_{(q)t-h+\left(\frac{i}{n}\right)h} - \hat{\beta}_{(l,q)t,h}r^{2}_{(q)t-h+\left(\frac{i}{n}\right)h}$.

The proof of this result can be found in Barndorff-Nielsen and Shephard (2004a).

Thus, a feasible α-percent confidence interval of the estimated beta can be comfortably recovered from the data.

9.5.2 Realized correlation

Another transformation of the realized covariation matrix which is an important tool for practitioners, is the correlation measure. While correlation is important for portfolio diversification and risk management, understanding its dynamics can have a crucial impact, as correlation changes in time.

The realized covariation matrix provides a very convenient and simple-to-use method which measures the non-parametrically observed correlation from the data. The concept is similar to realized beta.

Definition 23 *Estimator of realized correlation*
The realized correlation between the l-th and q-th asset over $[t-h, t]$, for $0 \le h \le t \le T$, can be estimated as

$$\hat{\rho}_{(l,q)t,h} = \frac{\sum_{i=1}^{n} r_{(l)t-h+\left(\frac{i}{n}\right)h}r_{(q)t-h+\left(\frac{i}{n}\right)h}}{\sqrt{\sum_{i=1}^{n} r^{2}_{(l)t-h+\left(\frac{i}{n}\right)h}\sum_{i=1}^{n} r^{2}_{(q)t-h+\left(\frac{i}{n}\right)h}}}, \tag{9.20}$$

where n is the number of observations.

The probability limit follows from the theory of quadratic variation, as shown, for example, by Andersen et al. (2001).

Proposition 27 *Consistency of realized correlation estimator*

$$\text{plim}_{n\to\infty}\,\hat{\rho}_{(l,q)t,h} = \rho_{(l,q)t,h} = \frac{\int_{t-h}^{t}\Sigma_{(l,q)u}du}{\sqrt{\int_{t-h}^{t}\Sigma_{(l,l)u}du\int_{t-h}^{t}\Sigma_{(q,q)u}du}} \tag{9.21}$$

Under appropriate regularity conditions that allow for non-stationarity in the series, Barndorff-Nielsen and Shephard (2004a) derive the limiting distribution of the realized correlation.

Proposition 28 *With sampling frequency $n \to \infty$,*

$$\frac{\hat{\rho}_{(l,q)t,h} - \rho_{(l,q)t,h}}{\sqrt{\left(\sum_{i=1}^{n} r^2_{(l)t-h+\left(\frac{i}{n}\right)h}\sum_{i=1}^{n} r^2_{(q)t-h+\left(\frac{i}{n}\right)h}\right)^{-1}\hat{g}_{(l,q)t,h}}} \sim N(0,1) \tag{9.22}$$

where $\hat{g}_{(l,q)t,h} = \sum_{i=1}^{n} x^2_{t-h+\left(\frac{i}{n}\right)h} - \sum_{i=1}^{n-1} x_{t-h+\left(\frac{i}{n}\right)h}x_{t-h+\left(\frac{i+1}{n}\right)h}$,
where $x_{t-h+\left(\frac{i}{n}\right)h} = r_{(l)t-h+\left(\frac{i}{n}\right)h}r_{(q)t-h+\left(\frac{i}{n}\right)h} - \frac{1}{2}\hat{\beta}_{(l,q)t,h}r^2_{(l)t-h+\left(\frac{i}{n}\right)h} - \frac{1}{2}\hat{\beta}_{(l,q)t,h}r^2_{(q)t-h+\left(\frac{i}{n}\right)h}$,
with $\hat{\beta}_{(l,q)t,h}$ *defined by Definition 22.*

The proof of this result can be found in Barndorff-Nielsen and Shephard (2004a).

Thus, a feasible α-percent confidence interval of correlation can be easily estimated from the data. In the following part, we will build on these transformations of the realized covariation matrix and decompose them using realized wavelet estimators.

9.5.3 Wavelet-based realized correlation

Finally, we will use our estimators to generalize the realized correlations and beta measures so that we can decompose them into several investment horizons. Using wavelet analysis, we are able to decompose the estimated realized correlation (Definition 23) to get valuable knowledge (linear dependencies) about the two examined processes in the time-frequency domain.

Plugging our results from the previous sections, we can simply use the MODWT decomposition to estimate the correlation using our JWTSCV and JWTSRV estimators.

Definition 24 *Wavelet-based realized correlation*

$$\hat{\rho}^{(JWR)}_{(l,q)t,h} = \frac{\widehat{RC}^{(JWTSCV)}_{(l,q)t,h}}{\sqrt{\widehat{RV}^{(JWTSRV)}_{(l)t,h}\widehat{RV}^{(JWTSRV)}_{(q)t,h}}}, \tag{9.23}$$

where $\widehat{RC}^{(JWTSCV)}_{(l,q)t,h}$ *is defined by Definition 9.14 and estimates the covariation between the l-th and q-th series, and* $\widehat{RV}^{(JWTSRV)}_{(l)t,h}$ *and* $\widehat{RV}^{(JWTSRV)}_{(q)t,h}$ *are defined by Definition 3.44 and estimate the variation of the l-th and q-th series, respectively.*

In contrast to the energy of the squared wavelet coefficients or the wavelet covariance, we cannot simply decompose the wavelet correlation on all scales to get the correlation of the two processes examined. To obtain the contribution of every particular level j to the overall correlation, we have to weight the wavelet correlations at each level by their energy contributions.

Definition 25 *Decomposition of wavelet-based realized correlation*
The wavelet correlation $\hat{\rho}^{(JWR)}_{(l,q)t,h}$ *between the l-th and q-th process can be further decomposed into levels* $j = 1, \ldots, J_s + 1$ *according to their energy contributions as*

$$\widehat{\Upsilon}^{(JWR)}_{(l,q)t,h,j} = \eta_{t,h,j}\widehat{\rho}^{(JWR)}_{(l,q)t,h,j}, \tag{9.24}$$

where $\eta_{t,h,j}$ *is the weight vector of the contributions, defined as*

$$\eta_{t,h,j} = \sqrt{\frac{\sum_{k=1}^{n}\left(\widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h}\right)^2}{\sum_{j=1}^{J_s+1}\sum_{k=1}^{n}\left(\widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h}\right)^2}\frac{\sum_{k=1}^{n}\left(\widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}\right)^2}{\sum_{j=1}^{J_s+1}\sum_{k=1}^{n}\left(\widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}\right)^2}}, \tag{9.25}$$

which is the geometrical mean of the energy contributions of the l-th and q-th process on all scales, so the vector must sum to one: $\sum_{j=1}^{J}\eta_j = 1$.

Hence, after summation of all parts of the wavelet realized correlations, we get the overall correlation. The JWTSRV is an unbiased and consistent estimator of the RV and the JWTSCV is an unbiased and consistent estimator of the RC. The wavelet-based realized correlation converges in probability to the realized correlation.

$$\text{plim}_{n\to\infty}\sum_{j=1}^{J_s+1}\widehat{\Upsilon}^{(JWR)}_{(l,q)t,h,j} = \rho_{(l,q)t,h} \tag{9.26}$$

9.5.4 Wavelet-based realized beta

Similarly to the wavelet-based realized correlation, we define the wavelet-based realized beta based on our estimators.

Definition 26 *Wavelet-based realized beta*
The wavelet-based realized beta of q-th and l-th asset can be estimated as

$$\hat{\beta}^{(JWR)}_{(l,q)t,h} = \frac{\widehat{RC}^{(JWTSCV)}_{(l,q)t,h}}{\widehat{RV}^{(JWTSRV)}_{(q)t,h}}, \tag{9.27}$$

where n is number of observations.

The wavelet-based realized beta can be simply decomposed in the same way as the JWTSRV and JWTSCV. The JWTSRV is an unbiased and consistent estimator of the RV and the JWTSCV is an unbiased and consistent estimator of the RC. The wavelet-based realized beta converges in probability to the realized beta

$$\text{plim}_{n\to\infty}\,\hat{\beta}^{(JWR)}_{(l,q)t,h} = \beta_{(l,q)t,h} \tag{9.28}$$

9.6 Numerical study on small sample performance of the estimators

In this section, we would like to perform a simulation analysis to assess the case where we do not have $n \to \infty$. Small sample studies are important for these types of estimators, as we are always left only with a data sample of finite length. In the simulation, we follow the setup of Barndorff-Nielsen et al. (2011) and simulate a bivariate factor stochastic volatility model for $X_{i,t}$, for $i = \{1, 2\}$ and $t \in [0, 1]$. Moreover, we consider jumps:

$$\begin{aligned} dX_{i,t} &= \mu_i dt + \gamma_i \sigma_{i,t} dB_{i,t} + \sqrt{1-\gamma_i^2}\sigma_{i,t} dW_t + c_{i,t} dN_{i,t} \\ d\sigma_{i,t} &= \exp(\beta_0 + \beta_1 v_{i,t}) \\ dv_{i,t} &= \alpha v_{i,t} dt + dB_{i,t}, \end{aligned} \tag{9.29}$$

where the elements of $B_{i,t}$ are independent standard Brownian motions and are also independent of W_t, and $c_{i,t}dN_{i,t}$ are independent compound Poisson processes with random jump sizes distributed as $N \sim (0, \sigma_{1,J})$. We simulate the processes using the Euler scheme at a time interval of $\delta = 1s$, each with $6.5 \times 60 \times 60$ steps $n = 23,400$, corresponding to a 6.5 trading hour day. The parameters are set to $(\mu_1, \mu_2, \beta_0, \beta_1, \alpha, \gamma_1, \gamma_2) = (0, 0, -5/16, 1/8, -1/40, -0.3, -0.3)$. Each day is restarted with the initial value of $v_{i,t}$ drawn from a normal distribution $N(0, (-2\alpha)^{-1})$. On each simulated path, we estimate $\widehat{\Sigma}_t$ over $T = 1$ day. The results are computed for sampling of 1 minute, 5 minutes, 30 minutes and 1 hour for the realized covariance defined in 8.7, the bipower realized covariance defined by 8.21, the two-scale realized covariance defined by 8.15, the multivariate realized kernel defined by 8.16 and our jump wavelet two-scale realized covariance estimator defined by 9.14. For convenience, we refer to the estimators in the description of the results as RC, BC, TSCV, MRK and JWTSCV respectively.

We repeat the simulations with different levels of noise as well as different numbers of jumps, assuming the market microstructure noise, ϵ_t, to be normally distributed with different standard deviations: $(E[\epsilon^2])^{1/2}$=$\{0,0.0015\}$. Thus, we consider simulations with zero noise and 0.15% of the value of the asset price level noise. We also add different levels of jumps, controlled by intensity λ from the Poisson process $c_{i,t}dN_{i,t}$, starting with $\lambda = 0$, and continue adding jumps with a size implying a one standard deviation jump change. Moreover, co-jumps and individual jumps are separated as these have very

different impacts on the covariance and correlation measures. We start by simulating prices with only a single co-jump, followed by the addition of one jump to each of the bivariate series which are independent of each other, as these two types have different impacts on the covariation and correlation. Individual jumps cause positive bias in the realized covariance measures, so if the estimator is not able to deal with it, the correlation will suffer from large negative bias, as the denominator of the fraction will be larger. On the other hand, co-jumps will logically cause positive bias to the correlation measure, as the numerator, the covariance, will be positively biased.

Finally, we compare the bias of all the estimators for covariance and for correlation.

The true spot correlation between $X_{1,t}$ and $X_{2,t}$ without noise and jumps is used for comparison reference: $\sqrt{(1-\gamma_1^2)(1-\gamma_2^2)}$, which is equal to 0.91 in our case. The full spot covariance matrix $\Sigma_t = \begin{pmatrix} \Sigma_t^{11} & \Sigma_t^{12} \\ \Sigma_t^{12} & \Sigma_t^{22} \end{pmatrix} = \begin{pmatrix} \sigma_{(1)t}^2 & \sigma_{(1,2),t} \\ \sigma_{(1,2),t} & \sigma_{(2)t}^2 \end{pmatrix}$, where $\sigma_{(1,2),t} = \sigma_{1,t}\sigma_{2,t}\rho_t$, is used for computing the correlation bias in the simulations. We also report the covariation bias, Σ_t^{12}.

The results for covariation are reported in Table 9.1. We can see that our JWTSCV proves to be the most efficient estimator, as it is robust to noise as well as jumps at all sampling frequencies. The bipower realized covariance measure (BC) is able to deal with jumps to some extent, while, as expected, the two-scale realized covariance (TSCV) suffers from bias introduced by the co-jumps. The multivariate realized kernels (MRK) also suffers from bias introduced by the co-jumps. However, it is less pronounced with decreasing sampling frequency.

It is interesting to note that the realized covariance is significantly affected by co-jumps in processes, while individual jumps do not have such an effect. Our JWTSCV estimator may thus be very important for portfolio theory, as it is able to consistently estimate the true covariation of the processes as well as all jumps and co-jumps. In the application section, we will use this approach and see how it can improve the results on real data. Interestingly, the sampling frequencies do not reveal any patterns. This is probably caused by the effect of quite large jumps in the simulations. In the case without jumps and noise we can see that the covariance measures get more biased with increasing sampling frequency.

The biases of all the estimators when calculating the correlation are reported in Table 9.2. The results confirm our expectations. The realized covariance measure (RC) is hugely biased when a single large jump is put to the series independently. The bias is really huge, as a true correlation of 0.91 is considered and a correlation of only about 0.4 is estimated. On the other hand, co-jumps cause slight positive bias to the correlation measure. The TSCV and MRK estimators generate similar results. The BC reduces the bias significantly and demonstrates nice finite sample performance under both individual jumps and co-jumps. Our JWTSCV estimator provides significantly lower bias to the BC estimator. The results show that it has by far the lowest bias and standard error.

We can therefore conclude that the JWTSCV estimator proves to be the most efficient estimator, as it is able to consistently estimate the individual jumps, co-jumps and noise from the processes. It is able to provide the true covariation of the processes, and also the

Table 9.1: Σ_t^{12} Bias (variance in parenthesis) $\times 10^4$ of all estimators from 10,000 simulations of jump-diffusion model with $\epsilon_1 = 0$, $\epsilon_2 = 0.0015$, zero and one common jump as well as zero and one independent jump. RC – realized covariance estimator, BC – bipower covariance estimator, TSCV – two-scale realized covariance, MRK – multivariate realized kernel, JWTSCV – jump wavelet two-scale realized covariance with different sampling of 1 min, 5 min, 30 min and 1 hour.

			RC	**BC**	**TSCV**	**MRK**	**JWTSCV**
			Zero Noise (ϵ_1)				
Zero IJ	Zero CJ	1-min	-0.001 (0.015)	-0.002 (0.017)	-0.005 (0.013)	-0.006 (0.042)	-0.005 (0.013)
		5-min	0.001 (0.035)	-0.002 (0.040)	-0.002 (0.029)	-0.008 (0.069)	-0.002 (0.029)
		30-min	-0.001 (0.085)	-0.015 (0.090)	-0.015 (0.067)	-0.040 (0.112)	-0.015 (0.067)
		1-hour	0.002 (0.124)	-0.032 (0.124)	-0.030 (0.091)	-0.080 (0.129)	-0.030 (0.091)
	One CJ	1-min	0.990 (1.786)	0.047 (0.089)	0.969 (1.755)	0.982 (1.805)	-0.004 (0.012)
		5-min	0.988 (1.811)	0.107 (0.245)	0.962 (1.772)	0.960 (1.834)	-0.005 (0.029)
		30-min	1.019 (2.041)	0.241 (0.577)	0.895 (1.705)	0.743 (1.617)	-0.018 (0.065)
		1-hour	1.001 (1.925)	0.272 (0.745)	0.753 (1.564)	0.444 (1.335)	-0.036 (0.090)
One IJ	Zero CJ	1-min	-0.003 (0.042)	0.035 (0.042)	-0.006 (0.036)	-0.000 (0.155)	-0.004 (0.012)
		5-min	-0.006 (0.115)	0.063 (0.093)	-0.008 (0.090)	-0.014 (0.218)	-0.005 (0.028)
		30-min	0.012 (0.326)	0.097 (0.209)	-0.007 (0.266)	-0.021 (0.467)	-0.014 (0.066)
		1-hour	-0.008 (0.568)	0.069 (0.341)	-0.038 (0.384)	-0.096 (0.547)	-0.035 (0.090)
	One CJ	1-min	0.926 (1.624)	0.084 (0.107)	0.907 (1.593)	0.917 (1.632)	-0.005 (0.012)
		5-min	1.002 (1.795)	0.197 (0.343)	0.988 (1.781)	0.968 (1.860)	-0.005 (0.028)
		30-min	1.012 (1.892)	0.417 (0.758)	0.910 (1.768)	0.771 (1.800)	-0.018 (0.069)
		1-hour	1.013 (2.097)	0.493 (1.113)	0.797 (1.730)	0.469 (1.586)	-0.038 (0.091)
			Noise (ϵ_2)				
Zero IJ	Zero CJ	1-min	0.000 (0.015)	-0.000 (0.017)	-0.004 (0.013)	-0.002 (0.045)	-0.004 (0.013)
		5-min	-0.002 (0.035)	-0.004 (0.040)	-0.005 (0.028)	-0.009 (0.069)	-0.005 (0.028)
		30-min	0.004 (0.091)	-0.016 (0.095)	-0.015 (0.071)	-0.036 (0.130)	-0.015 (0.071)
		1-hour	-0.000 (0.124)	-0.036 (0.125)	-0.036 (0.087)	-0.086 (0.123)	-0.036 (0.087)
	One CJ	1-min	1.016 (1.745)	0.047 (0.068)	0.993 (1.710)	0.999 (1.739)	-0.005 (0.013)
		5-min	0.882 (1.597)	0.099 (0.252)	0.866 (1.605)	0.874 (1.691)	-0.004 (0.028)
		30-min	1.024 (1.850)	0.261 (0.632)	0.948 (1.774)	0.831 (1.838)	-0.018 (0.062)
		1-hour	0.982 (1.834)	0.292 (0.719)	0.789 (1.615)	0.490 (1.371)	-0.035 (0.093)
One IJ	Zero CJ	1-min	0.001 (0.049)	0.037 (0.045)	-0.003 (0.039)	-0.001 (0.196)	-0.004 (0.012)
		5-min	0.007 (0.094)	0.068 (0.099)	-0.001 (0.084)	-0.014 (0.248)	-0.005 (0.029)
		30-min	0.015 (0.362)	0.097 (0.211)	0.002 (0.268)	-0.030 (0.523)	-0.018 (0.066)
		1-hour	0.017 (0.536)	0.072 (0.307)	-0.028 (0.370)	-0.074 (0.449)	-0.033 (0.092)
	One CJ	1-min	0.832 (1.443)	0.076 (0.084)	0.815 (1.418)	0.831 (1.472)	-0.005 (0.012)
		5-min	1.042 (1.818)	0.228 (0.473)	1.031 (1.806)	1.015 (1.995)	-0.004 (0.029)
		30-min	0.977 (1.865)	0.448 (0.782)	0.886 (1.678)	0.763 (1.704)	-0.018 (0.067)
		1-hour	0.993 (1.957)	0.501 (1.080)	0.812 (1.698)	0.515 (1.617)	-0.037 (0.088)

correlation. This makes the JWTSCV a very powerful tool for modeling the covariance and correlation of assets. With the theoretical results in hand, we can move to empirical examples and use the JWTSCV to study empirical data sets.

Table 9.2: Correlation bias (variance in parenthesis) of all estimators from 10,000 simulations of jump-diffusion model with $\epsilon_1 = 0$, $\epsilon_2 = 0.0015$, zero and one common jump as well as zero and one independent jump. RC – realized covariance estimator, BC – bipower covariance estimator, TSCV – two-scale realized covariance, MRK – multivariate kealized kernel, JWTSCV – jump wavelet two-scale realized covariance with different sampling of 1 min, 5 min, 30 min and 1 hour.

			RC	**BC**	**TSCV**	**MRK**	**JWTSCV**
			Zero Noise (ϵ_1)				
Zero IJ	Zero CJ	1-min	-0.001 (0.009)	-0.001 (0.013)	-0.001 (0.007)	-0.003 (0.027)	-0.001 (0.007)
		5-min	-0.001 (0.020)	-0.001 (0.029)	-0.000 (0.016)	-0.005 (0.045)	-0.000 (0.016)
		30-min	-0.007 (0.056)	-0.004 (0.091)	-0.003 (0.042)	-0.019 (0.111)	-0.003 (0.042)
		1-hour	-0.013 (0.095)	-0.011 (0.164)	-0.005 (0.069)	-0.030 (0.207)	-0.005 (0.069)
	One CJ	1-min	0.037 (0.036)	0.010 (0.020)	0.037 (0.036)	0.036 (0.041)	0.000 (0.007)
		5-min	0.037 (0.039)	0.014 (0.037)	0.037 (0.038)	0.035 (0.051)	-0.001 (0.017)
		30-min	0.037 (0.052)	0.029 (0.076)	0.037 (0.048)	0.031 (0.081)	-0.005 (0.044)
		1-hour	0.027 (0.087)	0.017 (0.157)	0.029 (0.072)	-0.007 (0.213)	-0.013 (0.078)
One IJ	Zero CJ	1-min	-0.492 (0.283)	-0.021 (0.055)	-0.492 (0.282)	-0.494 (0.301)	-0.001 (0.007)
		5-min	-0.464 (0.306)	-0.035 (0.092)	-0.460 (0.300)	-0.466 (0.342)	-0.001 (0.016)
		30-min	-0.483 (0.364)	-0.102 (0.276)	-0.473 (0.342)	-0.488 (0.461)	-0.002 (0.043)
		1-hour	-0.504 (0.456)	-0.194 (0.450)	-0.476 (0.398)	-0.512 (0.620)	-0.012 (0.078)
	One CJ	1-min	-0.331 (0.301)	-0.007 (0.055)	-0.330 (0.300)	-0.333 (0.312)	-0.001 (0.007)
		5-min	-0.343 (0.313)	-0.021 (0.136)	-0.342 (0.312)	-0.355 (0.351)	-0.002 (0.016)
		30-min	-0.340 (0.355)	-0.060 (0.258)	-0.332 (0.337)	-0.346 (0.435)	-0.007 (0.047)
		1-hour	-0.368 (0.434)	-0.111 (0.430)	-0.347 (0.391)	-0.398 (0.593)	-0.012 (0.076)
			Noise (ϵ_2)				
Zero IJ	Zero CJ	1-min	-0.000 (0.009)	-0.000 (0.013)	0.000 (0.007)	-0.002 (0.028)	0.000 (0.007)
		5-min	-0.002 (0.020)	-0.001 (0.030)	-0.001 (0.016)	-0.005 (0.046)	-0.001 (0.016)
		30-min	-0.002 (0.052)	-0.005 (0.086)	-0.002 (0.043)	-0.014 (0.107)	-0.002 (0.043)
		1-hour	-0.011 (0.093)	-0.011 (0.171)	-0.006 (0.067)	-0.048 (0.249)	-0.006 (0.067)
	One CJ	1-min	0.039 (0.037)	0.010 (0.019)	0.039 (0.037)	0.038 (0.041)	-0.000 (0.007)
		5-min	0.037 (0.039)	0.015 (0.034)	0.037 (0.037)	0.035 (0.049)	-0.001 (0.017)
		30-min	0.036 (0.054)	0.021 (0.082)	0.036 (0.049)	0.027 (0.091)	-0.005 (0.043)
		1-hour	0.028 (0.092)	0.018 (0.153)	0.033 (0.066)	0.010 (0.187)	-0.012 (0.078)
One IJ	Zero CJ	1-min	-0.477 (0.287)	-0.018 (0.058)	-0.477 (0.286)	-0.479 (0.310)	-0.000 (0.007)
		5-min	-0.470 (0.302)	-0.042 (0.106)	-0.469 (0.298)	-0.481 (0.343)	-0.002 (0.016)
		30-min	-0.489 (0.355)	-0.112 (0.264)	-0.476 (0.328)	-0.485 (0.445)	-0.006 (0.045)
		1-hour	-0.482 (0.428)	-0.154 (0.463)	-0.467 (0.392)	-0.483 (0.594)	-0.012 (0.079)
	One CJ	1-min	-0.344 (0.293)	-0.011 (0.047)	-0.344 (0.293)	-0.346 (0.306)	-0.001 (0.007)
		5-min	-0.347 (0.311)	-0.014 (0.141)	-0.345 (0.305)	-0.358 (0.349)	-0.002 (0.016)
		30-min	-0.345 (0.359)	-0.051 (0.269)	-0.337 (0.345)	-0.347 (0.441)	-0.006 (0.044)
		1-hour	-0.360 (0.417)	-0.133 (0.415)	-0.345 (0.381)	-0.368 (0.567)	-0.012 (0.077)

Chapter 10

Decomposition of empirical multivariate volatility

In this chapter, we apply our theoretical results to real-world data. The JWTSCV proved to be the most efficient estimator of integrated covariance and correlation, thus we also expect it to improve our understanding of the true relationship between assets. In addition, we use the power of wavelets to decompose the realized measures into several investment horizons.

10.1 Data description

Foreign exchange future contracts are traded on the Chicago Mercantile Exchange (CME) on a 24-hour basis. These markets are among the most liquid, so they are suitable for testing our estimator. We will estimate the realized covariance of British pound (GBP), Swiss franc (CHF) and euro futures (EUR), while we will focus on the GBP-CHF, GBP-EUR and CHF-EUR futures pairs. After estimating the covariance, we will study the correlations between the currencies. All contracts are quoted in the unit value of the foreign currency in US dollars, which makes them comparable. The cleaned data are available from Tick Data, Inc. [1]

It is important to understand the trading system before we begin the study. In August 2003, CME launched the Globex trading platform, which generated a large increase in the liquidity of currency futures. For the first time ever in a single month, the trading volume on the electronic trading platform exceeded 1 million contracts every day. On Monday, December 18, 2006, the CME Globex(R) electronic platform started offering 23-hours-a-day trading. The weekly trading cycle begins at 5:00 pm on Sunday and ends at 4:00 pm on Friday, while every day the trading is interrupted for one hour from 4:00 pm until 5:00 pm. These changes in the trading system had a dramatic impact on trading activity. For this reason, we restrict ourselves to a sample period extending from

[1] http://www.tickdata.com/

January 5, 2007 through November 17, 2010, which contains the most recent financial crisis. The futures contracts we use are automatically rolled over to provide continuous price records, so we do not have to deal with different maturities.

The tick-by-tick transactions are recorded in Chicago Time, referred to as Central Standard Time (CST). Therefore, in a given day, trading activity starts at 5:00 pm CST in Asia, continues in Europe followed by North America, and finally closes at 4:00 pm in Australia. We exclude potential jumps due to the one hour gap in trading from our analysis by redefining the day in accordance with the electronic trading system. Moreover, we eliminate Saturdays and Sundays, US federal holidays, December 24 to December 26, and December 31 to January 2 because of the very low activity on these days, which would bias the estimates. Finally, we are left with 944 days in the sample.

For the analysis of relations between the currencies, it is crucial that they are synchronized in time. As discussed in Section 8.3, we use the refresh time scheme to synchronize the data. Looking more closely at the higher frequencies, we find that a large amount of transactions have a common time stamp, so we use the arithmetic average for all observations with the same time stamp. Finally, we redefine the clock according to the refresh time scheme so that we can work with the data that are synchronized. We use the refresh time scheme for each pair separately in order to keep as much data as possible in the analysis.

10.2 Multivariate unconditional volatility distributions

Having prepared the data, we can use the estimators to study the dependencies. For each pair, we estimate the covariance using the realized covariance estimator defined by 8.7, the bipower realized covariance defined by 8.21, the two-scale realized covariance defined by 8.15, the multivariate realized kernel defined by 8.16 and our jump wavelet two-scale realized covariance estimator defined by 9.14. For convenience, we refer to the estimators in the description of the results as RC, BC, TSCV, MRK and JWTSCV, respectively, while the l-th and q-th processes will be described in words, i.e., we compute the RC for the EUR and GBP futures pair. The RC and the BC are estimated on 1-min and 5-min log-returns for comparison. The TSCV and the JWTSCV are estimated on the slow time scale of 5 minutes.

We extend the results from the study of univariate distributions contained in Chapter 5 by studying the relationships that may exist among the different measures of variation and covariation. One of the key questions in finance is whether the individual volatilities move together and whether we can find any positive relationship between them. The realized measures provide us with a relatively simple way of addressing these questions. Our wavelet-based realized measures further allow us to decompose the relationship, so we hope they will also help us to move forward our understanding of the dependence.

Table 10.1 provides the average estimated covariation and correlation among the three currencies. As the benchmark, we use unconditional open-to-close measures computed as

Table 10.1: The table reports the average covariation ($\times 10^4$) and correlation among the three currencies.

	Covariance			Correlation		
	GBP-CHF	GBP-EUR	CHF-EUR	GBP-CHF	GBP-EUR	CHF-EUR
RC 1-min	0.299	0.371	0.429	0.419	0.538	0.652
RC 5-min	0.305	0.384	0.434	0.472	0.605	0.738
BC 1-min	0.273	0.340	0.388	0.428	0.545	0.660
BC 5-min	0.285	0.358	0.398	0.478	0.611	0.738
TSCV 5-min	0.274	0.351	0.390	0.528	0.655	0.808
MRK 1-min	0.301	0.384	0.429	0.529	0.654	0.809
MRK 5-min	0.292	0.374	0.424	0.525	0.650	0.811
JWTSCV 5-min	0.249	0.322	0.346	0.506	0.629	0.770
Open-to-Close	0.245	0.325	0.4217	0.458	0.623	0.787

the outer products of the open-to-close returns. Interestingly, the unconditional measures are not far from the realized measures. This seems to be a feature of currency data, as other authors, e.g. Barndorff-Nielsen et al. (2011), have found significant differences on large samples of US stocks.

All the correlations are positive. The average relationship between the studied currencies is strong, pointing to a strong degree of integration among these European countries. Our findings are consistent with those of Aït-Sahalia et al. (2010), who use the same data set as we do, with the only difference that their data sample ends in June 2009.

When comparing the different, we observe the Epps effect. The RC and BC estimators show lower correlation on average when data with higher frequencies are used. Interestingly, this finding is not confirmed by the MRK estimator. The BC shows slightly stronger correlation than the RC, suggesting that some degree of co-jumping is present in the data. While the correlations are generally strong, it seems that co-jumps do not have such an impact on the currency data. Interestingly, the TSCV and MRK estimators, which are robust to noise, estimate the strongest correlations. When compared with the JWTSCV estimating only the dependence of the true, continuous part without jumps, it estimates the correlation to be a little lower. Thus the JWTSCV seems to be much more efficient in the estimation of jumps than the BC, as it is not biased with noise. Economically, these differences may lead to improved results in portfolio theory. We will study this impact in more depth in the last part of this thesis. All the estimated realized covariance series, moreover, show very similar behavior to the realized variances studied in Part I of this dissertation. In particular, all of them are highly persistent. We will build on these findings in the following sections.

Before we do so, let us look at the decomposed dependencies using wavelets. We decompose the covariance and correlation measures into four scales, corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes and 40-80 minutes, and the rest (80 minutes up to 1 day). We remind the reader that the sum of these components will always add to the estimator.

More precisely, the components of the JWTSCV corresponding to the various invest-

Table 10.2: The table reports the correlation of the daily logarithmic realized volatilities estimated by the RV, BV, TSRV, RK and JWTSRV for the GBP-CHF, GBP-EUR and CHF-EUR futures pairs. The table also reports the correlation between the decomposed volatility $JWTSRV_j$ on the different investment horizons $j = 1, \ldots, 5$, representing 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day).

						$JWTSRV_j$				
	RV	BV	TSRV	RK	JWTSRV	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
GBP-CHF	0.827	0.841	0.827	0.766	0.858	0.865	0.832	0.779	0.749	0.659
GBP-EUR	0.923	0.930	0.922	0.894	0.941	0.940	0.925	0.902	0.868	0.836
CHF-EUR	0.928	0.931	0.931	0.898	0.939	0.937	0.921	0.891	0.845	0.815

ment horizons will be referred to as $JWTSCV_{(l,q)j}$:

$$JWTSCV_{(l,q)} = \sum_{j=1}^{5} JWTSCV_{(l,q)j}, \qquad j = 1, \ldots, 5. \tag{10.1}$$

Similarly, the components of the correlations computed using the JWTSCV and JWTSRV estimators will be transformed to the correlations according to Definition 25, and we will assign them the name $WRCorr$ in the empirical analysis:

$$WRCorr_{(l,q)} = \sum_{j=1}^{5} \eta_j \widehat{\rho}^{(JWR)}_{(l,q)j}. \tag{10.2}$$

For ease of notation, we drop q and l and will refer to the estimators as JWTSCV and WRCorr, always noting the currencies between which the dependence is studied.

In addition to the results from Table 10.1, which suggest that the currencies move together, we would like to carry out a more rigorous analysis of the dependence. We follow Andersen et al. (2001), who first introduce this kind of analysis to study multivariate unconditional volatility distributions. Moreover, it will be interesting to see what the dependence on various investment horizons is.

Figure 10.1 contains scatter plots of the daily logarithmic realized volatilities of the pairs. For example, in column (a) of Figure 10.1, we have scatter plots of the daily logarithmic realized volatility of GBP futures against the logarithmic realized volatility of CHF futures. Moreover, we analyze the volatilities estimated using different estimators – the RV, the BV, the TSRV, the RK and our JWTSRV. The other columns of this Figure show the same relationship for the remaining currency pairs. It is evident that all the pairs of volatilities have a strong positive relation and they move together. Table 10.2 summarizes the correlations between the volatilities using all the estimators. It is interesting that the correlations between the volatilities using the RV, TSRV and RK estimators, which include jump variation, are lower than the correlations of the realized volatilities using the BV and the JWTSRV. Moreover, the relationship between the two volatilities is strongest when one uses our JWTSRV estimator for all pairs.

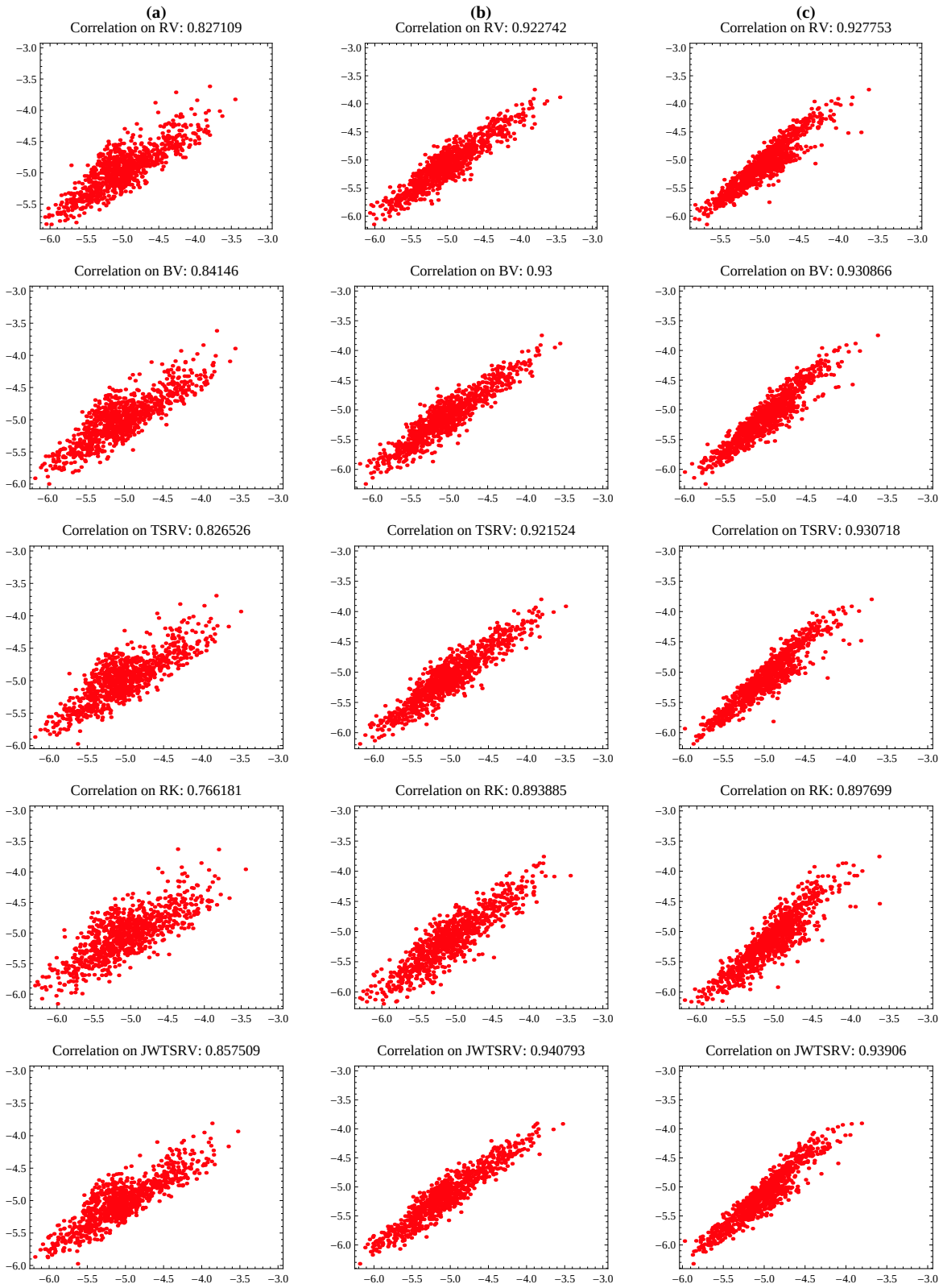


Figure 10.1: Scatter plots of l-th currency volatility on q-th currency volatility estimated by various estimators: RV, BV, TSRV, RK and JWTSRV for (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.

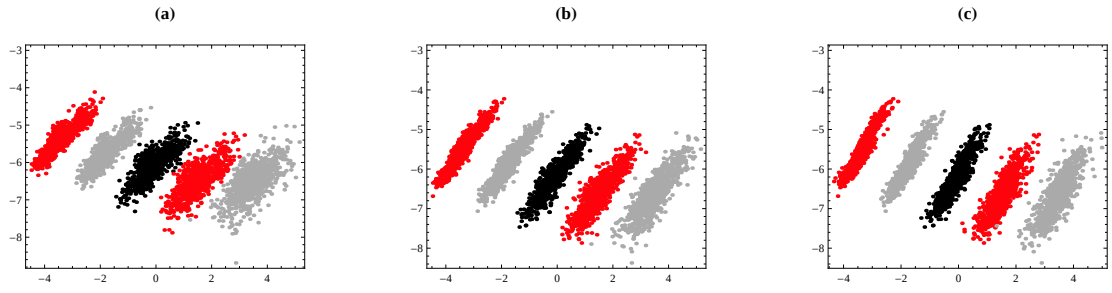


Figure 10.2: Scatter plots of components of l-th currency volatility on components of q-th currency volatility estimated by $JWTSRV_j$ on the different investment horizons $j = 1, \ldots, 5$, representing 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day). Note that x and y axis are not (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.

This result indicates that there is a significant number of independent jumps in the univariate series, which biases the true dependence in the bivariate setting. Thus, the JWTSRV estimator, which is robust to noise and jumps, can recover the true dependence of the volatilities. The difference is economically significant, and it may have an impact on portfolio valuation. We will test this assumption later in the analysis.

We also repeat the analysis on the decomposed volatility $JWTSRV_j$ on the different investment horizons $j = 1, \ldots, 5$, representing 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day). Figure 10.2 shows the scatter plots of the decomposition of the daily logarithmic realized volatilities of the pairs using $JWTSRV_j$. The scatter plots are shifted in the figure, so the dependence is visible and for this reason the x-axis cannot be interpreted. In addition, Table 10.2 summarizes the correlation coefficients between the volatility components. The results suggest that the dependence is stronger with higher frequency used. This is logical, as the higher frequencies make a larger contribution. For example, the first, 5–10 minute frequency ($j = 1$) makes a contribution of around 50% on average (for a more detailed discussion, see Section 5.3; volatility contribution plots can be found in Figure 5.2). Still, the result is economically interesting, as it suggests that most of the co-volatility comes from the higher frequencies on the currency markets.

Next, Figure 10.3 presents the volatility-in-correlation effect. We plot the average realized daily correlations for the q-th currency, $1/m \sum_i WRCorr_{(q,i)t}$ for $i \neq q$, against the logarithmic standard deviation for the l-th. These findings are also in line with Andersen et al. (2001), showing a positive correlation in the first row. The second row of Figure 10.3 again shows the dependence of the decomposed volatility-in-correlation effect. We can see that the effect is lower in the higher frequencies and vice versa. The dependence suggests that standard mean-variance efficiency calculations based on constant correlations are misguided. It is much more interesting to look at Figure 10.4, which contains scatter plots of the average realized daily correlations of the q-th and l-th currencies, i.e., $1/m \sum_i WRCorr_{(q,i)t}$ for $i \neq q$ and $1/m \sum_l WRCorr_{(l,i)t}$ for $i \neq l$.

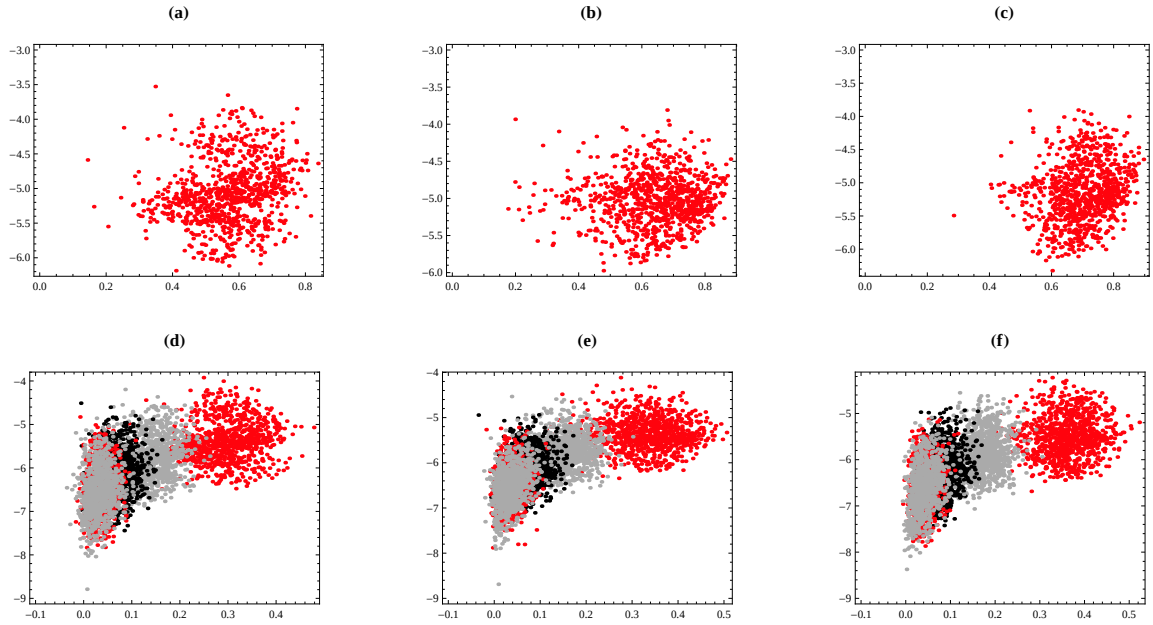


Figure 10.3: Scatter plots of average realized daily correlations for q-th currency, $1/n \sum_i WRCorr_{(q,i)t}$ for $i \neq q$, against logarithmic standard deviation for l-th currency estimated by JWTSRV. (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair subfigures correspond to total measures, while (d) GBP-CHF futures pair, (e) GBP-EUR futures pair and (f) CHF-EUR futures pair subfigures are estimators decomposed to different investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day).

Clearly, there is a strong similarity in the comovement across the individual currencies. Andersen et al. (2001) suggest that this behavior might be credited to a low-dimensional factor structure driving the second moment characteristics of the joint distribution. The second row of Figure 10.4 confirms similar behavior on the decomposed correlations.

The higher correlations are more related than the lower correlations. The second row of Figure 10.4 shows nicely that the dependencies are greatest on the higher frequencies. Our results thus generalize the findings of Andersen et al. (2001), who find a strong positive relationship between volatilities on broad market data with analysis of intraday dependencies on higher frequencies. Although the approach is nonparametric, thus making it difficult to use for predicting future correlations, it tells us a lot about the dependence. A wrong assumption about the dependence process will have a large impact on the portfolio valuation. In the next sections, we will look more closely at the covariance and correlation estimates and we will use these findings to develop a forecasting model. In the last part, we will apply them in asset pricing theory.

But before we do so, we would like to formally wrap up the interesting empirical findings presented in this section by proposing a simple low-dimensional latent factor structure in volatility which explains the commonalities found.

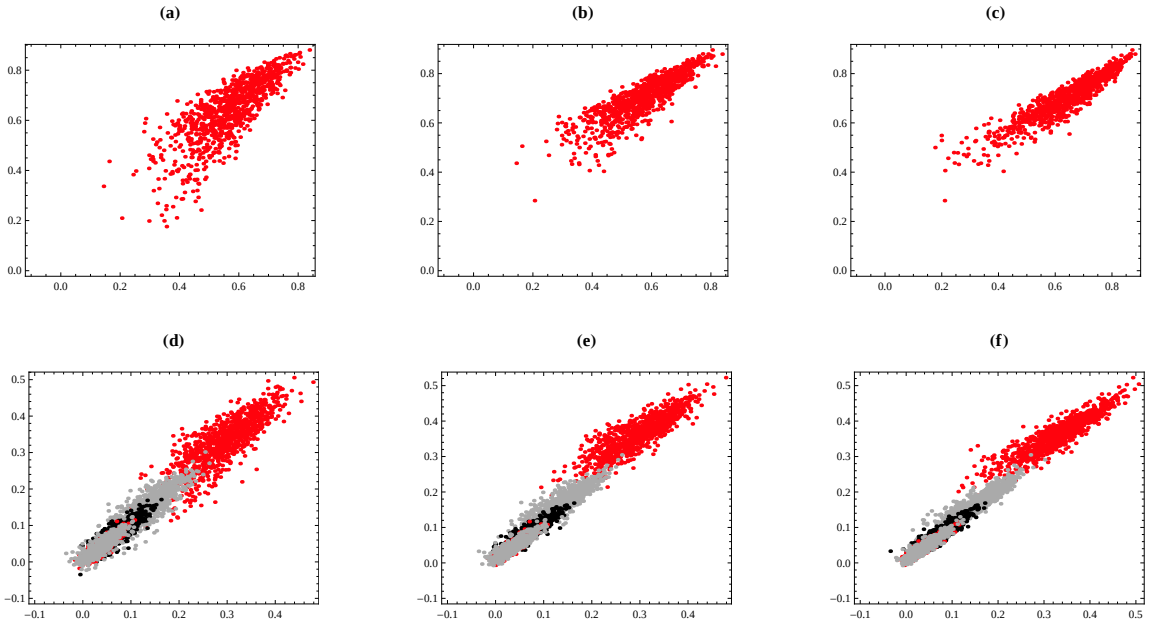


Figure 10.4: Scatter plots of average realized daily correlations of l-th and q-th currencies. (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair subfigures correspond to total measures, while (d) GBP-CHF futures pair, (e) GBP-EUR futures pair and (f) CHF-EUR futures pair subfigures are estimators decomposed to different investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day).

10.3 Latent factor structure in volatility

Following Andersen et al. (2001), we briefly focus on the low-dimensional factor structure to complete the previous analysis. Based on the observation of the tendency of volatilities to move together, the tendency of correlations to be higher when the corresponding volatilities are high, and the tendency for an arbitrary correlation to be high when other correlations are also high, the latent factor model is explored to model these phenomena.

Consider an m-dimensional diffusion process for log price $\mathbf{p_{t,h}}$ with a single-factor representation, $\mathbf{p_{t,h}} = (p_{(1)t,h}, \ldots, p_{(m)t,h})$, evolving continuously over the time interval $[t-h, t]$, $0 \leq h \leq t \leq T$. Each element of the returns dp_t is driven partially by a single latent factor with stochastic volatility and partially by an orthogonal idiosyncratic noise.

$$d\mathbf{p_t} = \lambda \sigma_t dW_t + \Theta dV_t, \tag{10.3}$$

where λ is an m-dimensional vector of loadings on the volatility factor σ_t, dW_t and dV_t are m-dimensional standard Brownian motions with mutually independent elements, and the diagonal matrix Θ contains m individual volatilities.

The returns $\mathbf{r_{t,h}}$, evolving continuously over the time interval $[t-h, t]$, $0 \leq h \leq t \leq T$, are then driven by

$$\mathbf{r_{t,h}} = \int_{t-h}^{t} \lambda \sigma_s dW_s + \int_{t-h}^{t} \Theta dV_s. \tag{10.4}$$

The quadratic covariation process associated with the returns $\mathbf{r_{t,h}}$ between the l-th and q-th price process over $[t-h, t]$, for $0 \leq h \leq t \leq T$, are then

$$CV_{(l,q)t,h} = \lambda\lambda' \int_{t-h}^{t} \sigma^2_{1+s} ds + \Theta_{(l,q)}. \tag{10.5}$$

So, the conditional variances and covariances inherit the dynamics from σ_t.

The continuous time latent volatility model can be directly approximated by the discrete time model

$$r_{(q)t} = \lambda_{(q)} f_{(q)} + \nu_{(q),t}, \tag{10.6}$$

where $f_t|I_t \sim (0, h_t)$, $\nu_{(q)t}$ is an *iid* process with zero mean and variance $\epsilon^2_{(q)}$, and $cov(\nu_{(q)t}\nu_{(l)t'}) = 0$ for all $q \neq l$, and $t \neq t'$, where $q, l = 1, \ldots, m$ and time $t = 1, \ldots, T$.

The volatilities of this factor model move together. The l-th and q-th time conditional variances of the l-th and q-th price process are themselves stationary stochastic processes driven entirely by movements in volatility

$$\begin{aligned} h_{(l)t} &= \lambda^2_{(l)} h_t + \epsilon^2_{(l)} \\ h_{(q)t} &= \lambda^2_{(q)} h_t + \epsilon^2_{(q)}. \end{aligned} \tag{10.7}$$

The unconditional covariance between $h_{(l)t}$ and $h_{(q)t}$ is $cov(h_{(l)t}, h_{(q)t}) = \lambda^2_{(l)}\lambda^2_{(q)}(h_t - E(h_t))^2$, which is always positive.

Finally, representation of the conditional correlation for this model will help us to see why a factor structure leads directly to high correlations when there is high volatility. The (l, q)-th conditional correlation can be computed as

$$Corr_{(l,q)} = \frac{\lambda_{(l)}\lambda_{(q)} h_t}{\sqrt{\lambda^2_{(l)} h_t + \epsilon^2_{(l)}}\sqrt{\lambda^2_{(q)} h_t + \epsilon^2_{(q)}}}. \tag{10.8}$$

Finally, it is also easy to see that the factor structure leads directly to high correlations between two arbitrary chosen stocks when the correlation between other stocks is also high. In fact, if all stocks load positively off the common factor, all correlations will be increasing in volatility. These findings hold not only for the price process, but also for the decomposed price process. In this way, we can simply develop a latent factor structure model for decomposed volatilities as well, with similar implications.

Imposing a simple latent factor structure on volatility can explain all the empirical findings from this section. Of course it is straightforward to extend these results to richer factor structures, including models with dynamics in Θ, or models with multiple factors. These are beyond the scope of this work, as we simply used the latent factor structure to explain our findings. Let's turn to studying the dynamics of the dependence between the currencies.

10.4 Dynamics of decomposed dependencies

The previous section provided us with a basic statistical overview of the dependence between the currencies. While looking at the averages, we did not show the considerable variation of all the measures. Such variation points to interesting dynamics, which we uncover here. In addition, we take advantage of wavelet theory and study the dynamics of the decomposed measures as well. More specifically, we decompose the covariance and correlation measures into four scales corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20-40 minutes and 40–80 minutes, and the rest (80 minutes up to 1 day). Finally, we use wavelet theory to disentangle co-jumps and individual jumps from the series.

Figure 10.6 provides us with the decomposition of the estimated covariance for all the currency pairs. The first row provides the bivariate time series plots and the second row the covariance estimated by our JWTSCV estimator. The third row presents the decomposition of the covariance into the various investment horizons, while the last three rows give estimates of the co-jumps and individual jump variations of both series.

As in the univariate case, the most of the covariance comes from the 5–10 minute frequency, which accounts for about 50% of the total covariance, and the 10–20 minute frequency, which accounts for about 25% of the total, which is strikingly similar to the univariate case. The full picture of the contributions for all pairs can be seen in Figure 10.5

Our method of estimation also allows us to study jumps and co-jumps in the currencies. Interestingly, the jump variation is much stronger than the co-jump variation in the studied currencies. Still, the co-jumps are significant and should not be ignored in any further analysis. These results suggest that if the jumps are ignored, the covariation will be downward biased, as we saw in the previous analysis (Table 10.1). We will return to

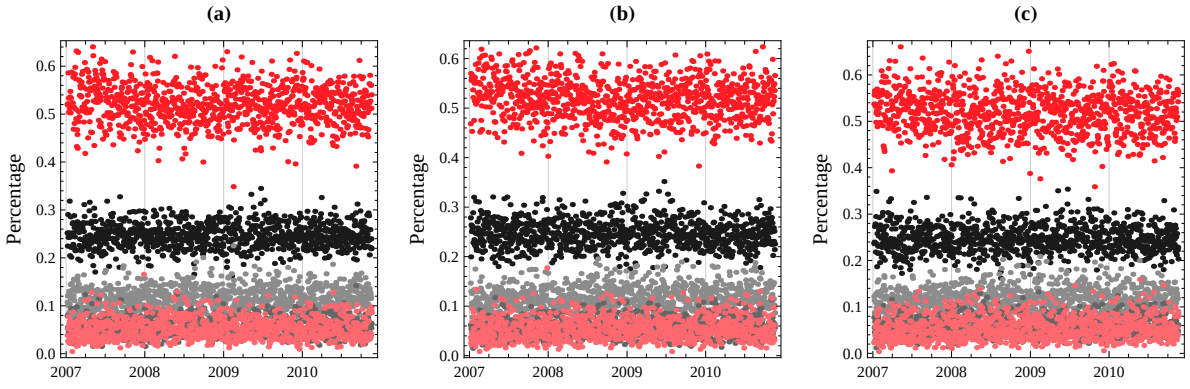


Figure 10.5: $JWTSCV_j$, $j = 1, \ldots, 5$, contributions of components of integrated covariation CV_t corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day. (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF - EUR futures pair.

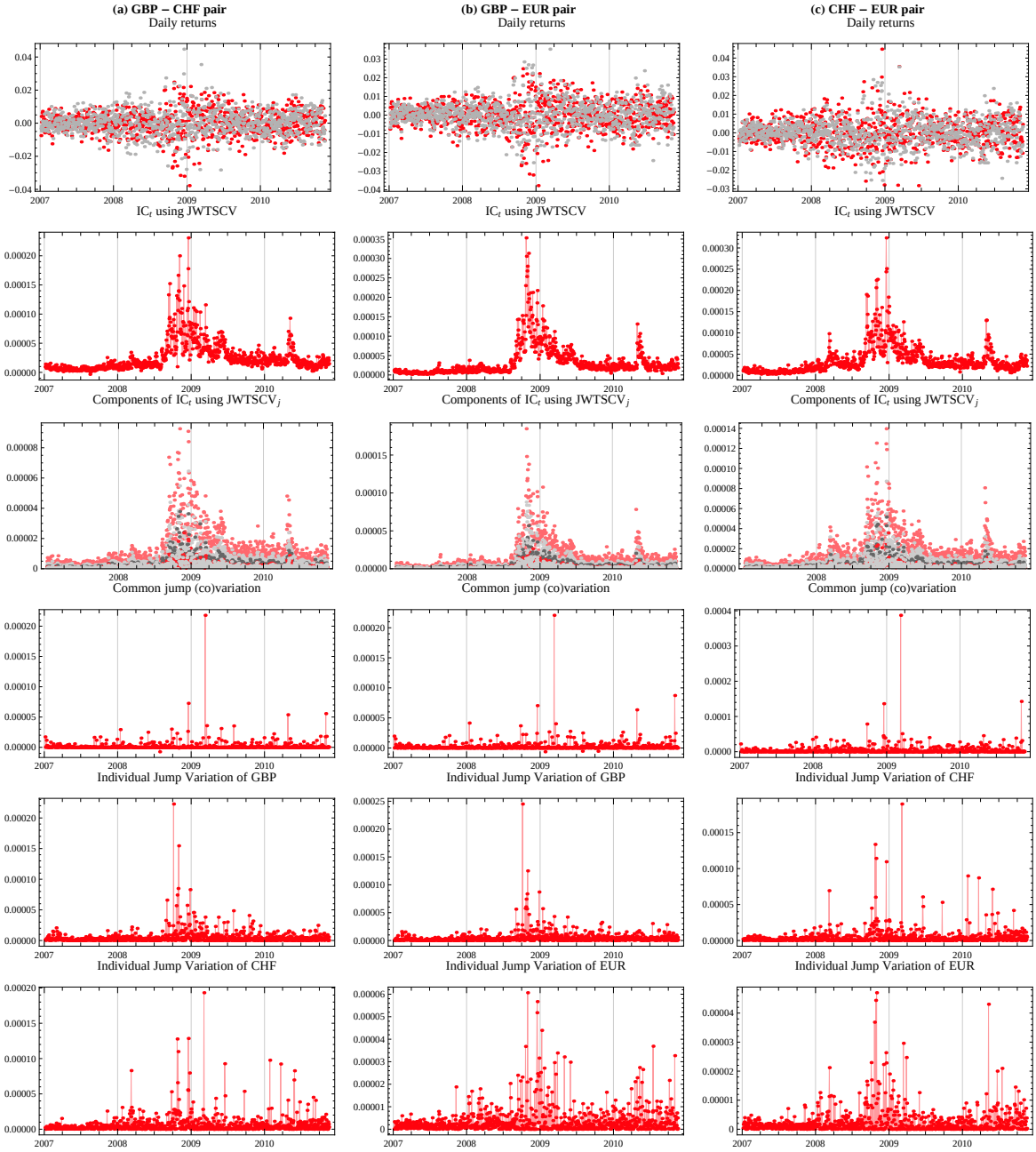


Figure 10.6: Daily returns, CV_t estimated by JWTSCV, decomposition of CV_t using $JWTSCV_j$ for $j = 1, \ldots, 5$ corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day, JWTSCV estimated common jump variation, individual jump variations of both time series. (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.

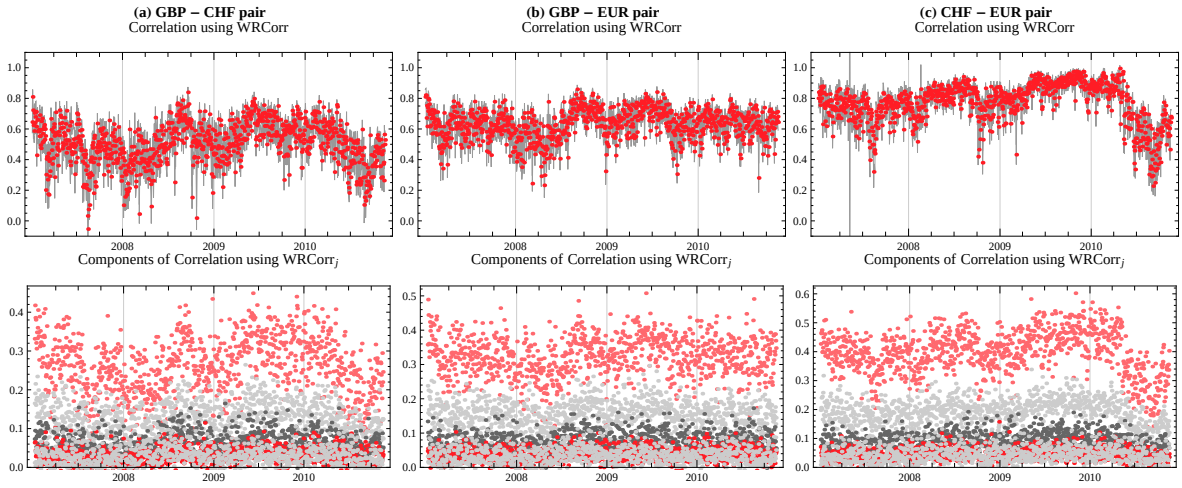


Figure 10.7: Correlations with 95% confidence interval and decomposition of correlations using $WRCorr_j$ for $j = 1, \ldots, 5$ corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day. (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.

the impact of jumps later in the text.

Having computed the variances and covariances, we can take a look at the correlation dynamics. Figure 10.7 presents the estimate of WRCorr complete with 95% confidence intervals, as well as its decomposition. We can see that the correlation of all the pairs vary substantially. While during 2007, the correlation of all three currencies was decreasing, it increased during 2008. At the end of 2008, during the largest stock market falls, which lasted approximately two weeks, the dependence in the currencies weakened. This finding is interesting, as the correlations are expected to grow during large drops. While the currencies show a strong degree of common dependence with the European Union, it seems that the recent financial crisis did not affect the dependence, while it of course substantially increased the variation of all series. Interestingly, the correlation of CHF with both GBP and EUR weakened substantially during 2010.

The decomposition of the correlations again shows an interesting result. Most of the correlation comes from the highest scale of 5–10 minutes. For example, of the total 0.506 average correlation of the GBP-CHF pair, the correlation on the 5–10 minute horizon is 0.26, the correlation on the 10–20 minute horizon is 0.13, and the rest corresponds to 0.06, 0.03 and 0.03 (note that by simply summing these correlations we get the total correlation for the pair).

Figure 10.8 provides a comparison of the correlation dynamics computed using two estimators: the basic realized correlation and our jump-adjusted wavelet correlation (WRCorr) estimator. It is noticeable that our WRCorr estimator provides an estimate with lower variance (basically due to jumps) and confidence intervals.

To be precise, the GBP-CHF futures pair, the GBP-EUR futures pair and the CHF-EUR futures pair have average estimated WRCorr correlations of 0.506 ($\pm$0.069), 0.629

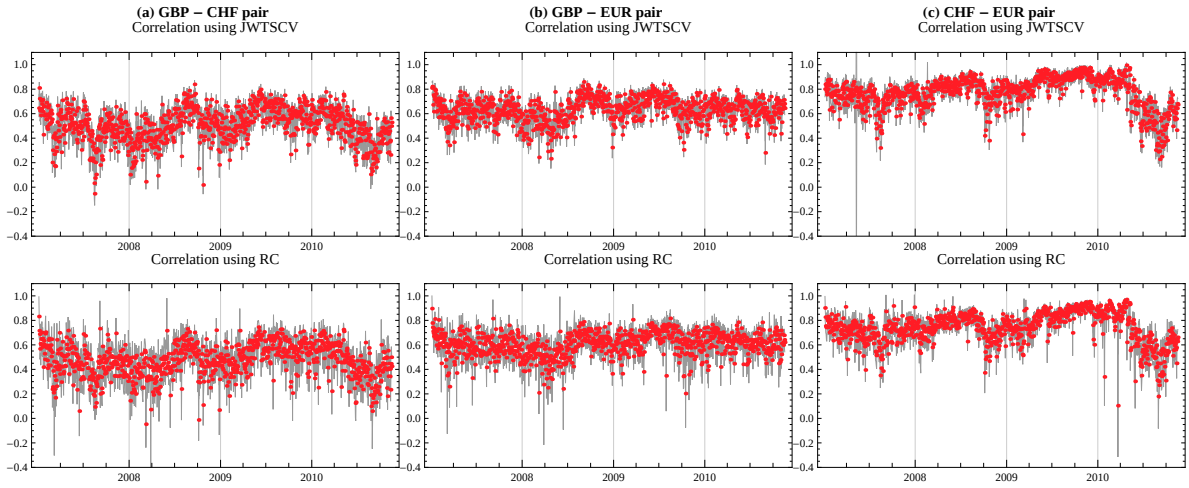


Figure 10.8: Comparison of correlations with 95% confidence interval using WRCorr in first row and using RC in second row. (a) GBP-CHF futures pair, (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.

(±0.053) and 0.769 (±0.051), respectively (95% confidence intervals in parentheses). The average correlations for the same pairs estimated using the standard RC method are 0.47 (±0.1), 0.602 (±0.086) and 0.738 (±0.062), respectively. Even though the correlations change significantly over time, the average correlation estimated using our method is approximately 0.03 larger than that using the simple RC. This result is economically significant and can have direct impact on portfolio diversification. Moreover, our method provides much narrower confidence intervals for the estimates.

10.4.1 Temporal dependence and long memory in correlations

While long memory in volatility has been widely discussed in the literature, the behavior of correlations is not so well founded. Maybe the reason is the considerable difficulty in estimating parametric multivariate volatility models. A striking feature of the previous analysis is the high degree of dependence in the dynamics, pointing to their predictability. We would like to briefly continue the analysis and look at the long memory feature of the correlations more closely.

Figures 11.4 and 11.5, available in the Appendix to Part II, show the autocorrelation functions of the decomposed covariance and correlations of all pairs. Let's start with the covariance. Similarly to the variance of the currencies, the covariance between the pairs shows a strong long memory at all investment horizons. This means that the construction of a forecasting model similar to the Heterogeneous Autoregressive (HAR)-type model used in the univariate part should help us to accurately forecast the covariance.

The autocorrelation function of the correlations does not provide the same story. Strong memory is a feature of the highest frequency, but disappears at lower frequencies. To be able to characterize the long memory of the processes, we compute the Hurst

Table 10.3: The table reports Hurst exponents, with standard deviations in parenthesis, of the covariance (JWTSCV), the correlations (WRCorr) and their components $JWTSCV_j$ and $WRCorr_j$ for $j = 1, \ldots, 5$, representing the different investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and the rest (80 minutes up to 1 day), for the GBP-CHF, GBP-EUR and CHF-EUR futures pairs.

	JWTSCV	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
GBP - CHF	0.937	0.946	0.928	0.850	0.811	0.781
	(0.003)	(0.001)	(0.005)	(0.018)	(0.022)	(0.019)
GBP - EUR	0.953	0.951	0.951	0.935	0.902	0.883
	(0.002)	(0.002)	(0.001)	(0.005)	(0.006)	(0.010)
CHF - EUR	0.933	0.938	0.928	0.886	0.847	0.831
	(0.001)	(0.002)	(0.002)	(0.011)	(0.018)	(0.013)
	WRCorr	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
GBP - CHF	0.878	0.858	0.805	0.748	0.654	0.623
	(0.009)	(0.016)	(0.016)	(0.011)	(0.029)	(0.021)
GBP - EUR	0.838	0.751	0.736	0.708	0.572	0.528
	(0.013)	(0.021)	(0.022)	(0.013)	(0.033)	(0.026)
CHF - EUR	0.943	0.901	0.860	0.755	0.623	0.583
	(0.005)	(0.015)	(0.019)	(0.019)	(0.026)	(0.014)

exponent.[2]

Table 10.3 shows the estimated Hurst exponents for the second moment ($q = 2$) with standard deviations obtained by estimates on different investment horizons from 5 to 19. Both the total covariance and the correlation show strong long memory (close around 0.9). All the components of the covariance estimates also show a strong long memory feature, which is declining for lower frequencies. This implies that the HAR-type model should be able to provide a good approximation of the temporal dependence in the covariation and thus it should serve as a reliable model for forecasting. Moreover, decomposition of the covariance series should also yield an improvement in its forecasting.

Interestingly, the correlation shows a different pattern. While its temporal dependence is a little lower than in the case of the covariation, it breaks at lower frequencies. Still, the long memory feature is strikingly strong. We remind the reader that all the results are based on our JWTSCV measure of realized variation, covariation and correlation, meaning that based on the previous analysis we are dealing with the dynamics after noise reduction as well as jump cleaning from the series. We also note that all the data scales, so it makes sense to estimate the Hurst exponent. Scaling plots are available from the authors on request.

Based on these results, we build a forecasting model in the next section.

[2]We use the generalized Hurst exponent method popularized by Tiziana Di Matteo. We do not explain all the details here, but leave the reader with the reference, i.e., Di Matteo et al. (2003).

10.5 Forecasting model based on decomposed integrated covariances

Motivated by the results from the previous analysis, we turn to building a forecasting model for covariances. Since the realized covariances show strikingly similar long memory behavior to the realized variances we make us of this feature to build an ARFIMA-type long memory model. Moreover, we decompose the covariance into several investment horizons and jumps, similarly to the univariate model.

If we assume that covariation series belong to the class of ARFIMA processes introduced into econometrics by Granger and Joyeux (1980) then the dth difference of each series is a stationary and invertible ARMA process where parameter d may be any real number such that $-1/2 < d < 1/2$ to ensure stationarity and invertibility. More precisely, $\sigma_{(1)t}\sigma_{(2)t}$ is an ARFIMA(p, d, q) process if it follows:

$$\alpha(L)(1-L)^d(\sigma_{(1)t}\sigma_{(2)t} - \mu) = \beta(L)v_t, \tag{10.9}$$

where $\alpha(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ and $\beta(z) = 1 + \beta_1 z + \cdots + \beta_q z^q$ are polynomials of order p and q, respectively, in the lag operator $L(L\sigma_{(1)t}\sigma_{(2)t} = \sigma_{(1)t-1}\sigma_{(2)t-1})$, which roots strictly outside the unit circle, v_t is iid with zero mean and σ_v^2 variance, and $(1-L)^d$ is defined by its binomial expansion

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j \tag{10.10}$$

using gamma function, $\Gamma(.)$.

The parameter d determines the memory of the process. For $d > 0$, the process is said to have long memory, since its autocorrelations die out at a hyperbolic rate and are no longer absolutely summable, in contrast to the much faster exponential rate in the weak dependence case of $d = 0$, where the process captures the behavior of the short-memory ARMA model.

Once we have estimated the ARFIMA(p, d, q) model with the Haslett and Raftery (1989) maximum likelihood estimator, forecasting is carried out by extrapolating the estimated model. As in the univariate counterpart, we estimate a simple ARFIMA$(1, d, 1)$ model on both the realized covariation and its logarithmic transform.

10.5.1 Forecast evaluation

To analyze the forecast efficiency and information content of the different covariance estimators, we employ the popular approach of Mincer and Zarnowitz (1969) regressions on both the realized covariance and its logarithmic transformation. The regression takes the form:

$$V_{t+1}^{(m)} = \alpha + \beta_1 V_t^{(k)ARFIMA} + \epsilon_t, \tag{10.11}$$

with $V_{t+1}^{(m)}$ being the integrated covariance (or its logarithmic transformation) estimated using the mth estimator, namely, realized covariance defined by 8.7 , the bipower realized

covariance defined by 8.21, the two-scale realized covariance defined by 8.15, the multivariate realized kernel defined by 8.16 and our jump wavelet two-scale realized covariance estimator defined by 9.14. $V_t^{(k)ARFIMA}$ denotes the 1-day ahead forecast of $V_{t+1}^{(m)}$ using the kth estimator based on ARFIMA$(1, d, 1)$, while we consider the same estimators. We report in-sample as well as rolling out-of-sample results.

After testing the forecasting efficiency of the different covariance estimators, we would also like to test the information content of the wavelet decomposition of the realized covariance. For this purpose, we separately estimate ARFIMA$(1, d, 1)$ for all components $JWTSCV_j$ for $j = 1, \ldots, 5$ of the realized covariance as well as the estimated jumps. We should note that in the case of logarithmic transformation of the realized covariance, we also take logarithms of the decomposed levels $JWTSCV_j$. After obtaining the forecast for each level, we transform the forecasts back to be able to compare the results. For convenience, we refer to the estimators in the description of the results as RC, BC, TSCV, MRK and JWC, while BC and JWC are referring to the continuous part of the realized covariance estimators, and finally $\sum JWC$ referring to the sum of the individual forecasts of the decomposed realized covariations.

Finally, we test the information content of the separate decomposed realized covariances by running the following regressions:

$$JWC_{t+1} = \alpha + \beta_1 W_{t,j}^{ARFIMA} + \epsilon_t, \tag{10.12}$$

where $W_{t,j}^{ARFIMA}$ denotes the one-day ahead forecasts of the individual components $JWTSCV_j$ for $j = 1, \ldots, 5$, corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day, respectively, and

$$JWC_{t+1} = \alpha + \beta_1 J_t^{ARFIMA} + \epsilon_t, \tag{10.13}$$

where J_t^{ARFIMA} denotes the forecasts of the jumps. For now, we consider J_t to include both co-jumps and individual jumps and we will test its separate impact in the following section. Thus we test the information content of the long memory forecasts of the realized covariance estimators using the coefficient of determination, R^2, of the regression.

10.5.2 Does decomposition bring any improvement in covariation forecasting?

We use the period from January 5, 2007 to December 31, 2009 to perform the estimations of all the models. We refer to this period as the in-sample period and it contains the GBP-CHF, GBP-EUR and CHF-EUR pairs. The year 2010 is saved for comparison of the out-of-sample forecasts, which are done on a rolling basis.

Tables 10.4 and 10.5 present the results of the logarithmic transform of the realized covariation and the realized covariation, respectively. A striking result is that JWC is the easiest to forecast in terms of having the highest R^2 for all cases except the GBP-EUR pair, where BC results in a slightly higher R^2. Thus JWC seems to carry the most significant information in comparison with the other estimators. BC is the second

Table 10.4: Results for $\log \widehat{RC}_t$*:* R^2 *for the Minzer-Zarnowitz regressions regressing ARFIMA forecasts of RC, BC, TSCV, MRK, JWC and* $\sum JWC$ *on its estimates ,* W_j *denotes* $JWTSCV_j$*,* $j = 1, \ldots, 5$ *components of realized covariance and Jumps all jumps including co-jumps and individual jumps.*

GBP-CHF

	in-sample							out-of-sample						
	RC	BC	TSCV	MRK	JWC	∑JWC	Avg	RC	BC	TSCV	MRK	JWC	∑JWC	Avg
RC	0.744	0.746	0.748	0.738	0.752	0.723	0.742	0.336	0.297	0.364	0.377	0.342	0.358	0.346
BC	0.759	0.761	0.762	0.748	0.766	0.738	0.756	0.327	0.294	0.340	0.328	0.323	0.335	0.325
TSCV	0.736	0.736	0.748	0.747	0.756	0.720	0.741	0.291	0.264	0.306	0.338	0.292	0.309	0.300
MRK	0.684	0.684	0.697	0.701	0.702	0.669	0.689	0.222	0.217	0.228	0.241	0.240	0.241	0.232
JWC	0.788	0.791	0.800	0.798	0.807	0.774	**0.793**	0.349	0.343	0.352	0.369	0.355	0.373	**0.357**
Avg	0.742	0.744	0.751	0.746	**0.757**	0.725		0.305	0.283	0.318	**0.331**	0.310	0.323	
	W1	W2	W3	W4	W5	Jumps		W1	W2	W3	W4	W5	Jumps	
JWC	0.770	0.766	0.759	0.690	0.622	0.544		0.329	0.330	0.295	0.119	0.083	0.122	

GBP-EUR

	in-sample							out-of-sample						
	RC	BC	TSCV	MRK	JWC	∑JWC	Avg	RC	BC	TSCV	MRK	JWC	∑JWC	Avg
RC	0.834	0.836	0.836	0.828	0.843	0.833	0.835	0.329	0.328	0.305	0.333	0.322	0.348	0.328
BC	0.840	0.843	0.842	0.833	0.848	0.840	0.841	0.381	0.377	0.352	0.374	0.365	0.393	**0.373**
TSCV	0.828	0.831	0.832	0.827	0.840	0.830	0.831	0.254	0.260	0.233	0.259	0.253	0.277	0.256
MRK	0.789	0.791	0.793	0.792	0.799	0.789	0.792	0.196	0.201	0.174	0.197	0.196	0.219	0.197
JWC	0.860	0.863	0.865	0.858	0.871	0.862	**0.863**	0.338	0.339	0.315	0.321	0.316	0.352	0.330
Avg	0.830	0.833	0.834	0.828	**0.840**	0.831		0.300	0.301	0.276	0.297	0.290	**0.318**	
	W1	W2	W3	W4	W5	Jumps		W1	W2	W3	W4	W5	Jumps	
JWC	0.867	0.864	0.816	0.634	0.378	0.522		0.310	0.297	0.262	0.043	0.053	0.015	

CHF-EUR

	in-sample							out-of-sample						
	RC	BC	TSCV	MRK	JWC	∑JWC	Avg	RC	BC	TSCV	MRK	JWC	∑JWC	Avg
RC	0.773	0.776	0.776	0.768	0.785	0.784	0.777	0.406	0.374	0.396	0.398	0.395	0.394	0.394
BC	0.777	0.780	0.779	0.769	0.787	0.785	0.780	0.433	0.395	0.421	0.427	0.412	0.401	0.415
TSCV	0.769	0.771	0.776	0.770	0.783	0.783	0.775	0.285	0.269	0.260	0.278	0.266	0.275	0.272
MRK	0.721	0.722	0.729	0.727	0.733	0.735	0.728	0.302	0.282	0.277	0.288	0.280	0.285	0.285
JWC	0.826	0.828	0.831	0.821	0.838	0.838	**0.830**	0.484	0.466	0.433	0.453	0.450	0.457	**0.457**
Avg	0.773	0.776	0.778	0.771	**0.785**	**0.785**		**0.382**	0.357	0.357	0.369	0.361	0.362	
	W1	W2	W3	W4	W5	Jumps		W1	W2	W3	W4	W5	Jumps	
JWC	0.837	0.830	0.815	0.724	0.676	0.391		0.428	0.402	0.385	0.084	0.086	0.231	

best, which confirms that the continuous part of the realized covariance has the highest information content. The other estimators, which are not robust to jumps, are more difficult to forecast.

JWC also forecasts all the other realized covariance estimates best, again except for the case of the GBP-EUR pair. When we decompose the realized covariation, forecast its components individually and then use the sum of the forecasts, it does not seem to bring the improvement seen in the univariate case of forecasting realized volatility. The separate realized covariances also carry quite a large information content, as the first three are able to forecast the realized covariance similarly well. In other words, the 5–10 minute covariation component is able to forecast the total covariation JWC with a similar

Table 10.5: Results for $\widehat{RC}_t$: R^2 for the Minzer-Zarnowitz regressions regressing ARFIMA forecasts of RC, BC, TSCV, MRK, JWC and $\sum$JWC on its estimates , W_j denotes $JWTSCV_j$, $j = 1, \ldots, 5$ components of realized covariance and Jumps all jumps including co-jumps and individual jumps.

GBP-CHF

	in-sample							out-of-sample						
	RC	BC	TSCV	MRK	JWC	∑JWC	Avg	RC	BC	TSCV	MRK	JWC	∑JWC	Avg
RC	0.733	0.730	0.739	0.734	0.738	0.737	0.735	0.338	0.332	0.341	0.348	0.354	0.322	0.339
BC	0.740	0.738	0.747	0.741	0.748	0.745	0.743	0.343	0.333	0.335	0.337	0.350	0.322	0.337
TSCV	0.720	0.716	0.731	0.732	0.733	0.734	0.728	0.312	0.303	0.306	0.319	0.311	0.282	0.306
MRK	0.668	0.663	0.680	0.684	0.680	0.682	0.676	0.263	0.262	0.252	0.258	0.262	0.240	0.256
JWC	0.773	0.772	0.785	0.786	0.787	0.787	**0.782**	0.419	0.419	0.399	0.401	0.402	0.378	**0.403**
Avg	0.727	0.724	0.736	0.735	**0.737**	**0.737**		0.335	0.330	0.326	0.333	**0.336**	0.309	
	W1	W2	W3	W4	W5	Jumps		W1	W2	W3	W4	W5	Jumps	
JWC	0.788	0.777	0.749	0.742	0.736	0.593		0.377	0.391	0.384	0.231	0.208	0.134	

GBP-EUR

	in-sample							out-of-sample						
	RC	BC	TSCV	MRK	JWC	∑JWC	Avg	RC	BC	TSCV	MRK	JWC	∑JWC	Avg
RC	0.836	0.835	0.838	0.833	0.842	0.837	0.837	0.365	0.387	0.347	0.349	0.366	0.365	0.363
BC	0.842	0.841	0.845	0.838	0.849	0.844	0.843	0.421	0.440	0.399	0.394	0.412	0.410	**0.413**
TSCV	0.830	0.828	0.834	0.830	0.839	0.835	0.833	0.301	0.326	0.289	0.299	0.307	0.305	0.305
MRK	0.798	0.796	0.803	0.800	0.806	0.804	0.801	0.245	0.268	0.231	0.237	0.249	0.255	0.247
JWC	0.862	0.862	0.867	0.863	0.871	0.867	**0.865**	0.415	0.433	0.398	0.396	0.402	0.393	0.406
Avg	0.834	0.832	0.837	0.833	**0.842**	0.837		0.349	**0.371**	0.333	0.335	0.347	0.346	
	W1	W2	W3	W4	W5	Jumps		W1	W2	W3	W4	W5	Jumps	
JWC	0.864	0.864	0.844	0.829	0.830	0.715		0.394	0.385	0.335	0.276	0.339	0.187	

CHF-EUR

	in-sample							out-of-sample						
	RC	BC	TSCV	MRK	JWC	∑JWC	Avg	RC	BC	TSCV	MRK	JWC	∑JWC	Avg
RC	0.741	0.743	0.741	0.728	0.747	0.743	0.740	0.413	0.406	0.383	0.377	0.401	0.383	0.394
BC	0.755	0.759	0.755	0.741	0.765	0.758	0.755	0.426	0.415	0.392	0.388	0.403	0.390	0.403
TSCV	0.725	0.724	0.731	0.722	0.733	0.731	0.728	0.359	0.354	0.328	0.330	0.338	0.321	0.338
MRK	0.674	0.671	0.682	0.675	0.680	0.681	0.677	0.311	0.307	0.278	0.283	0.289	0.281	0.291
JWC	0.796	0.799	0.801	0.789	0.806	0.802	**0.799**	0.516	0.508	0.468	0.461	0.468	0.451	**0.479**
Avg	0.738	0.739	0.742	0.731	**0.746**	0.743		0.405	**0.398**	0.370	0.368	0.380	0.365	
	W1	W2	W3	W4	W5	Jumps		W1	W2	W3	W4	W5	Jumps	
JWC	0.805	0.796	0.766	0.757	0.755	0.510		0.451	0.449	0.445	0.261	0.322	0.277	

forecasting power as if the total JWC was used. Thus, even though decomposition does not bring an overall improvement, we can see that the realized covariation at the higher frequency carries the most important information. In other words, the main part of the realized covariation comes from the highest frequency.

Finally, we can see that jumps carry important information which may help to forecast the realized covariation. All the estimated parameters are significantly different from zero and the in-sample fits describe the data well. For reasons of space, we do not provide all the results here and we proceed to test the impact of further decomposition of the jumps into individual jump and co-jump components.

10.5.3 Impact of jumps and co-jumps on the covariance forecasts

In the small sample study, we have shown that co-jumps cause large bias in the covariance estimates, while individual jumps cause much lower bias. Thus, we would like to see if further decomposition to co-jumps and individual jumps can help to forecast the realized covariances. For this purpose, we construct an ARFIMA$(1, d, 1)$ model for the jump and co-jump components of the realized covariance estimated using our methodology and test for the informational efficiency of each of them to the realized covariance forecast using the encompassing regression:

$$JWC_{t+1} = \alpha + \beta_1 JWC_t^{ARFIMA} + \beta_2 J_{co-jumps,t}^{ARFIMA} + \beta_3 J_{1,t}^{ARFIMA} + \beta_4 J_{2,t}^{ARFIMA} + \epsilon_t, \quad (10.14)$$

where JWC_t^{ARFIMA} denotes the one-day ahead forecast of JWC_{t+1} and $J_{co-jumps,t}^{ARFIMA}$ denotes the forecast of co-jumps, while $J_{1,t}^{ARFIMA}$ and $J_{2,t}^{ARFIMA}$ denote the forecasts of individual jumps of both assets in the forecasted pair.

With the help of the encompassing regressions, we can test if jumps contain any information relevant to the covariation forecasts. We will first test the information content of $J_{co-jumps,t}^{ARFIMA}$, $J_{1,t}^{ARFIMA}$ and $J_{2,t}^{ARFIMA}$ separately by setting all other βs to zero. Then, we will add parameters to the regression, starting with α and β_1, and adding β_2, β_3 and β_4 gradually to see if they bring any information which is not contained in the realized covariation forecast itself. If, for example, common jumps carry information important for the forecast, parameter β_2 will be significantly different from zero, even if parameter β_1 is significantly different from zero.

Table 10.6 summarizes the results for the logarithmic transform of the realized covariances. The results vary, but in the individual regressions, the jumps seem to carry some significant information in the case of the GBP-CHF pair and the CHF-EUR pair. When testing their significance in the presence of the realized covariance forecast JWC_t^{ARFIMA}, we can see that in most cases they are not significant, even though they slightly improve the R^2 of the regressions.

Table 10.7 shows the results for the realized covariances, which are more interesting. Striking evidence of the significance co-jumps for the forecasts is found in all cases (in only two cases the parameter has p-values of 0.103 and 0.104, so we can consider it to be marginally significant at the 89% level of significance). The presence of co-jumps in the encompassing regression also significantly improves the R^2 in comparison with the JWC estimate.

To conclude, we have shown that the decomposition of the realized covariation into a continuous part and co-jumps using our wavelet-based methods can help improve the forecasting significantly.

10.5.4 Forecasting of correlations

While co-jumps cause large bias in the covariance measures, individual jumps may cause bias to the correlation. Thus, we would like to complete our forecasting exercise by

Table 10.6: R^2 from encompassing regression of ARFIMA on $\log \widehat{RC}_t^{1/2}$ estimator JWC, co-jumps (Jcom) and individual jumps (J1 and J2). p-values of estimated parameters in parentheses.

	const.	JWC	Jcom	J1	J2	R^2
GBP-CHF	**-6.360** (0.000)		**-0.337** (0.000)			0.101
	-4.304 (0.000)			**0.210** (0.029)		0.038
	-4.687 (0.000)				**0.152** (0.094)	0.023
	-1.296 (0.014)	**0.769** (0.000)				0.355
	-1.717 (0.014)	**0.727** (0.000)	-0.080 (0.353)			0.360
	-1.932 (0.010)	**0.763** (0.000)	-0.079 (0.358)	-0.069 (0.425)		0.363
	-2.196 (0.007)	**0.773** (0.000)	-0.098 (0.271)	-0.046 (0.610)	-0.071 (0.401)	0.367
	const.	JWC	Jcom	J1	J2	R^2
GBP-EUR	**-5.198** (0.000)		0.040 (0.421)			0.005
	-5.284 (0.000)			0.009 (0.929)		0.000
	-4.819 (0.000)				0.087 (0.241)	0.011
	-1.592 (0.002)	**0.704** (0.000)				0.316
	-1.610 (0.002)	**0.729** (0.000)	-0.043 (0.316)			0.321
	-2.562 (0.000)	**0.794** (0.000)	-0.010 (0.815)	**-0.229** (0.016)		0.353
	-2.498 (0.000)	**0.791** (0.000)	-0.011 (0.800)	**-0.230** (0.016)	0.015 (0.805)	0.354
	const.	JWC	Jcom	J1	J2	R^2
CHF-EUR	**-5.654** (0.000)		-0.087 (0.143)			0.018
	-8.297 (0.000)			**-0.491** (0.000)		0.230
	-6.468 (0.000)				**-0.187** (0.023)	0.042
	-0.952 (0.032)	**0.823** (0.000)				0.450
	-0.890 (0.086)	**0.828** (0.000)	0.011 (0.814)			0.450
	-2.548 (0.004)	**0.716** (0.000)	0.037 (0.426)	**-0.192** (0.021)		0.474
	-2.553 (0.005)	**0.717** (0.000)	0.037 (0.429)	**-0.191** (0.042)	-0.003 (0.969)	0.474

Table 10.7: R^2 from encompassing regression of ARFIMA on $\widehat{RC}_t^{1/2}$ estimator JWC, co-jumps (Jcom) and individual jumps (J1 and J2). p-values of estimated parameters in parentheses.

	const.	JWC	Jcom	J1	J2	R^2
GBP-CHF	**0.003** (0.000)		**1.342** (0.000)			0.118
	0.003 (0.000)			0.677 (0.174)		0.015
	0.005 (0.000)				-0.286 (0.409)	0.006
	0.001 (0.002)	**0.722** (0.000)				0.402
	0.001 (0.004)	**0.665** (0.000)	**0.608** (0.034)			0.424
	0.001 (0.234)	**0.667** (0.000)	0.585 (0.103)	0.052 (0.913)		0.424
	0.001 (0.262)	**0.664** (0.000)	0.586 (0.104)	0.065 (0.892)	-0.071 (0.794)	0.424
	const.	JWC	Jcom	J1	J2	R^2
GBP-EUR	**0.004** (0.000)		**1.031** (0.000)			0.104
	0.004 (0.000)			0.676 (0.138)		0.018
	0.000 (0.975)				**2.909** (0.000)	0.160
	0.001 (0.000)	**0.694** (0.000)				0.402
	0.001 (0.004)	**0.648** (0.000)	**0.627** (0.006)			0.439
	0.002 (0.018)	**0.654** (0.000)	**0.697** (0.004)	-0.296 (0.434)		0.442
	0.002 (0.162)	**0.656** (0.000)	**0.703** (0.012)	-0.300 (0.443)	-0.031 (0.963)	0.442
	const.	JWC	Jcom	J1	J2	R^2
CHF-EUR	**0.003** (0.000)		**1.695** (0.000)			0.267
	0.006 (0.000)			**-0.610** (0.054)		0.030
	0.005 (0.000)				-0.217 (0.696)	0.001
	0.001 (0.016)	**0.794** (0.000)				0.468
	0.001 (0.015)	**0.670** (0.000)	**0.615** (0.019)			0.491
	0.001 (0.106)	**0.666** (0.000)	**0.614** (0.020)	-0.044 (0.852)		0.491
	-0.000 (0.951)	**0.725** (0.000)	**0.504** (0.056)	-0.191 (0.432)	**0.932 (0.033)**	0.511

creating a forecasting model of the realized correlations, and we again construct an ARFIMA$(1, d, 1)$ model for the realized correlation.

In the previous sections, we have shown that realized correlation estimated using a wavelet-based estimator is much smoother with lower confidence intervals than the correlation estimated using the standard realized variance and covariance measures. Thus we would like to see if our estimate carries better information for forecasting correlations. For this purpose, we again employ encompassing regression. This time, we will test the informational efficiency of each of the two measures. Moreover, we would like to see if decomposition of realized correlation generates any significant improvement. Thus we will forecast the decomposed correlations individually, and then compare the sum of the forecasts with the latter two estimates in the following way:

$$Corr_{t+1} = \alpha + \beta_1 Rcorr_t^{ARFIMA} + \beta_2 WRCorr_t^{ARFIMA} + \beta_3 \sum_{i=1}^{5} WRcorr_{t,i}^{ARFIMA} + \epsilon_t, \tag{10.15}$$

where $Rcorr_t^{ARFIMA}$ denotes the one-day ahead forecast of correlation using the standard realized correlation, $WRCorr_t^{ARFIMA}$ denotes the forecast using wavelet-based correlation and $\sum_{i=1}^{5} WRcorr_{t,i}^{ARFIMA}$ denotes the sum of the individual forecasts of decomposed correlation using our wavelet estimator.

We also run individual regressions testing the forecasting power of the individual estimators:

$$Corr_{t+1}^{(m)} = \alpha + \beta_1 V_t^{(k)ARFIMA} + \epsilon_t, \tag{10.16}$$

where $Corr_{t+1}^{(m)}$ is the realized correlation estimated using the mth estimator, and $V_t^{(k)ARFIMA}$ denotes the one-day ahead forecast of $Corr_{t+1}^{(m)}$ using the kth estimator, while we consider the same three estimators, $Rcorr_t^{ARFIMA}$, $WRCorr_t^{ARFIMA}$ and $\sum_{i=1}^{5} WRcorr_{t,i}^{ARFIMA}$.

Table 10.8 summarizes the results of the individual regressions as well as the encompassing regressions for both the in-sample and the out-of-sample periods, which are the same as in the covariance forecasting exercise.

The results from the in-sample fits tell us that our WRcorr is a very efficient estimator for forecasting of realized correlations, as its coefficient is significantly different from zero but is not significantly different from 1, while the forecast is unbiased as the constant coefficient is not significantly different from zero, except in some cases. Moreover, the WRcorr forecasts also carry the highest R^2. The sum of the individual correlation forecasts do not seem to be as efficient as the WRcorr estimator and it also gives slightly biased results. The realized correlation also seems to be quite an efficient and unbiased estimator, even though its coefficient is rather higher than 1 in some cases. It still has the lowest R^2. When looking at the results from the encompassing regressions, we can see that the WRcorr estimator remains the only significant estimator in the regression. Its coefficient is slightly lower than 1, but the coefficients of the other two estimators are not significantly different from zero. This means that these estimators do not generate any other significant information for the correlation forecasts.

Table 10.8: R^2 M-Z regression of ARFIMA on Correlations. Rcorr denotes realized correlation estimate, WRcorr wavelet-based realizd correlation and $\sum$WRcorr sum of individual forecasts of decomposed correlation. p-values of estimated parameters in parentheses.

GBP-CHF

	in-sample					out-of-sample				
	const.	RCorr	WRcorr	$\sum$WRcorr	R^2	const.	RCorr	WRcorr	$\sum$WRcorr	R^2
RCorr	-0.02 (0.47)	**1.04** (0.00)			0.33	0.04 (0.46)	**0.83** (0.00)			0.22
WRcorr	-0.00 (0.92)		**1.01** (0.00)		0.42	0.04 (0.42)		**0.85** (0.00)		0.26
$\sum$WRcorr	**-0.08** (0.00)			**1.16** (0.00)	0.41	-0.02 (0.78)			**0.98** (0.00)	0.24
WRcorr	-0.02 (0.55)	0.13 (0.29)	**0.91** (0.00)		0.42	0.03 (0.63)	0.29 (0.40)	**0.61** (0.05)		0.27
WRcorr	0.01 (0.76)	0.21 (0.14)	**1.17** (0.00)	-0.38 (0.21)	0.43	0.09 (0.30)	0.44 (0.25)	**1.14** (0.07)	-0.79 (0.32)	0.27

GBP-EUR

	in-sample					out-of-sample				
	const.	RCorr	WRcorr	$\sum$WRcorr	R^2	const.	RCorr	WRcorr	$\sum$WRcorr	R^2
RCorr	-0.04 (0.26)	**1.07** (0.00)			0.28	0.12 (0.44)	**0.80** (0.00)			0.08
WRcorr	-0.01 (0.75)		**1.02** (0.00)		0.36	**0.23** (0.06)		**0.62** (0.00)		0.07
$\sum$WRcorr	**-0.20** (0.00)			**1.31** (0.00)	0.33	0.09 (0.66)			**0.86** (0.01)	0.06
WRcorr	-0.02 (0.64)	0.04 (0.72)	**0.99** (0.00)		0.36	-0.05 (0.80)	0.13 (0.67)	**0.96** (0.03)		0.11
WRcorr	-0.00 (0.93)	0.07 (0.63)	**1.03** (0.00)	-0.08 (0.74)	0.36	0.07 (0.79)	0.43 (0.48)	**0.97** (0.03)	-0.51 (0.57)	0.11

CHF-EUR

	in-sample					out-of-sample				
	const.	RCorr	WRcorr	$\sum$WRcorr	R^2	const.	RCorr	WRcorr	$\sum$WRcorr	R^2
RCorr	-0.00 (0.92)	**1.00** (0.00)			0.50	0.06 (0.29)	**0.86** (0.00)			0.38
WRcorr	0.00 (0.93)		**1.00** (0.00)		0.56	0.08 (0.17)		**0.85** (0.00)		0.41
$\sum$WRcorr	**-0.16** (0.00)			**1.20** (0.00)	0.51	0.02 (0.78)			**0.90** (0.00)	0.34
WRcorr	-0.00 (0.94)	0.08 (0.33)	**0.93** (0.00)		0.57	-0.03 (0.64)	0.08 (0.77)	**0.96** (0.00)		0.46
WRcorr	0.03 (0.36)	0.12 (0.18)	**1.04** (0.00)	-0.19 (0.19)	0.57	0.00 (0.96)	0.28 (0.48)	**0.93** (0.00)	-0.22 (0.50)	0.46

When looking at the results for the out-of-sample period, which are much more important as these are the real forecasts, we still have a very similar picture. WRcorr is unaffected in the encompassing regressions, being the only significant estimator. In the individual regressions, the sum of the decomposed forecasts surprisingly seems to be the most efficient estimator, as its coefficient is closest to one, but it has a lower coefficient of determination, R^2, than the WRcorr estimator. To summarize the results from this section, we show that the wavelet-based estimator of the realized correlation is able to bring a significant improvement to the forecasting of correlation.

10.6 Application to portfolio: Realized beta decomposition

In the final section of the empirical application, we would like to use our theory in the estimation and decomposition of the realized beta. The theory presented by Barndorff-Nielsen and Shephard (2004a) shows that the beta of an asset can be consistently estimated by the ex-post realized regression coefficient obtained as the ratio of the unbiased estimators of the average realized covariance and the realized market variance – see Section 9.5 for the theory. For the purpose of this section, we construct an artificial equi-weighted portfolio consisting of all three currencies used in this analysis, and compute the realized beta for each of the currencies in the portfolio.

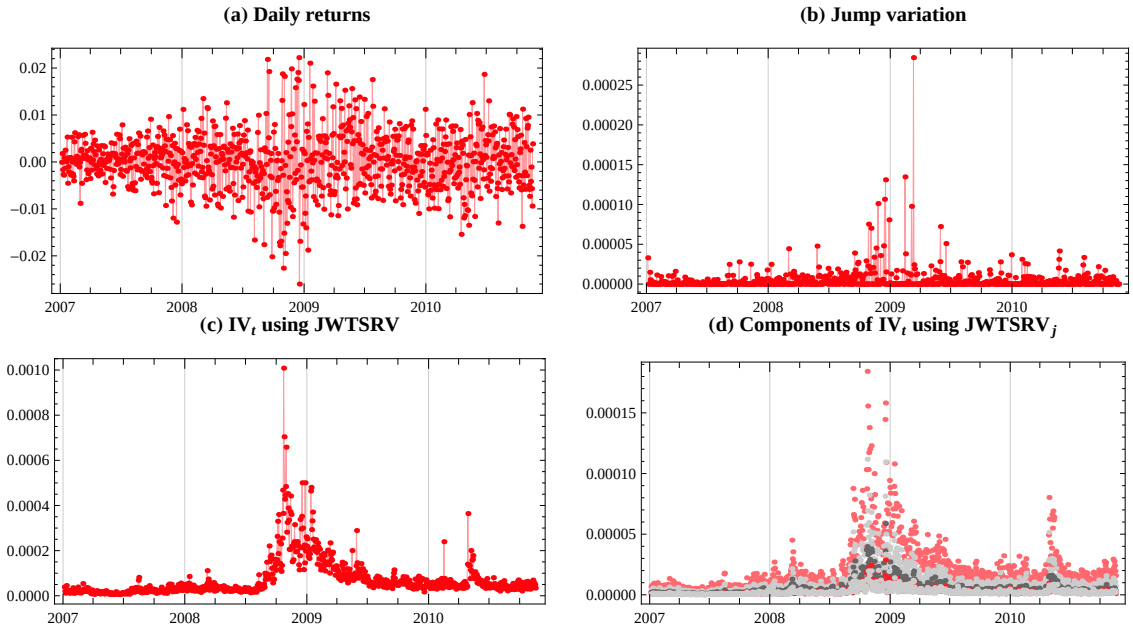


Figure 10.9: Portfolio: (a) daily returns, (b) JWTSRV estimated jump variation, (c) IV_t estimated by JWTSRV (d) decomposition of IV_t using $JWTSRV_j$ for $j = 1, \ldots, 5$ components corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day.

Using intra-day prices, we essentially observe the daily covariances and can model them non-parametrically. Thus instead of estimating nonlinear models via maximum likelihood procedures, we can simply use the data and compute the observed daily covariances. Common problem of this approach is that in large portfolios the realized covariance matrix is not positive definite. In our low-dimensional setting, this is not the case. Specifically, if number of assets in portfolio will be larger than number of intraday observations, $m > 1/\delta$, where $\delta = 1\ M$ with sampling frequency M, the estimated covariance matrix will not be of full rank and it will fail to be positive definite, as discussed by Andersen et al. (2003). For example the use of thirty-minute returns, corresponding to $1/\delta = 48$ intraday observations will require number of assets in the portfolio m no larger that 48. Generally, our approach do not solve this problem, as wavelet estimator only decomposes the realized measures asymptotically, but as we analyze low-dimensional portfolio, the positive semidefinitness of variance-covariance matrix is guaranteed.

We would like to test for price co-jumps in a equally weighted portfolio constructed from the currencies. Using our methodology described in Section 9.3, we will study the impact of co-jumps on the portfolio.

For the ease of the notation in our illustration, we consider equally weighted portfolio, but the results hold without loss of generality for any well-diversified portfolio. Consider the i-th intraday return on an equally weighted portfolio of m assets

$$r_{(EP)t,i} = \frac{1}{m}\sum_{j=1}^{m} r_{i,t,j} \tag{10.17}$$

The realized variation for our portfolio can be estimated as:

$$\widehat{RV}_{(EP)t} = \sum_{i=1}^{M} r^2_{(EP)t,i} = \sum_{i=1}^{M}\left(\frac{1}{m}\sum_{j=1}^{m} r_{i,t,j}\right)^2, \tag{10.18}$$

While with increasing sampling frequency M to ∞, we have

$$\begin{aligned}\text{plim}_{M\to\infty}\,\widehat{RV}_{(EP)t} &= \underbrace{\frac{1}{m^2}\sum_{j=1}^{m}\int_{t-1}^{t}\sigma^2_{j,s}ds + \frac{1}{m^2}\sum_{j=1}^{m}\sum_{l=1,l\neq j}^{m}\int_{t-1}^{t}\sigma_{j,s}\sigma_{j,s}ds}_{\text{Continuous Part}} \\ &+ \underbrace{\frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{N_j,t}J^2_{j,t,k} + \frac{1}{m^2}\sum_{j=1}^{m}\sum_{l=1,l\neq j}^{m}\sum_{k=1}^{N^*_t}J_{j,t,k}J_{l,t,k}}_{\text{Jump Part}},\end{aligned} \tag{10.19}$$

where N^*_t is counting process of co-jumps occuring simultaneously across all m assets. While we know that we are able to consistently estimate jumps in all assets using our theory, we can then separate quadratic variation of portfolio into its continuous part, and jump part. With large m, the first part of the jump part become negligible, the second

part - co-jumps is of interest. Using our theory, we can easily identify and estimate all co-jumps of the portfolio using Definition 20.

Figure 10.9 shows the estimated continuous part of the realized variance for our equally weighted portfolio of all three currencies, together with total jump variation, as in Equation 10.19. Moreover, Figure 10.9 also brings the decomposition of the continuous part of the variance of portfolio.

Figure 10.10 shows the beta estimates with their 95% confidence intervals. We can see that for all three currencies, the beta is changing significantly, hence it is worth working with a time-varying beta in the analysis. Moreover, Figure 10.10 contains the decomposition of the beta into several investment horizons. We again observe most activity coming from the fastest frequency of 5–10 minutes, with an average beta of 0.47 for GBP for example. The other scales have betas of 0.23, 0.11, 0.054 and 0.054. A simple sum of the betas across the scales gives us a total beta for GBP of 0.92. Thus, the results from the previous analysis suggesting that most of the action comes from fast time scales are again confirmed.

Figure 10.11 compares the beta estimators. Specifically, it compares our wavelet-based beta estimator and the simple realized beta estimator. The figure has fixed axes, so the two estimators can be compared easily even at a quick glance. We can see that our wavelet-based beta has much lower variance, thus bringing much more precision into the estimated result.

To be precise, GBP, CHF and EUR have betas computed using the wavelet-based realized beta (with 95% confidence intervals in parentheses) of 0.923 ($\pm$0.1), 1.02 ($\pm$0.098) and 0.94 ($\pm$0.78), while the realized betas computed with the simple estimator are 0.95

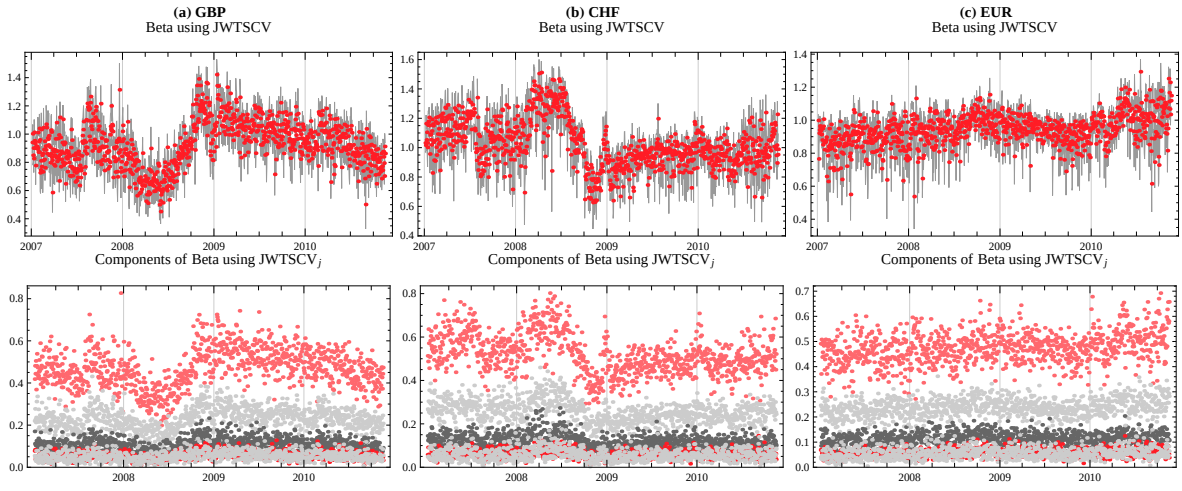


Figure 10.10: Beta estimates with 95% confidence intervals in first row and decomposition of beta using $JWTSCV_j$ for $j = 1, \ldots, 5$ corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day in the second row. (a) GBP futures, (b) CHF futures and (c) EUR futures pair.

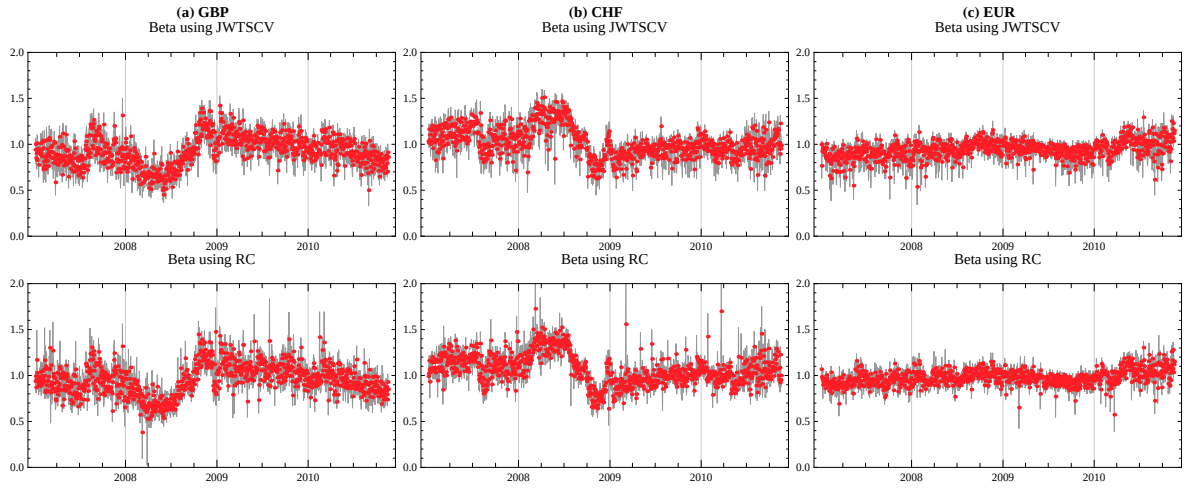


Figure 10.11: Beta estimates with 95% confidence intervals using wavelet-based estimator in first row and realized beta in second row. (a) GBP futures, (b) CHF futures and (c) EUR futures pair.

(±0.1), 1.06 (±0.09) and 0.97 (±0.07), respectively. Our estimator thus estimates the beta to be lower on average by 0.03.

The result again has a significant economic impact for asset pricing and portfolio diversification, as our estimator generates more precise estimates. Another aspect is to look at the dynamics of the realized beta and its decomposition, which give us a better understanding of the stock market generating process.

Finally, let us note that in our work we simply use the daily decomposition for illustration, but the methodology is much more powerful, as it may be used for any period $[t-h,t]$. In fact, the level of decomposition is constrained only by the time horizon used – in our case one day. For example, considering one month, we can simply use our estimator of realized measures to study investment horizons ranging from intraday, say one-minute, horizons up to the one-month horizon. This way, we have a very general method of estimation of realized measures not only in the time domain, but also in the frequency domain.

Chapter 11

Conclusion

The second part of this dissertation follows the structure of the first part closely. After the necessary introduction of the generalized multivariate framework for modeling the covariation structure between processes, we build a new, wavelet-based realized covariation theory by extending the findings from the univariate part, and we define the wavelet-based realized estimator of covariance together with its properties. We use wavelets to disentangle jumps from co-jumps, which is crucial in the study of multivariate dependencies. Having defined the estimators of variance and covariance, we also define the transformations of interest for portfolio theory: the wavelet-based realized correlation measure and the wavelet-based realized beta. Similarly to the univariate findings, the presented theory provides a new type of multivariate estimators in the time-frequency domain.

To support the theoretical results, we again run a numerical study of the small sample behavior of the estimators. In the study, we simulate prices using bivariate jump-diffusion stochastic volatility process and compare the performance of the wavelet-based realized covariation and correlation estimators with the popular realized covariance, bipower realized covariance and two-scale realized covariance estimators. The study proves that in this generalized setting as well, our wavelet-based realized theory is able to outperform other methods of estimation, as it displays the lowest bias under different amounts of simulated noise and jumps in the bivariate process.

The last chapter applies the multivariate theory and studies the decomposition of integrated covariation, correlation and beta on the forex markets. Our estimator is able to separate jumps, co-jumps and *true* covariation from the data. It is also robust to the Epps effect caused by noise in the data. The results suggest that understanding jumps and co-jumps in a multivariate setting may be crucial for studying the dependencies. While individual jumps bring some bias to the covariance, co-jumps introduce large bias into the covariation measure. The impact on correlation is even more crucial. Individual jumps in the processes bring large downward bias to the correlation measure, while co-jumps introduce upward bias with a smaller magnitude.

The empirical part also contains an interesting study of multivariate unconditional

volatility distributions and their decomposition into several investment horizons. While the multivariate volatilities show strong dependence, a volatility-in-correlation effect suggests that the standard mean-variance efficiency calculations based on constant correlations are misguided. Our results have significant economic value, as a wrong assumption about the dependence process will have a direct impact on the portfolio valuation. The dynamics of the decomposed dependencies reveal interesting results as well. Our wavelet-based realized theory generates a more precise correlation measure with narrower confidence intervals than the standard realized correlations. A study of the temporal dependence in the decomposed correlations reveals that a similar heterogeneous autoregression type model as in the univariate case should be used for forecasting.

Similarly to the univariate part, therefore, we build a forecasting model for covariation and correlation based on wavelet decomposition. Our model again outperforms all other models in-sample as well as out-of-sample. We bring the study on the impact of co-jumps and individual jumps on the covariance and correlation forecasts and we find that proper accounting for jumps and co-jumps bring significant improvement in the forecasting of covariance and correlation measures.

In conclusion, this dissertation presents a new theoretical framework generalizing the popular concept of realized variance and covariance. The work also contributes to the literature by providing interesting empirical findings from our time-frequency realized measures.

References

Aït-Sahalia, Y., J. Fan, and D. Xiu (2010). High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association 105*, 1504–1517.

Aït-Sahalia, Y. and J. Jacod (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics 37*(1), 184–222.

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

Andersen, T. and L. Benzoni (2007). Realized volatility. In T. Andersen, R. Davis, J. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*. Springer Verlag.

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

Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

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

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* (71), 579–625.

Andersen, T., T. Bollerslev, and X. Huang (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics 160*(1), 176–189.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics* (4), 701–720.

Andersen, T. G., T. Bollerslev, and D. Dobrev (2007). No-arbitrage semi- martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics* (138), 125–180.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2009). Jump-robust volatility estimation using nearest neighbor truncation. *Working paper, National Bureau of Economic Research*.

Antoniou, I. and K. Gustafson (1999). Wavelets and stochastic processes. *Mathematics and Computers in Simulation 49*, 81–104.

Arneodo, A., J. Muzy, and D. Sornette (1998). Casual cascade in stock market from the "infrared" to the "ultraviolet". *European Physical Journal B* (2), 277–282.

Audrino, F. and F. Corsi (2010). Modeling tick-by-tick realized correlations. *Computational Statistics and Data Analysis 54*, 2372–2382.

Back, K. (1991). Asset prices for general processes. *Journal of Mathematical Economics* (20), 317–395.

Bandi, F. and J. Russell (2006a). Separating microstructure noise from volatility. *Journal of Financial Economics* (79), 655–692.

Bandi, F. and J. Russell (2006b). Volatility. In J. Birge and V. Linetsky (Eds.), *Handbook of Financial Engineering*. Elsevier.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics 162*(2), 149–169.

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

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

Barndorff-Nielsen, O. and N. Shephard (2002b). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* (17), 457–477.

Barndorff-Nielsen, O. and N. Shephard (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

Barndorff-Nielsen, O. and N. Shephard (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* (2), 1–48.

Barndorff-Nielsen, O. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* (4), 1–30.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* (81), 637–654.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economic and Statistics 69*, 542–547.

Brillinger, D. R. (1981). *Time Series: Data analysis and Theory*. Holden-Day Series in Time Series Analysis. San Francisco: Holden-Day.

Calderón, A. P. (1964). Intermediate spaces and interpolation, the complex method. *Studia Math. 24*, 113–190.

Capobianco, E. (2004). Multiscale stochastic dynamics in finance. *Physica A* (344), 122–127.

Chui, C. (1992). *An Inroduction to Wavelets*. Academic Press, New York.

Comte, F. and E. Renault (1999). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* (8), 291–323.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Corsi, F., D. Pirino, and R. Renò (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* (2), 276–288.

Dacorogna, M., U. Müller, R. Nagler, R. Olsen, and O. Pictet (2001). *An Introduction to High-Frequency Finance*. Academic Press, San Diego.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics 41*, 909–996.

Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM.

Deo, R. S., C. M. Hurvich, and Y. Lu (2006). Forecasting realized volatility using a long memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics 131*(1-2), 29–58.

Di Matteo, T., T. Aste, and M. Dacorogna (2003). Scaling behaviors in differently developed markets. *Physica A* (324).

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association* (74), 291–296.

Fama, E. (1965). The behavior of stock market prices. *Journal of Business 38*, 34–105.

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

Fan, J. and Y. Wang (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association* (102), 1349–1362.

Fleming, J. and B. Paye (2011). High-frequency returns, jumps and the mixture of normals hypothesis. *Journal of Econometrics 160*(1), 119–128.

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

Gençay, R., F. Selçuk, and B. Whitcher (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press.

Gençay, R., N. Gradojevic, F. Selçuk, and B. Whitcher (2010). Asymmetry of information flow between volatilities across time scales. *Quantitative Finance*, 1–21.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*(2), 227–238.

Granger, C. W. J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* (1), 15–29.

Griffin, J. and R. Oomen (2011). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics 160*(1), 58–68.

Hansen, P. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics 24*(2), 127–161.

Harris, F., T. McInish, G. Shoesmith, and R. Wood (1995). Cointegration, error correction and price discovery on informationally-linked security markets. *Journal of Financial and Quantitative Analysis 30*, 563–581.

Haslett, J. and A. E. Raftery (1989). Space-time modelling with long-memory dependence: Assessing ireland's wind power resource. *Applied Statistics 38*, 1–50.

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

Hø g, E. and A. Lunde (2003). Wavelet estimation of integrated volatility. *Working Paper. Aarhus School of Business.*.

Holschneider, M., R. Kronland-Martinet, J. Morlet, and P. Tchamitchian (1989). Areal-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets, Time-Frequency Methods and Phase Space*, pp. 289–297. Springer-Verlag.

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics 3*, 456–499.

Igari, S. (1998). *Real analysis - With an Introduction to Wavelet Theory*, Volume 177 of *Mathematical Monographs*. AMS.

Jiang, G, J. and C. Oomen, R (2008). Testing for jumps when asset prices are observed with noise - a swap variance approach. *Journal of Econometrics* (144), 352–370.

Lanne, M. (2007). Forecasting realized exchange rate volatility by decomposition. *International Journal of Forecasting* (23), 307–320.

LeBaron, B. (2001). Stochastic volatility as a simple generator of financial power laws and long memory. *Quantitative Finance* (1), 621–631.

Lee, S. and P. A. Mykland (2008). Jumps in financial markets: a new nonpara- metric test and jump dynamics. *The Review of Financial Studies 21*, 2525–2563.

Lee, S. S. and J. Hannig (2010). Detecting jumps from lèvy jump diffusion processes. *Journal of Financial Economics 96*(2), 271–290.

Lindsay, R., D. Percival, and D. Rothrock (1996). The discrete wavelet transform and the scale analysis of the surface properties of sea ice. *Transactions on Geoscience and Remote Sensing 34*, 771–787.

Lintner, J. (1965). Security prices, risk and maximal gains from diversification. *Journal of Finance* (20), 587–615.

Lynch, P. and G. Zumbach. (2003). Market heterogeneities and the causal structure of volatility. *Quantitative Finance* (3), 320–331.

Mallat, S. (1998). *A wavelet tour of signal processing*. Academic Press.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics 36*, 270–296.

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

Marinucci, D. and P. Robinson (1999). Alternative forms of fractional brownian motion. *Journal of Statistical Planning and Inference* (80), 111–122.

Martens, M. and J. Zein (2004). Predicting financial volatility: High-frequency time-series forecasts vis–vis implied volatility. *The Journal of Futures Markets* (11), 1005–1028.

McAleer M, M. M. (2008). Realized volatility: A review. *Econometric Reviews* (27), 10–45.

Merton, R. (1976a). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* (3), 125–144.

Merton, R. C. (1976b). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 125–144.

Mincer, J. and V. Zarnowitz (1969). *The evaluation of economic forecasts*. New York: National Bureau of Economic Research.

Müller, U., M. Dacorogna, R. Dav, R. Olsen, O. Pictet, and J. von Weizsacker (1997). Volatilities of different time resolutions - analysing the dynamics of market components. *ournal of Empirical Finance* (4), 213–239.

Müller, U., M. Dacorogna, R. Dav, O. Pictet, R. Olsen, and J. Ward (1993). Fractals and intrinsic time – a challenge to econometricians. In *39th International AEA Conference on Real Time Econometrics*.

Najzar, K. (2004). *Základy teorie waveletů*. Učební texty Univerzity Karlovy v Praze. Karolinum.

Nielsen, M. and P. Frederiksen (2008). Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. *Journal of Empirical Finance* (15), 265–286.

Percival, D. B. (1995). On estimation of the wavelet variance. *Biometrika 82*, 619–631.

Percival, D. B. and H. Mofjeld (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association 92*(439), 886–880.

Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time series Analysis*. Cambridge University Press.

Protter, P. (1992). *Stochastic integration and differential equations: A new approach*. New York: Springer-Verlag.

Raimondo, M. (1998). Minimax estimation of sharp change points. *The Annals of Statistics* (26), 1379–1397.

Serroukh, A., A. T. Walden, and D. B. Percival (2000). Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. *Journal of the American Statistical Association 95*, 184–196.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science* (9), 227–293.

Subbotin, A. (2008). A multi-horizon scale for volatility. Technical report, Documents de travail du Centre d'Economie de la Sorbonne, Université Panthéon-Sorbonne (Paris 1).

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

Wang, Y. (1995). Jump and sharp cusp detection via wavelets. *Biometrika* (82), 385–397.

Whitcher, B., P. Guttorp, and D. B. Percival (1999). Mathematical background for wavelets estimators for cross covariance and cross correlation. *Tech. Rep. 38, Natl. Res. Cent. for stat. and the Environ.*.

Whitcher, B., P. Guttorp, and D. B. Percival (2000). Wavelet analysis of covariance with application to atmosferic time series. *Journal of Geophysical Research 105*, 941–962.

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

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

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

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

Appendix

11A Technical appendix

11A.1 Wavelets introduction

For a family of Daubechies wavelets $D(L)$, where L is assumed to be even, the scaling filter $m_0(\xi)$ associated with the scaling function ψ has the following form:

$$m_0(\xi) = \left(\frac{1+e^{i\xi}}{L/2}\right)^{L/2} \mathcal{L}(\xi), \tag{11A.1}$$

where $\mathcal{L}(\xi)$ denotes a trigonometric polynomial of order $(L/2)-1$ such that the following holds:

$$|\mathcal{L}(\xi)|^2 \;=\; p_{L/2}\left(\sin^2\left(\frac{\xi}{2}\right)\right) \quad \text{and} \quad \mathcal{L}(0) = 1, \tag{11A.2}$$

where $p_{L/2}$ is a polynomial of order $(L/2)-1$,

$$p_{L/2}(x) = \sum_{n=0}^{(L/2)-1} \binom{(L/2)-1+n}{n} x^n = \sum_{k=0}^{(L/2)-1} \binom{L-1}{k} x^k (1-x)^{(L/2)-1-k}. \tag{11A.3}$$

The Daubechies wavelet $D(L)$ has $L/2$ vanishing moments, i.e.,

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0 \quad k = 0, ..., (L/2)-1 \tag{11A.4}$$

Furthermore, for $(L/2) \geq 2$ the Daubechies wavelets are continuous with a compact support of length L, hence for a wavelet and scaling function we can write:

$$\text{supp}\,\psi = \langle -(L/2)+1, (L/2)\rangle \quad \text{supp}\,\phi = \langle 0, L-1\rangle. \tag{11A.5}$$

For $L = 2$, the length of a Daubechies wavelet is 2, and it corresponds to the Haar wavelet.

11A.2 Daubechies wavelets

The squared gain function for the associated Daubechies family wavelet filter is defined as:

$$\mathcal{H}_f \equiv 2\sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f). \tag{11A.6}$$

The squared gain function for the corresponding Daubechies family scaling filter can be obtained as

$$\mathcal{G}_{1,L}(f) \equiv \mathcal{H}_{1,L}(1/2 - f), \tag{11A.7}$$

hence we get

$$\mathcal{G}_{1,L} \equiv 2\cos^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \sin^{2l}(\pi f). \tag{11A.8}$$

Note that at the first level wavelet decomposition $J = 1$ the wavelet filter is an approximation of an ideal high-pass filter $|f| \in [1/4, 1/2]$, i.e., only the high-frequency part of the signal (or process) passes through the filter. Conversely, the scaling filter is an approximation of an ideal low-pass filter $|f| \in [0, 1/4]$, so only the low-frequency part of the signal passes through.

The Daubechies D(4) DWT coefficients for the scaling filter g and for the wavelet filter h have the following values:

$$g_0 = \frac{1+\sqrt{3}}{4\sqrt{2}},\ g_1 = \frac{3+\sqrt{3}}{4\sqrt{2}},\ g_2 = \frac{3-\sqrt{3}}{4\sqrt{2}},\ g_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}, \tag{11A.9}$$

$$h_0 = \frac{1-\sqrt{3}}{4\sqrt{2}},\ h_1 = \frac{-3+\sqrt{3}}{4\sqrt{2}},\ h_2 = \frac{3+\sqrt{3}}{4\sqrt{2}},\ h_3 = \frac{-1-\sqrt{3}}{4\sqrt{2}}. \tag{11A.10}$$

11A.3 Proof of Theorem 1

The scaling filter $m_0(\xi)$ associated with the scaling function ψ for the Daubechies D(4) wavelet has the following form:

$$m_0(\xi) = \left(\frac{1+e^{i\xi}}{2}\right)^2 \mathcal{L}(\xi), \tag{11A.11}$$

where $\mathcal{L}(\xi)$ denotes a trigonometric polynomial of order 1 such that the following holds:

$$|\mathcal{L}(\xi)|^2 = p_2\left(\sin^2\frac{\xi}{2}\right) \quad \text{and} \quad \mathcal{L}(0) = 1, \tag{11A.12}$$

where p_2 represents a trigonometric polynomial of order 1. From 11A.3 we easily get:

$$p_2(x) = 1 + 2x. \tag{11A.13}$$

Furthermore, there exists a constant C such that

$$|\mathcal{L}(\xi)| \leq 1 + C|\xi|. \tag{11A.14}$$

Hence from 11A.11 it follows that

$$|m_0(\xi)|^2 = \left(\frac{1+e^{i\xi}}{2}\right)^4 p_2\left(\sin^2\left(\frac{\xi}{2}\right)\right) = \frac{1}{4}\left(1+\cos\xi\right)^2\left(2-\cos\xi\right). \tag{11A.15}$$

Further, we get

$$\left|m_0\left(\frac{\xi}{2}+\pi\right)\right|^2 = \frac{1}{4}\left(1-\cos\frac{\xi}{2}\right)^2\left(2+\cos\frac{\xi}{2}\right) = \left|m_0\left(-\frac{\xi}{2}+\pi\right)\right|^2. \tag{11A.16}$$

Since the Daubechies $D(4)$ wavelet is compactly supported, it can be used for construction of the multiresolution analysis with scaling function $\varphi \in L^2(\mathbb{R})$. Then, using the Fourier transform, we define function ψ, which is the wavelet associated with this multiresolution analysis (Mallat, 1998; Najzar, 2004)

$$\hat{\psi}(\xi) = e^{-i\left(\frac{\xi}{2}+\pi\right)}\overline{m_0\left(\frac{\xi}{2}+\pi\right)}\hat{\varphi}\left(\frac{\xi}{2}\right). \tag{11A.17}$$

The squared gain function of the wavelet $\psi(\xi)$ takes the form

$$\left|\hat{\psi}(\xi)\right|^2 = \left|m_0\left(\frac{\xi}{2}+\pi\right)\right|^2\left|\hat{\varphi}\left(\frac{\xi}{2}\right)\right|^2. \tag{11A.18}$$

The Fourier transform of a scale function φ takes the form Igari (1998, p.231)

$$|\hat{\varphi}(\xi)| = \prod_{k=1}^{+\infty}\cos^2(\pi 2^{-k}\xi)|L(2^{-k}\xi)| \le C_1(1+|\xi|)^{-2}|\xi|^{\alpha}, \quad \xi \in \mathbb{Z}, \tag{11A.19}$$

where $\alpha = \log 3/(2\log 2) \approx 0,7925$ and C_1 is a constant independent of ξ (Igari, 1998). Therefore, the following estimate holds:

$$\left|\hat{\psi}(\xi)\right|^2 = \left|m_0\left(\frac{\xi}{2}+\pi\right)\right|^2\left|\hat{\varphi}\left(\frac{\xi}{2}\right)\right|^2 \le C_1^2\left(1+\left|\frac{\xi}{2}\right|\right)^{-4}\left|\frac{\xi}{2}\right|^{2\alpha}(1+C\ |\xi|)^2. \tag{11A.20}$$

From 11A.20 we can easily show that the squared gain function of the wavelet $\psi(\xi)$ is finite, i.e.,

$$\int_0^{+\infty}\left|\hat{\psi}(s)\right|^2\frac{1}{s}\,\mathrm{d}s = \int_0^1 ... + \int_1^{+\infty} ... < +\infty. \tag{11A.21}$$

To complete the proof we have to show that the following equality holds:

$$\left|\hat{\psi}(\xi)\right|^2 = \left|\hat{\psi}(-\xi)\right|^2. \tag{11A.22}$$

It is sufficient to show that a similar relation holds for the squared gain function of the scaling function $\varphi(.)$, i.e.,

$$\left|\hat{\varphi}\left(\frac{\xi}{2}\right)\right|^2 = \left|\hat{\varphi}\left(-\frac{\xi}{2}\right)\right|^2. \tag{11A.23}$$

Then, 11A.23 follows from the formula

$$\hat{\varphi}(\xi) = \prod_{j=1}^{+\infty} m_0\left(\frac{\xi}{2^j}\right), \quad \xi \in \mathbb{R} \tag{11A.24}$$

and from the equality

$$|m_0(\xi)|^2 \;=\; |m_0(-\xi)|^2 \,. \tag{11A.25}$$

Since,

$$\hat{\varphi}\left(\frac{\xi}{2}\right) = \prod_{j=2}^{+\infty} m_0\left(\frac{\xi}{2^j}\right), \quad \xi \in \mathbb{R}. \tag{11A.26}$$

This implies that the extra conditions hold, i.e.,

$$\int_0^{+\infty} \left|\hat{\psi}(s)\right|^2 \frac{1}{s}\,\mathrm{d}s = \int_0^{+\infty} \left|\hat{\psi}(-s)\right|^2 \frac{1}{s}\,\mathrm{d}s = \frac{1}{2}C_\psi. \tag{11A.27}$$

□

11A.4 Proof of Proposition 8

This proof readily follows from the Calderón reconstruction formula, Theorem 1, and Proposition 7, where we specify exact conditions for the Daubechies $D(4)$ wavelet.

□

11A.5 Proof of Proposition 9

For a time series X_i and its discrete Fourier transform $\mathcal{X}_k$, with frequency $f_k \equiv k/N$, we get the following relationship (using Parseval's theorem, see e.g. Percival and Walden, 2000, p.72):

$$\sum_{i=0}^{N-1} |X_i|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2, \tag{11A.28}$$

where $|\mathcal{X}_k|^2/N$ establishes an energy spectrum at frequencies $f_k \equiv k/N$.

For a time series X_i, using Parseval's theorem we can write for the j-th level MODWT wavelet and scaling vectors of coefficients:

$$\|\widetilde{\mathbf{W}}_j\|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |\widetilde{H}_j(k/N)|^2 |\mathcal{X}_k|^2 \tag{11A.29}$$

$$\|\widetilde{\mathbf{V}}_j\|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |\widetilde{G}_j(k/N)|^2 |\mathcal{X}_k|^2. \tag{11A.30}$$

Summation of the vectors yields:

$$\|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_j\|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2 (|\widetilde{H}_j(k/N)|^2 + |\widetilde{G}_j(k/N)|^2). \tag{11A.31}$$

Following Percival and Mofjeld (1997) and Percival and Walden (2000) we can write for any $j \geq 2$

$$\begin{aligned}
\|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_j\|^2 &= \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2 \left(|\widetilde{H}_j(k/N)|^2 + |\widetilde{G}_j(k/N)|^2\right) \\
&= \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2 |\widetilde{H}(2^{j-1}k/N)|^2 \prod_{l=0}^{j-2}\left|\widetilde{G}(2^l k/N)\right|^2 + \prod_{l=0}^{j-1}\left|\widetilde{G}(2^l k/N)\right|^2 \\
&= \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2 \left(|\widetilde{H}(2^{j-1}k/N)|^2 + |\widetilde{G}(2^{j-1}k/N)|^2\right) \prod_{l=0}^{j-2}\left|\widetilde{G}(2^l k/N)\right|^2 \\
&= \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2 \left(\widetilde{\mathcal{H}}(2^{j-1}k/N) + \widetilde{\mathcal{G}}(2^{j-1}k/N)\right) \left|\widetilde{G}_{j-1}(k/N)\right|^2 \\
&= \frac{1}{N}\sum_{k=0}^{N-1} |\mathcal{X}_k|^2 |\widetilde{G}_{j-1}(k/N)|^2 \\
&= \|\widetilde{\mathbf{V}}_{j-1}\|^2, \qquad \text{(11A.32)}
\end{aligned}$$

where we have used the fact that

$$\widetilde{H}\left(2^{j-1}k/N\right)\widetilde{H}\left(2^{j-1}k/N\right) = \left|\widetilde{H}\left(2^{j-1}k/N\right)\right|^2 = \widetilde{\mathcal{H}}\left(2^{j-1}k/N\right), \qquad \text{(11A.33)}$$

$$\left|\widetilde{G}\left(2^{j-1}k/N\right)\right|^2 = \widetilde{\mathcal{G}}\left(2^{j-1}k/N\right) \qquad \text{(11A.34)}$$

and

$$\widetilde{\mathcal{H}}(f) + \widetilde{\mathcal{G}}(f) = 1, \text{ for all } f. \qquad \text{(11A.35)}$$

Using the above result for $j \geq 2$, by induction we obtain

$$\|\widetilde{\mathbf{V}}_1\|^2 = \sum_{j=2}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2, \qquad J_0 \geq 2, \qquad \text{(11A.36)}$$

which in fact says that we can decompose $\|\widetilde{\mathbf{V}}_2\|^2$ further to higher levels. To complete the proof of the energy decomposition using the MODWT wavelet and scaling coefficients we have to show that the following holds (Percival and Walden, 2000):

$$\|\mathbf{X}\|^2 = \|\widetilde{\mathbf{W}}_1\|^2 + \|\widetilde{\mathbf{V}}_1\|^2. \qquad \text{(11A.37)}$$

Using Parseval's theorem we can write the vectors of the MODWT wavelet and scaling

coefficients at the first level in the following form:

$$\begin{aligned} \|\widetilde{\mathbf{W}}_1\|^2 &= \frac{1}{N}\sum_{k=0}^{N-1}|\widetilde{H}_1(k/N)|^2|\mathcal{X}_k|^2 \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\widetilde{\mathcal{H}}(k/N)\mid\mathcal{X}_k|^2, \end{aligned} \tag{11A.38}$$

$$\begin{aligned} \|\widetilde{\mathbf{V}}_1\|^2 &= \frac{1}{N}\sum_{k=0}^{N-1}|\widetilde{G}_1(k/N)|^2|\mathcal{X}_k|^2 \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\widetilde{\mathcal{G}}(k/N)|\mathcal{X}_k|^2. \end{aligned} \tag{11A.39}$$

Summation of the vectors yields

$$\|\widetilde{\mathbf{W}}_1\|^2 + \|\widetilde{\mathbf{V}}_1\|^2 = \frac{1}{N}\sum_{k=0}^{N-1}(\widetilde{\mathcal{H}}(k/N) + \widetilde{\mathcal{G}}(k/N))|\mathcal{X}_k|^2. \tag{11A.40}$$

With $\widetilde{\mathcal{H}}(f) + \widetilde{\mathcal{G}}(f) = 1$ for all f we obtain

$$\|\widetilde{\mathbf{W}}_1\|^2 + \|\widetilde{\mathbf{V}}_1\|^2 = \frac{1}{N}\sum_{k=0}^{N-1}|\mathcal{X}_k|^2 = \|\mathbf{X}\|^2. \tag{11A.41}$$

Hence, using 11A.41 we finally get the energy decomposition of a time series $(X_t)_{t\in[0,N-1]}$ and maximum scale $J \geq 1$

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J_0}\|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2. \tag{11A.42}$$

□

11A.6 Proof of Proposition 10 and Proposition 11

The proof of the unbiasedness and consistency of the $\widehat{RV}_{t,h}^{(WRV)}$ estimator comes readily from the introduced theory. Let us summarize the logic in the following lines.

From definition 2.12 and propositions about the $\widehat{RV}_{t,h}^{(sparse)}$ estimator 3 and 4, we know that $\widehat{RV}_{t,h}^{(sparse)} = \sum_{i=1}^{n} r^2_{t-h+\left(\frac{i}{n}\right)h}$ is an unbiased and consistent estimator of $RV_{t,h} = \int_{t-h}^{t}\sigma_s^2 ds$ on $[t-h,t]$, as $E\left[RV_{t,h}|\mathcal{F}_t\right] = E\left[\widehat{RV}_{t,h}^{(sparse)}|\mathcal{F}_t\right]$ and $\text{plim}_{n\to\infty}\,\widehat{RV}_{t,h}^{(sparse)} = RV_{t,h}$ with uniform convergence in probability.

Moreover, from Proposition 3.24 we know that we can conveniently decompose the energy of the process $r_{t,h}$ over $[t-h,t]$, for $0 \leq h \leq t \leq T$, using the MODWT coefficients (proof in Appendix 11A.5). Using equation 3.26 we can define the intraday returns over $[t-h,t]$ as

$$\|\mathbf{r}\|^2 = \sum_{j=1}^{J_s+1} \|\widetilde{\mathcal{W}}_j\|^2. \tag{11A.43}$$

As $\|\mathbf{r}\|^2 = \sum_{i=1}^{n} r^2_{t-h+\left(\frac{i}{n}\right)h}$, we can directly apply this decomposition to the realized variance estimator:

$$\widehat{RV}^{(sparse)}_{t,h} = \sum_{i=1}^{n} r^2_{t-h+\left(\frac{i}{n}\right)h} = \sum_{j=1}^{J_s+1} \sum_{k=1}^{n} \widetilde{\mathcal{W}}^2_{j,t-h+\frac{k}{n}h} = \widehat{RV}^{(WRV)}_{t,h}. \tag{11A.44}$$

Thus $E\left[RV_{t,h}|\mathcal{F}_t\right] = E\left[\widehat{RV}^{(sparse)}_{t,h}|\mathcal{F}_t\right] = E\left[\widehat{RV}^{(WRV)}_{t,h}|\mathcal{F}_t\right]$.

Moreover, based on the wavelet representation theorem in Proposition 8 proved in Appendix 11A.4, $\text{plim}_{n\to\infty} \widehat{RV}^{(sparse)}_{t,h} = \text{plim}_{n\to\infty} \widehat{RV}^{(WRV)}_{t,h} = \int_{t-h}^{t} \sigma_s^2 ds$ and $\widehat{RV}^{(WRV)}_{t,h}$ provides a consistent estimator with increasing sampling frequency $n \to \infty$.

□

11A.7 Proof of Proposition 14 and Proposition 15

All of the theory has been actually proved, we just bring it together in a new estimator. Thus we describe the logic of the proof here.

Consider a log-price process $(p_t)_{t\in[0,T]}$ that is contaminated with noise, i.e., $y_t = p_t + \epsilon_t$, where $(y_t)_{t\in[0,T]}$ is the observed log-price process. Moreover, p_t follows a jump-diffusion process

$$dp_t = \mu_t dt + \sigma_t dW_t + \xi_t dq_t, \tag{11A.45}$$

where q is a constant-intensity Poisson process with the same magnitude as process 2.8. Process 2.10 can be characterized as a Brownian semi-martingale with finite jump activity, and is also a special case of decomposition 1. Its quadratic return variation over $[t-h,t]$, for $0 \leq h \leq t \leq T$, is

$$QV_{t,h} = \underbrace{\int_{t-h}^{t} \sigma_s^2 ds}_{IV_{t,h}} + \underbrace{\sum_{t-h\leq s\leq t} J_s^2}_{\text{Jump Variation}}. \tag{11A.46}$$

Fan and Wang (2007) show that integrated variance and jump variation can be separated using the wavelet estimator defined by Definition 10. The authors prove that the $\widehat{WJV}$ estimator can consistently estimate the jump variation part, and that the jump-adjusted price $y^{(J)}_{t,h} = y_{t,h} - \widehat{WJV}$ converges in probability to its theoretical, continuous counterpart at a convergence rate of $n^{-1/4}$.

We have proved that the $\widehat{RV}_{t,h}^{(WRV)}$ estimator is able to estimate integrated variance consistently (Proposition 10 and Proposition 11 proved in Appendix 11A.6). Thus, the realized variance of process $y_{t,h}^{(J)}$ can be estimated using the energy decomposition of process:

$$\sum_{i=1}^{n}\left(y_{t-h+\left(\frac{i}{n}\right)h}^{(J)}\right)^2=\sum_{j=1}^{J_s+1}\sum_{k=1}^{n}\widetilde{\mathcal{W}}_{j,t-h+\frac{k}{n}h}^2, \tag{11A.47}$$

where $\widetilde{\mathcal{W}}_{j,t-h+\frac{k}{n}h}$ are wavelet coefficients estimated using the MODWT (the proof is exactly the same as the one in Appendix 11A.6).

Using Zhang et al. (2005)'s TSRV estimator, we consistently estimate the integrated variance of p_t from the noisy observed data y_t.

$$\widehat{RV}_{t,h}^{(tsrv)}=\underbrace{\widehat{RV}_{t,h}^{(average)}}_{\text{slow time scale}}-\frac{\bar{n}}{n}\underbrace{\widehat{RV}_{t,h}^{(all)}}_{\text{fast time scale}}, \tag{11A.48}$$

where $\widehat{RV}_{t,h}^{(average)}$ is the average of the RV estimators on a grid (see Section 2.3 for a full explanation). Finally, replacing the realized variance with the decomposition using wavelet realized variance on $y_{t,h}^{(J)}$ from Equation 11A.47 and putting it into the TSRV, we get

$$\widehat{RV}_{t,h}^{(JWTSRV)}=\widehat{RV}_{t,h}^{(WRV,J)}-\frac{\bar{n}}{n}\widehat{RV}_{t,h}^{(all,J)}, \tag{11A.49}$$

where $\widehat{RV}_{t,h}^{(WRV,J)}=\frac{1}{G}\sum_{g=1}^{G}\sum_{j=1}^{J_s+1}\sum_{k=1}^{n}\widetilde{\mathcal{W}}_{j,t-h+\frac{k}{n}h}^2$ obtained from the wavelet coefficient estimates on a grid of size $\bar{n}=n/G$ on the jump-adjusted observed data, $y_{t,h}^{(J)}=y_{t,h}-\sum_{l=1}^{N_t}J_l$.

We have shown that $\widehat{RV}_{t,h}^{(WRV,J)}$ converges to the integrated variation of process $y_{t,h}$, and Zhang et al. (2005) provide the proof for the TSRV.

□

11A.8 Proof of Proposition 19

To prove Proposition 19 we write the covariance of the MODWT wavelet coefficients in the form:

$$\gamma_{(XY)j}=\int_{-1/2}^{1/2}\widetilde{\mathcal{H}}_j(f)S_{(XY)}(f)df, \tag{11A.50}$$

where $\widetilde{\mathcal{H}}_j(f)$ denotes the squared gain function of the wavelet MODWT filter ($\tilde{h}_j$). The covariance of the scaling coefficients at level J (the last level of decomposition):

$$Cov(\widetilde{V}_{(X)J,i},\,\widetilde{V}_{(Y)J,i})=\int_{-1/2}^{1/2}\widetilde{\mathcal{G}}_J(f)S_{(XY)}(f)df \tag{11A.51}$$

where $\widetilde{\mathcal{G}}_J(f)$ denotes the squared gain function of the scaling MODWT filter $\tilde{g}_J$, such that $\widetilde{\mathcal{G}}_J(f) \equiv \prod_{l=0}^{J-1} \widetilde{\mathcal{G}}(2^l f)$. While $\widetilde{\mathcal{H}}(f) + \widetilde{\mathcal{G}}(f) = 1$ (Percival and Walden, 2000), the covariance decomposed by wavelets only at the first level ($J = 1$) is obtained as the sum of the wavelet and scaling MODWT coefficients' covariances,

$$Cov(X_i, Y_i) = \int_{-1/2}^{1/2} \left(\widetilde{\mathcal{H}}(f) + \widetilde{\mathcal{G}}(f)\right) S_{(XY)}(f)df = Cov(\widetilde{V}_{(X)1,i},\, \widetilde{V}_{(Y)1,i}) + \gamma_{(XY)1}. \tag{11A.52}$$

Further, we assume that it holds also for level $J-1$:

$$Cov(X_i, Y_i) = Cov(\widetilde{V}_{(X)J-1,i},\, \widetilde{V}_{(Y)J-1,i}) + \sum_{j=1}^{J-1} \gamma_{(XY)j}. \tag{11A.53}$$

Following Whitcher et al. (1999) we have

$$\begin{aligned}
Cov(\widetilde{V}_{(X)J-1,i},\, \widetilde{V}_{(Y)J-1,i}) &= \int_{-1/2}^{1/2} \widetilde{\mathcal{G}}_{J-1}(f) S_{(XY)}(f)df \\
&= \int_{-1/2}^{1/2} \left[\prod_{l=0}^{J-2} \widetilde{\mathcal{G}}(2^l f)\right] S_{(XY)}(f)df \\
&= \int_{-1/2}^{1/2} \left[\widetilde{\mathcal{G}}(2^{J-1} f) + \widetilde{\mathcal{H}}(2^{J-1} f)\right] \left[\prod_{l=0}^{J-2} \widetilde{\mathcal{G}}(2^l f)\right] S_{(XY)}(f)df \\
&= \int_{-1/2}^{1/2} \left[\widetilde{\mathcal{G}}_J(f) + \widetilde{\mathcal{H}}_J(f)\right] S_{(XY)}(f)df \\
&= Cov(\widetilde{V}_{(X)J,t},\, \widetilde{V}_{(Y)J,t}) + \gamma_{(XY)J}
\end{aligned} \tag{11A.54}$$

which proves, by induction, the wavelet covariance decomposition of (X_i, Y_i) for a finite number of scales J.

Further, we proof that as $J \to \infty$ then the covariance between between scaling coefficients goes to zero, therefore the covariance of X_t and Y_t depends only on the covariance of the wavelet coefficients $\gamma_{(XY)}$. Using the result (11A.54) we can write:

$$\begin{aligned}
Cov(\widetilde{V}_{(X)J-1,t},\, \widetilde{V}_{(Y)J-1,t}) &= Cov(\widetilde{V}_{(X)J,t},\, \widetilde{V}_{(Y)J,t}) + \gamma_{(XY)J} && \text{(11A.55)} \\
Cov(\widetilde{V}_{(X)J,t},\, \widetilde{V}_{(Y)J,t}) &= Cov(\widetilde{V}_{(X)J+1,t},\, \widetilde{V}_{(Y)J+1,t}) + \gamma_{(XY)J+1} && \text{(11A.56)} \\
\vdots &= \vdots && \text{(11A.57)} \\
Cov(\widetilde{V}_{(X)J+n-1,t},\, \widetilde{V}_{(Y)J+n-1,t}) &= Cov(\widetilde{V}_{(X)J+n,t},\, \widetilde{V}_{(Y)J+n,t}) + \gamma_{(XY)J+n} && \text{(11A.58)}
\end{aligned}$$

by summation we obtain

$$Cov(\widetilde{V}_{(X)J-1,t},\, \widetilde{V}_{(Y)J-1,t}) = Cov(\widetilde{V}_{(X)J+n,t},\, \widetilde{V}_{(Y)J+n,t}) + \sum_{j=0}^{n} \gamma_{(XY)J+j} \tag{11A.59}$$

for the part consisting of the wavelet coefficient covariance we have

$$\sum_{j=0}^{n} \gamma_{(XY)J+j} = Cov(\widetilde{V}_{(X)J-1,t},\, \widetilde{V}_{(Y)J-1,t}) - Cov(\widetilde{V}_{(X)J+n,t},\, \widetilde{V}_{(Y)J+n,t}). \tag{11A.60}$$

Let denote s_m as a sum of wavelet coefficients covariances up to a scale m, i.e.,

$$s_m = \sum_{j=0}^{m} \gamma_{(XY)j}. \tag{11A.61}$$

Then for any positive integer m, such that $m > J$ we have:

$$s_m = \sum_{j=0}^{J-1} \gamma_{(XY)j} + \sum_{j=0}^{m-J} \gamma_{(XY)J+j} \tag{11A.62}$$

$$= Cov(\widetilde{V}_{(X)J-1,t},\, \widetilde{V}_{(Y)J-1,t}) - Cov(\widetilde{V}_{(X)m,t},\, \widetilde{V}_{(Y)m,t}) + \sum_{j=0}^{J-1} \gamma_{(XY)j} \tag{11A.63}$$

Hence, for any two positive integers $m_1, m_2 > J$ we can write

$$|s_{m_1} - s_{m_2}| = |Cov(\widetilde{V}_{(X)m_1,t},\, \widetilde{V}_{(Y)m_1,t}) - Cov(\widetilde{V}_{(X)m_2,t},\, \widetilde{V}_{(Y)m_2,t})| \tag{11A.64}$$

Following result of Whitcher et al. (2000) (lemma 1, page 2): for any $\epsilon > 0$ there exists J_ϵ such that for a positive integer $m > J_\epsilon$ holds:

$$|Cov(\widetilde{V}_{(X)m,t},\, \widetilde{V}_{(Y)m,t})| < \epsilon. \tag{11A.65}$$

Using the result (11A.65) then for any $\epsilon > 0$ exists J_ϵ such that for positive integers $m_1, m_2 > J_\epsilon$ we obtain

$$|s_{m_1} - s_{m_2}| \leq 2\epsilon, \tag{11A.66}$$

so the sequence $\{s_m\}$ is Cauchy and has a limit:

$$\lim_{m\to\infty} s_m = \sum_{j=0}^{\infty} \gamma_{(XY)j} = Cov(\widetilde{V}_{(X)J-1,t},\, \widetilde{V}_{(Y)J-1,t}) + \sum_{j=0}^{J-1} \gamma_{(XY)j} \tag{11A.67}$$

then it follows

$$\sum_{j=J}^{\infty} \gamma_{(XY)j} = Cov(\widetilde{V}_{(X)J-1,t},\, \widetilde{V}_{(Y)J-1,t}) \tag{11A.68}$$

which implies (c.f. 11A.53)

$$Cov(X_t, Y_t) = \sum_{j=0}^{\infty} \gamma_{(XY)j}, \tag{11A.69}$$

which completes the proof. □

11A.9 Proof of Proposition 20 and Proposition 21

The realized covariance for the l-th and q-th asset return from an m-dimensional vector $\mathbf{r_{t,h}}$ over $[t-h,t]$, for $0 \leq h \leq t \leq T$, can be computed using Definition 8.7

$$\widehat{RC}_{(l,q)t,h} = \sum_{i=1}^{n} r_{(l)t-h+\left(\frac{i}{n}\right)h} r_{(q)t-h+\left(\frac{i}{n}\right)h}, \tag{11A.70}$$

while according to Propositions 16 and 17, 11A.70 is an unbiased and consistent estimator of realized covariance.

For a particular level j we define the realized wavelet covariance over $[t-h,t]$, for $0 \leq h \leq t \leq T$, as the sample covariance between the MODWT wavelet coefficients at level j, hence we have:

$$WRCov_{(l,q)t,h,j} = \sum_{t=L_j-1}^{N-1} \widetilde{W}_{(l)j,t-h+\frac{k}{n}h} \widetilde{W}_{(q)j,t-h+\frac{k}{n}h}, \tag{11A.71}$$

where we use the $M_j = N - L_j + 1 > 0$ wavelet coefficients at the j-th level for both processes which are unaffected by the boundary conditions.

In case $J \to \infty$, the realized covariance $\widehat{RC}_{(l,q)t,h}$ is simply the sum of all wavelet realized covariances:

$$RC_{(l,q)t,h} = \sum_{j=1}^{\infty} WRCov_{(l,q)j,t-h+\frac{k}{n}h}\,. \tag{11A.72}$$

Since we have datasets of a finite length, the contribution of the realized covariation of the scaling coefficients is still relatively high (we cannot ignore it), so we use a similar approach as with the wavelet covariance in 9.2, i.e., the realized wavelet covariance has two parts, the first one being the realized covariance of the MODWT scaling coefficients $\widetilde{V}_{(l)t,h,J}$ and $\widetilde{V}_{(q)t,h,J}$ at the maximum level of decomposition J, and the second one being the sum of the realized wavelet covariances up to the maximum level J. Thus, for the maximum level of decomposition $J \leq log_2(N)$ we have:

$$\begin{aligned} RCOV_{(l,q)t,h} &= \sum_{t=L_j-1}^{N-1} \widetilde{V}_{(l)J,t-h+\frac{k}{n}h} \widetilde{V}_{(q)J,t-h+\frac{k}{n}h} + \sum_{j=1}^{J} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{(l)j,t-h+\frac{k}{n}h} \widetilde{W}_{(q)j,t-h+\frac{k}{n}h} \\ &= \sum_{t=L_j-1}^{N-1} \widetilde{V}_{(l)J,t-h+\frac{k}{n}h} \widetilde{V}_{(q)J,t-h+\frac{k}{n}h} + \sum_{j=1}^{J} WRCov_{(lq)j,t-h+\frac{k}{n}h} \end{aligned} \tag{11A.73}$$

where for a specific level j we use only the $M_j = N - L_j + 1 > 0$ MODWT wavelet or scaling coefficients unaffected by the boundary conditions.

Denoting by $\widetilde{\mathcal{W}}_{(q)j,k}$ the MODWT coefficients defined in 3.26 on scales $j = 1, ..., J_s+1$, which include both parts of the wavelet covariance, we have

$$\widehat{RC}^{(WRC)}_{(l,q)t,h} = \sum_{j=1}^{J_s+1} \sum_{k=1}^{n} \widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h} \widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}, \tag{11A.74}$$

where n is the number of intraday observations and J_s is the number of scales considered.

Finally, from the presented theory we know that the $\widehat{RC}^{(WRC)}_{(l,q)t,h}$ estimator will converge to the integrated covariation as

$$\widehat{RC}_{(l,q)t,h} = \sum_{i=1}^{n} r_{(l)t-h+\left(\frac{i}{n}\right)h} r_{(q)t-h+\left(\frac{i}{n}\right)h} = \sum_{j=1}^{J_s+1} \sum_{k=1}^{n} \widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h} \widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h} = \widehat{RC}^{(WRC)}_{(l,q)t,h}. \tag{11A.75}$$

Thus, it is an unbiased estimator of integrated covariation:

$$E\left[RC_{(l,q)t,h}|\mathcal{F}_t\right] = E\left[\widehat{RC}_{(l,q)t,h}|\mathcal{F}_t\right] = E\left[\widehat{RC}^{(WRC)}_{(l,q)t,h}|\mathcal{F}_t\right]. \tag{11A.76}$$

As the wavelet-based covariance estimator is in fact the sample covariance without $1/M_j$ adjustment, $\text{plim}_{n\to\infty} \widehat{RC}_{(l,q)t,h} = \text{plim}_{n\to\infty} \widehat{RC}^{(WRC)}_{(l,q)t,h} = \int_{t-h}^{t} \Sigma_{(l,q)s} ds$ and $\widehat{RC}^{(WRC)}_{(l,q)t,h}$ provides a consistent estimator with increasing sampling frequency $n \to \infty$.

□

11A.10 Proof of Proposition 23 and Proposition 24

The construction of this proof is very similar to the univariate JWTSRV 11A.10. As all the theory has been introduced, we just use it to produce a new estimator which combines these ideas. We summarize the logic here.

Generalizing the idea of Definition 10 giving us the $\widehat{WJV}$ estimator, which can consistently estimate the jump variation part, we simply introduce the idea of an $\widehat{MWJC}$ estimator (Definition 19). Jumps estimated using this estimator enable us to work with jump-adjusted data, $\mathbf{y_{t,h}}^{(J)} = \mathbf{y}_{t,h} - \widehat{\mathbf{MWJC}}$, in a multivariate setting.

We have shown that $\widehat{RC}^{(WRC)}_{(l,q)t,h}$ is able to estimate integrated covariance consistently (Proposition 20 and Proposition 21 proved in Appendix 11A.9).

Finally, we plug the wavelet decomposition of the jump-adjusted vector into the TSCV estimator (Zhang, 2011), which is able to estimate realized covariance in the presence of noise.

□

11B Figures and Tables to Part I

Table 11.1: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 1,000 simulations of jump-diffusion model with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV - 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	-1.40 (0.73)	-3.15 (0.98)	-7.85 (0.47)	-0.77 (0.02)	-18.72 (2.49)	-7.97 (0.47)	-0.88 (0.02)
ϵ_2	97.75 (0.85)	95.62 (1.14)	-6.82 (0.45)	0.66 (0.52)	-9.22 (2.70)	-5.77 (0.46)	1.77 (0.53)
ϵ_3	399.80 (2.09)	416.37 (2.91)	-5.72 (0.40)	-0.12 (0.93)	16.53 (2.86)	-0.44 (0.43)	5.41 (0.97)
ϵ_4	890.46 (5.77)	954.81 (7.99)	-5.17 (0.36)	5.14 (1.39)	50.45 (2.61)	8.53 (0.52)	19.19 (1.56)
				One Jump			
ϵ_1	249.29 (20.25)	54.08 (2.32)	235.44 (18.93)	245.27 (18.84)	216.76 (22.03)	-6.18 (0.46)	0.13 (0.02)
ϵ_2	384.73 (26.80)	163.39 (3.08)	271.40 (24.19)	281.52 (23.10)	268.66 (27.64)	-1.35 (0.55)	3.42 (0.53)
ϵ_3	638.96 (18.70)	501.08 (4.97)	233.73 (17.03)	240.51 (17.90)	249.05 (20.97)	22.54 (0.62)	23.96 (1.15)
ϵ_4	1159.60 (29.31)	1051.00 (10.45)	261.24 (23.31)	270.92 (25.09)	334.67 (28.69)	58.81 (1.38)	68.92 (2.08)
				Two Jumps			
ϵ_1	473.69 (35.74)	114.12 (3.33)	459.90 (34.10)	475.33 (34.32)	447.28 (42.59)	-3.85 (0.40)	0.08 (0.01)
ϵ_2	595.94 (38.59)	236.80 (5.24)	485.90 (37.34)	495.89 (36.30)	470.30 (44.41)	-0.35 (0.45)	7.57 (0.57)
ϵ_3	917.26 (44.97)	575.46 (7.68)	506.49 (40.60)	523.27 (40.39)	528.93 (48.62)	39.20 (0.89)	47.26 (1.30)
ϵ_4	1374.30 (43.59)	1136.10 (13.59)	470.19 (34.92)	479.72 (36.24)	529.84 (45.45)	101.28 (2.25)	110.18 (2.69)
				Three Jumps			
ϵ_1	738.35 (58.31)	179.67 (5.74)	725.47 (56.65)	734.61 (53.77)	705.84 (67.91)	-4.67 (0.46)	0.46 (0.05)
ϵ_2	886.53 (70.19)	309.18 (6.53)	766.26 (63.93)	786.93 (63.10)	762.00 (83.57)	3.67 (0.50)	9.21 (0.54)
ϵ_3	1148.90 (63.54)	677.20 (11.39)	733.18 (57.31)	751.41 (61.15)	739.00 (69.94)	58.86 (1.07)	65.21 (1.28)
ϵ_4	1647.60 (71.55)	1252.80 (20.28)	743.86 (61.90)	765.78 (61.82)	795.67 (74.28)	155.70 (2.87)	164.84 (3.62)

Table 11.2: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 1,000 simulations of fractional stochastic volatility model with Hurst parameter $H = 0.5$ with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	17.76 (10.21)	-1.80 (13.23)	-17.01 (6.71)	-0.22 (0.23)	-60.64 (39.80)	-17.35 (6.71)	-0.57 (0.23)
ϵ_2	99.52 (10.82)	75.14 (14.20)	-23.83 (6.51)	2.32 (0.88)	-68.05 (39.44)	-22.92 (6.50)	3.74 (0.89)
ϵ_3	390.82 (13.85)	363.10 (17.60)	-22.83 (6.20)	4.98 (1.56)	-31.60 (41.33)	-15.51 (6.27)	11.14 (1.56)
ϵ_4	883.17 (21.89)	876.44 (28.62)	-41.16 (6.96)	0.97 (2.30)	-20.55 (42.94)	-28.56 (7.21)	13.87 (2.42)
				One Jump			
ϵ_1	251.63 (31.92)	91.82 (18.31)	207.41 (26.98)	252.77 (18.35)	159.10 (72.40)	-43.10 (7.05)	-1.11 (0.24)
ϵ_2	355.84 (33.53)	197.71 (19.68)	230.38 (26.77)	255.10 (17.83)	195.33 (67.11)	-20.75 (6.78)	8.81 (0.79)
ϵ_3	685.21 (38.50)	527.69 (25.34)	234.30 (28.12)	265.51 (23.09)	204.34 (68.12)	-2.04 (7.45)	20.55 (1.82)
ϵ_4	1153.70 (44.46)	1031.30 (37.67)	221.03 (26.04)	236.11 (21.12)	272.87 (77.80)	43.68 (8.11)	60.38 (3.03)
				Two Jumps			
ϵ_1	525.30 (48.88)	215.33 (21.51)	481.06 (42.64)	504.66 (34.58)	425.56 (91.20)	-22.70 (6.72)	-0.42 (0.23)
ϵ_2	641.30 (66.57)	345.94 (26.99)	496.61 (58.20)	518.55 (46.09)	465.15 (119.16)	-8.19 (7.12)	10.35 (0.92)
ϵ_3	979.96 (70.45)	651.90 (33.13)	534.67 (60.63)	561.12 (51.42)	521.03 (109.75)	14.74 (7.16)	42.87 (1.83)
ϵ_4	1425.70 (63.36)	1172.80 (42.42)	493.45 (47.37)	519.68 (41.74)	505.05 (104.60)	99.35 (9.70)	116.18 (3.89)
				Three Jumps			
ϵ_1	822.02 (85.39)	363.75 (31.43)	773.68 (77.55)	801.23 (66.61)	710.39 (138.28)	-23.16 (6.26)	0.09 (0.24)
ϵ_2	895.39 (79.02)	466.47 (31.31)	748.77 (68.16)	778.92 (59.09)	686.37 (125.24)	-9.88 (6.89)	6.35 (0.88)
ϵ_3	1141.80 (77.50)	753.44 (34.73)	691.23 (62.64)	732.50 (58.65)	689.81 (124.22)	49.86 (7.61)	69.87 (1.97)
ϵ_4	1689.80 (96.05)	1311.30 (50.06)	747.42 (73.52)	785.70 (64.97)	778.19 (133.15)	139.96 (9.98)	170.82 (4.76)

Table 11.3: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 1,000 simulations of fractional stochastic volatility model with Hurst parameter $H = 0.7$ with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	33.41 (10.21)	14.61 (13.38)	-4.71 (6.69)	0.31 (0.23)	-11.46 (41.49)	-5.17 (6.70)	-0.07 (0.23)
ϵ_2	111.39 (11.39)	84.94 (14.79)	-22.31 (6.39)	-6.03 (0.82)	-51.38 (38.70)	-21.12 (6.38)	-4.97 (0.82)
ϵ_3	398.87 (14.70)	365.88 (17.62)	-25.98 (6.06)	3.76 (1.51)	-30.57 (38.81)	-22.03 (6.12)	7.79 (1.53)
ϵ_4	909.86 (21.94)	893.59 (28.02)	-19.56 (7.07)	-0.03 (2.15)	44.98 (47.32)	-6.23 (7.34)	13.11 (2.29)
				One Jump			
ϵ_1	228.09 (32.71)	68.20 (19.47)	187.05 (26.87)	229.85 (19.44)	111.08 (64.60)	-49.22 (6.63)	-1.70 (0.24)
ϵ_2	370.74 (35.58)	192.60 (19.42)	229.98 (27.03)	248.81 (19.26)	214.97 (67.50)	-16.03 (6.84)	-0.73 (0.82)
ϵ_3	627.39 (39.13)	481.90 (24.90)	213.87 (29.65)	249.53 (25.19)	217.09 (74.34)	-5.98 (6.99)	23.71 (1.55)
ϵ_4	1173.90 (50.73)	1031.70 (36.92)	239.71 (30.68)	272.73 (23.04)	247.88 (71.83)	32.42 (8.07)	65.83 (3.02)
				Two Jumps			
ϵ_1	510.14 (56.62)	195.47 (22.69)	466.20 (49.21)	497.72 (38.33)	412.18 (95.05)	-31.37 (6.80)	-3.37 (0.23)
ϵ_2	613.93 (55.04)	318.39 (22.15)	471.20 (45.85)	497.28 (39.56)	440.70 (97.29)	-12.73 (6.51)	6.03 (0.81)
ϵ_3	909.74 (62.27)	616.91 (29.79)	477.39 (47.35)	519.41 (43.79)	468.19 (100.49)	18.29 (7.96)	45.83 (1.84)
ϵ_4	1396.40 (70.63)	1152.10 (42.36)	458.80 (49.48)	476.60 (40.01)	526.99 (114.35)	92.28 (9.40)	115.52 (3.96)
				Three Jumps			
ϵ_1	801.12 (78.94)	363.07 (29.03)	757.12 (70.00)	783.98 (62.78)	701.53 (134.25)	-20.43 (6.37)	-2.08 (0.24)
ϵ_2	860.96 (79.42)	444.32 (28.15)	724.42 (70.73)	746.75 (58.48)	646.22 (133.41)	-13.74 (6.32)	7.55 (0.84)
ϵ_3	1129.90 (84.57)	735.46 (37.73)	683.91 (67.18)	736.57 (58.84)	741.26 (129.72)	19.33 (7.28)	62.49 (1.91)
ϵ_4	1645.00 (89.77)	1295.20 (49.66)	718.02 (67.39)	749.09 (58.13)	705.71 (137.87)	135.97 (11.07)	155.57 (4.86)

Table 11.4: Bias (variance in parenthesis) $\times 10^4$ of all estimators from 1,000 simulations of fractional stochastic volatility model with Hurst parameter $H = 0.9$ with $\epsilon_1 = 0$, $\epsilon_2 = 0.0005$, $\epsilon_3 = 0.001$, $\epsilon_4 = 0.0015$. RV – 5 min. realized variance estimator, BV – 5 min. bipower variation estimator, TSRV – 5 min. two-scale realized volatility, JWTSRV – 5 min. jump wavelet two-scale realized variance. TSRV and JWTSRV* are minimum variance estimators (see 11A.48), and RK is Realized Kernel.*

	RV	BV	TSRV	TSRV*	RK	JWTSRV	JWTSRV*
				No Jumps			
ϵ_1	-9.17 (9.90)	-34.72 (12.98)	-38.81 (6.75)	-2.12 (0.23)	-76.15 (45.19)	-39.14 (6.74)	-2.59 (0.23)
ϵ_2	99.25 (11.88)	74.23 (15.02)	-30.66 (7.35)	0.38 (0.94)	-32.12 (43.09)	-29.23 (7.33)	1.60 (0.94)
ϵ_3	406.04 (15.51)	398.56 (20.27)	-30.26 (6.44)	-3.81 (1.43)	-28.34 (44.00)	-24.94 (6.52)	1.74 (1.45)
ϵ_4	892.51 (22.23)	889.12 (29.42)	-8.04 (6.89)	-1.53 (2.06)	38.46 (42.18)	7.30 (7.23)	12.12 (2.19)
				One Jump			
ϵ_1	291.99 (36.13)	114.01 (18.87)	248.20 (31.01)	266.01 (21.63)	227.66 (72.67)	-20.55 (7.05)	-1.77 (0.24)
ϵ_2	368.29 (32.88)	206.51 (19.30)	232.41 (26.22)	246.59 (20.31)	197.65 (63.26)	-11.03 (6.82)	1.14 (0.89)
ϵ_3	654.00 (36.65)	506.79 (24.24)	212.51 (27.24)	247.26 (20.99)	198.08 (68.56)	-6.69 (7.47)	24.13 (1.68)
ϵ_4	1168.90 (52.26)	1016.00 (38.88)	231.20 (30.97)	264.79 (23.35)	231.02 (82.51)	27.81 (8.22)	61.19 (3.28)
				Two Jumps			
ϵ_1	527.85 (58.34)	224.90 (23.98)	501.70 (51.47)	529.14 (44.35)	465.69 (108.41)	-14.98 (6.46)	1.79 (0.23)
ϵ_2	619.94 (58.42)	330.72 (28.42)	469.63 (48.08)	501.01 (37.23)	421.34 (95.87)	-25.11 (7.10)	5.70 (0.88)
ϵ_3	935.88 (60.95)	643.98 (30.22)	492.43 (46.43)	519.10 (41.37)	490.41 (100.82)	21.63 (7.80)	39.74 (1.85)
ϵ_4	1414.20 (76.88)	1170.70 (48.61)	490.20 (51.35)	515.42 (43.60)	486.79 (101.57)	86.89 (8.58)	111.84 (3.83)
				Three Jumps			
ϵ_1	771.89 (83.15)	343.66 (30.73)	738.00 (73.08)	774.26 (61.05)	633.39 (131.14)	-35.77 (7.12)	-0.66 (0.24)
ϵ_2	894.48 (92.02)	470.48 (35.26)	747.37 (77.69)	780.32 (66.10)	693.54 (141.13)	-15.18 (6.87)	9.00 (0.90)
ϵ_3	1114.00 (86.00)	728.42 (39.21)	687.64 (69.19)	759.26 (61.42)	648.02 (130.93)	24.56 (8.75)	64.78 (1.89)
ϵ_4	1625.20 (92.66)	1268.00 (46.45)	687.13 (69.80)	725.66 (59.25)	761.51 (135.40)	132.99 (9.97)	167.03 (4.73)

Table 11.5: The table summarizes the daily distributions of $JWTRSV_j$, $j = 1, \ldots, 5$, components of integrated volatility IV_t for GBP, CHF and EUR futures. JWTSRV is estimated using 5 minutes for the slow time scale. The sample period extends from January 5, 2007 through November 17, 2010, accounting for a total of 944 observations.

	Mean	St.dev.	Skew.	Kurt.	LB(5)	LB(10)	LB(15)	LB(20)
				GBP futures				
$j = 1$	0.000	0.000	3.398	19.351	3112.597	5778.592	8192.288	10187.605
$j = 2$	0.000	0.000	3.715	24.579	2858.212	5198.666	7366.679	9174.211
$j = 3$	0.000	0.000	4.201	30.508	2694.620	4868.030	6822.560	8255.797
$j = 4$	0.000	0.000	4.289	31.412	2456.556	4543.761	6254.518	7693.816
$j = 5$	0.000	0.000	4.210	28.236	1951.044	3399.126	4773.515	6095.179
				CHF futures				
$j = 1$	0.000	0.000	3.143	19.111	2406.406	4246.279	5718.564	7032.126
$j = 2$	0.000	0.000	3.299	19.451	2172.393	3690.242	4893.696	5998.477
$j = 3$	0.000	0.000	3.129	17.471	1774.281	2933.702	3953.451	4905.425
$j = 4$	0.000	0.000	3.492	20.947	1325.844	2228.800	2908.499	3522.840
$j = 5$	0.000	0.000	5.508	50.955	720.522	1353.422	1701.367	2032.318
				EUR futures				
$j = 1$	0.000	0.000	2.756	12.984	3184.932	5798.825	8163.729	10248.276
$j = 2$	0.000	0.000	2.891	14.252	2959.948	5293.381	7454.926	9415.475
$j = 3$	0.000	0.000	3.253	16.884	2734.376	4914.557	6885.595	8549.980
$j = 4$	0.000	0.000	3.736	23.860	2146.412	4027.802	5686.769	7139.671
$j = 5$	0.000	0.000	3.820	24.363	1839.767	3391.492	4800.807	6097.753

11C Figures and Tables to Part II

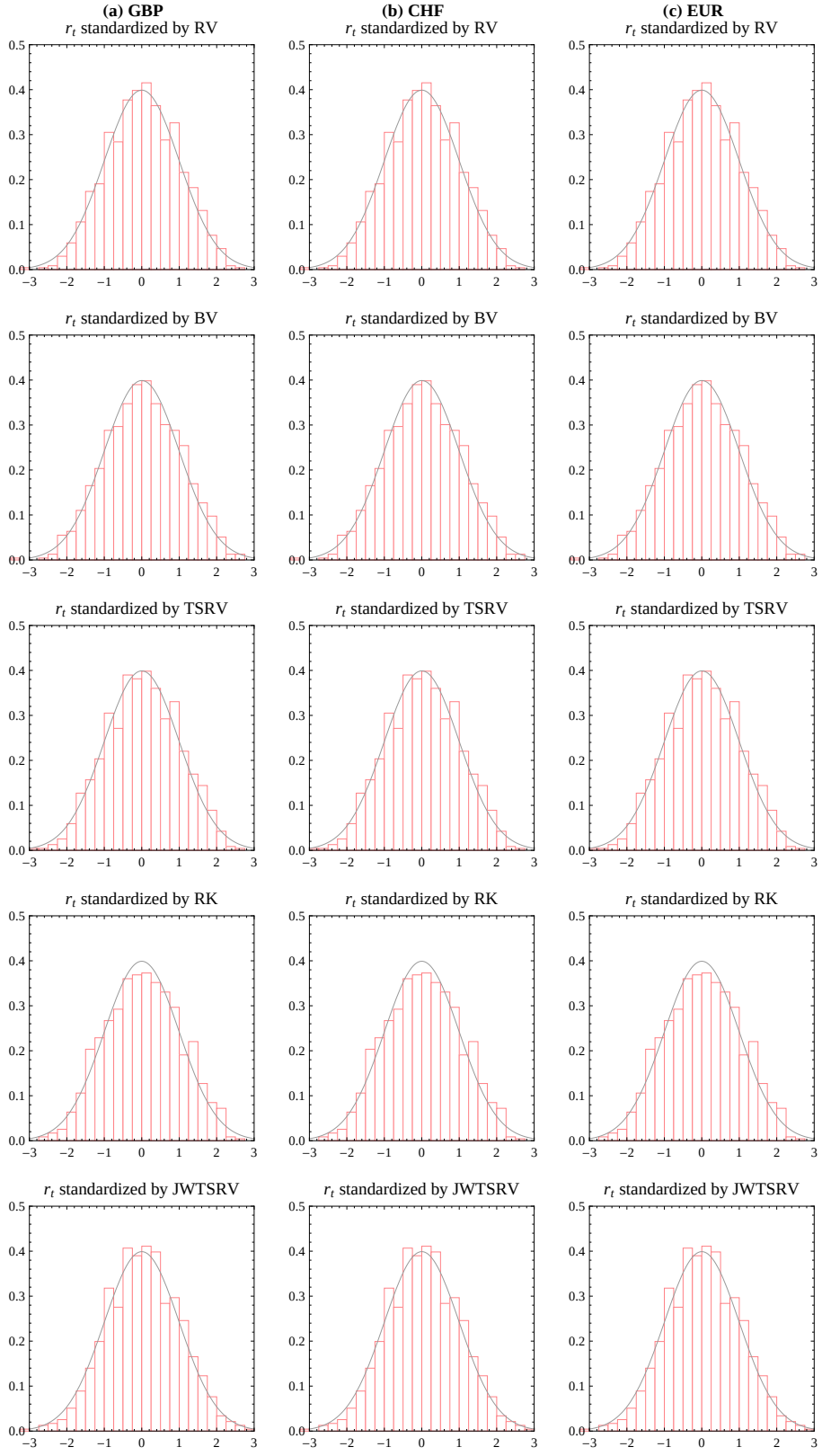


Figure 11.1: Histograms of normalized daily log-returns r_t by RV, BV, TSRV, RK and JWTSRV estimators. (a) GBP futures, (b) CHF futures and (c) EUR futures. PDF of $N(0,1)$ is shown for comparison.

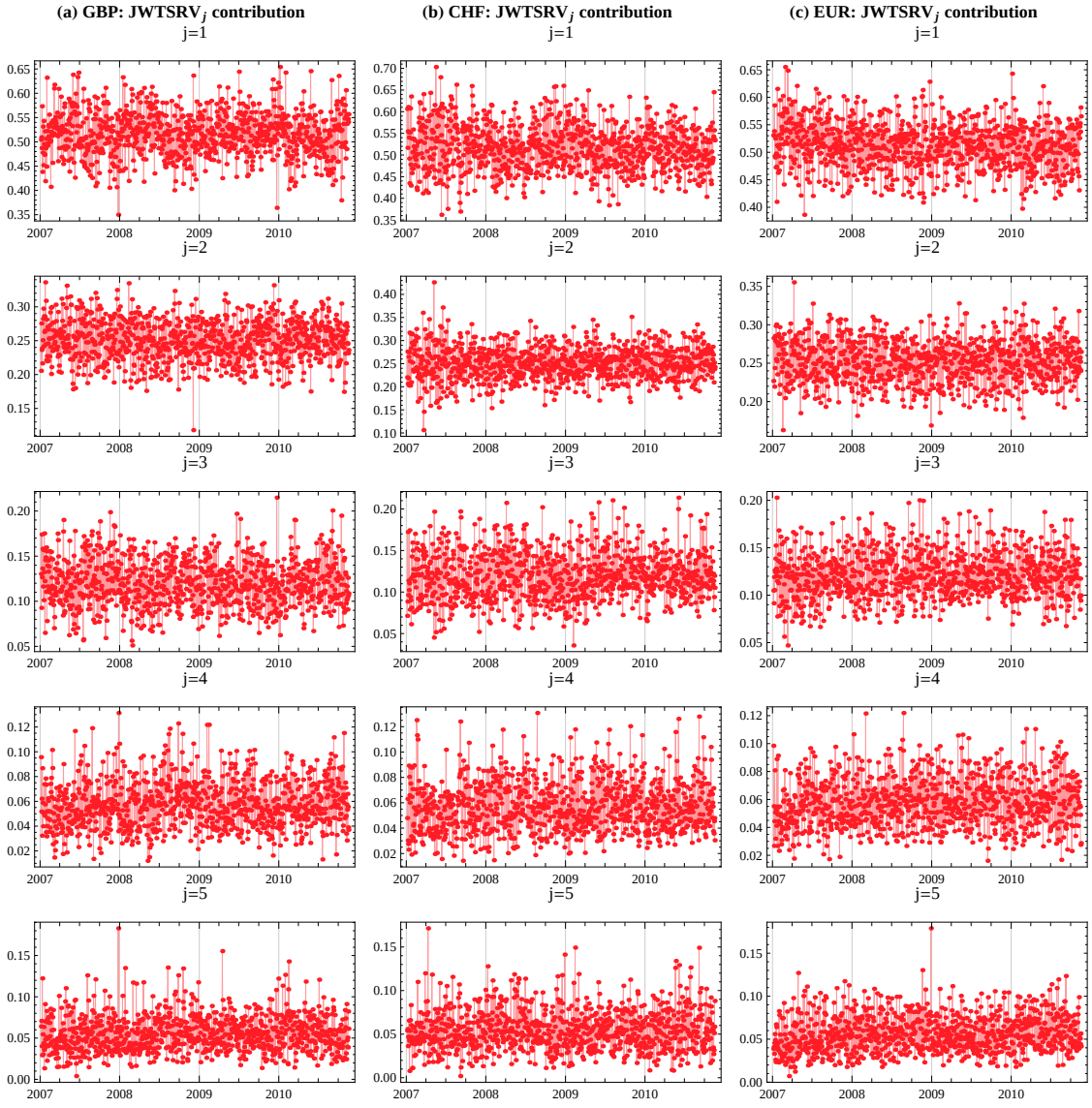


Figure 11.2: $JWTRSV_j$, $j = 1, \ldots, 5$, contributions of components of integrated variance IV_t corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day. (a) GBP futures (b) CHF futures (c) EUR futures.

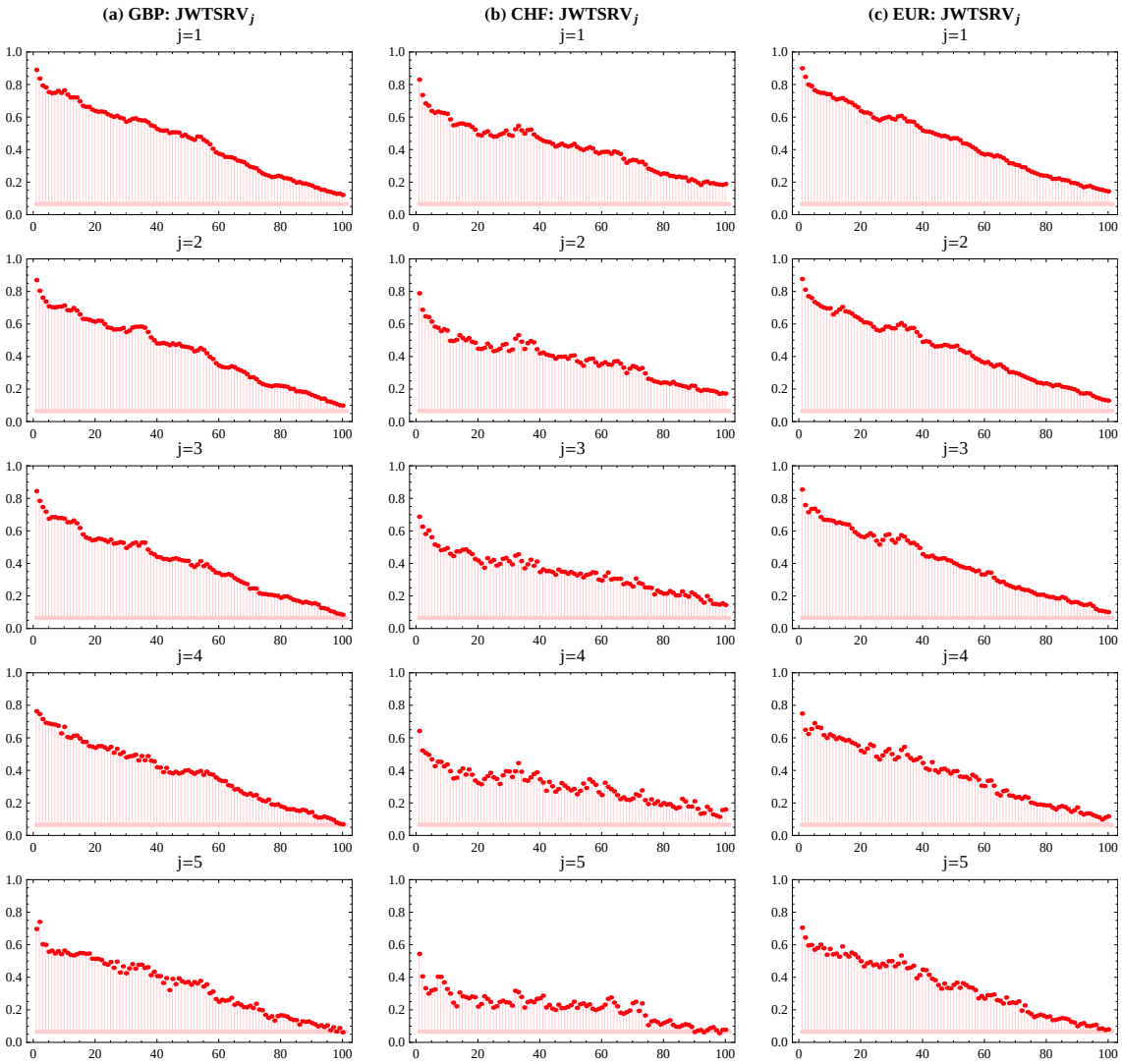


Figure 11.3: Autocorrelation functions of $JWTRSV_j$, $j = 1, \ldots, 5$, components of integrated variance IV_t corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day for (a) GBP futures (b) CHF futures and (c) EUR futures.

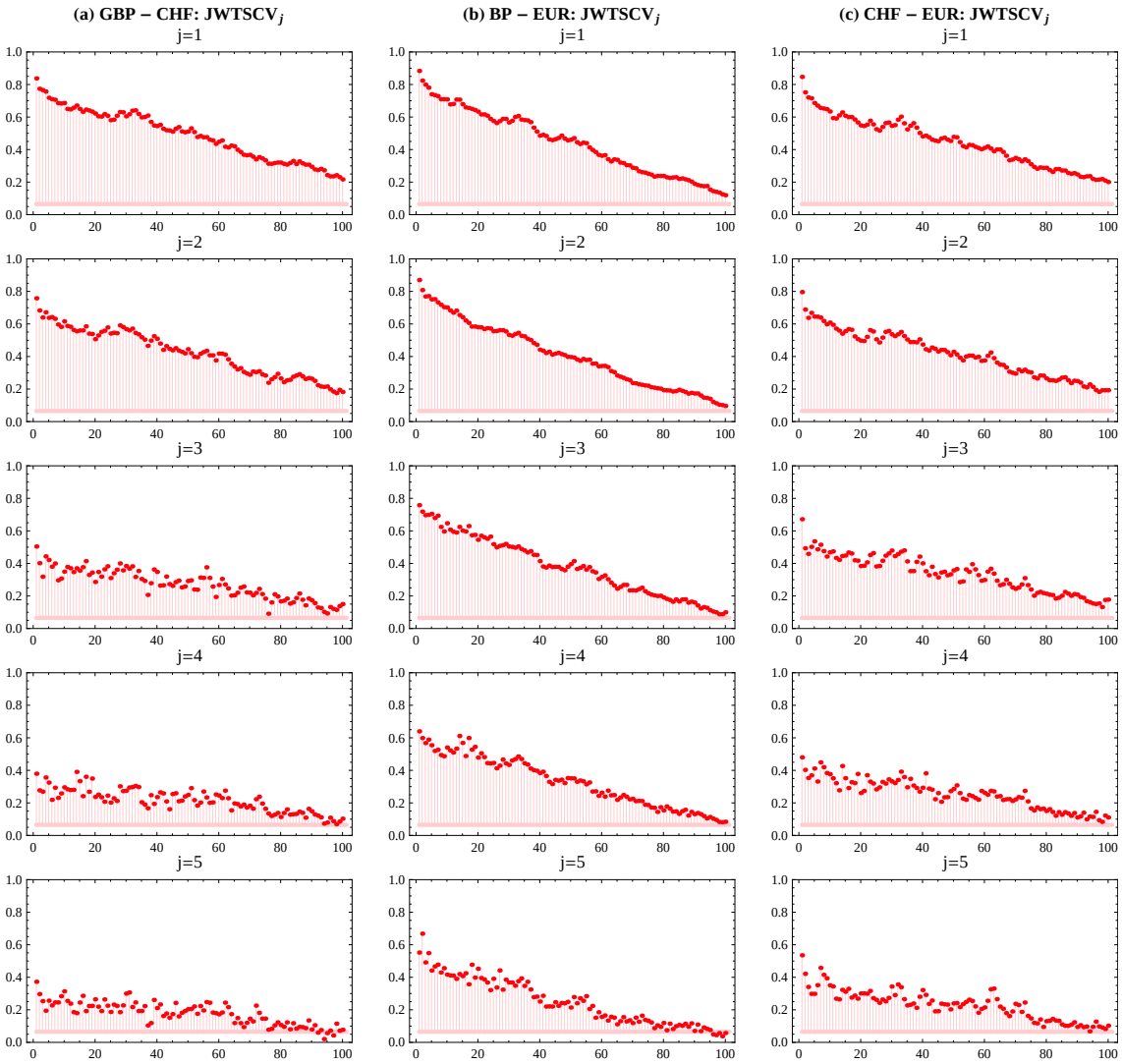


Figure 11.4: Autocorrelation functions of $JWTSCV_j$, $j = 1, \ldots, 5$, components of integrated covariance IC_t corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day for (a) GBP-CHF futures pair (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.

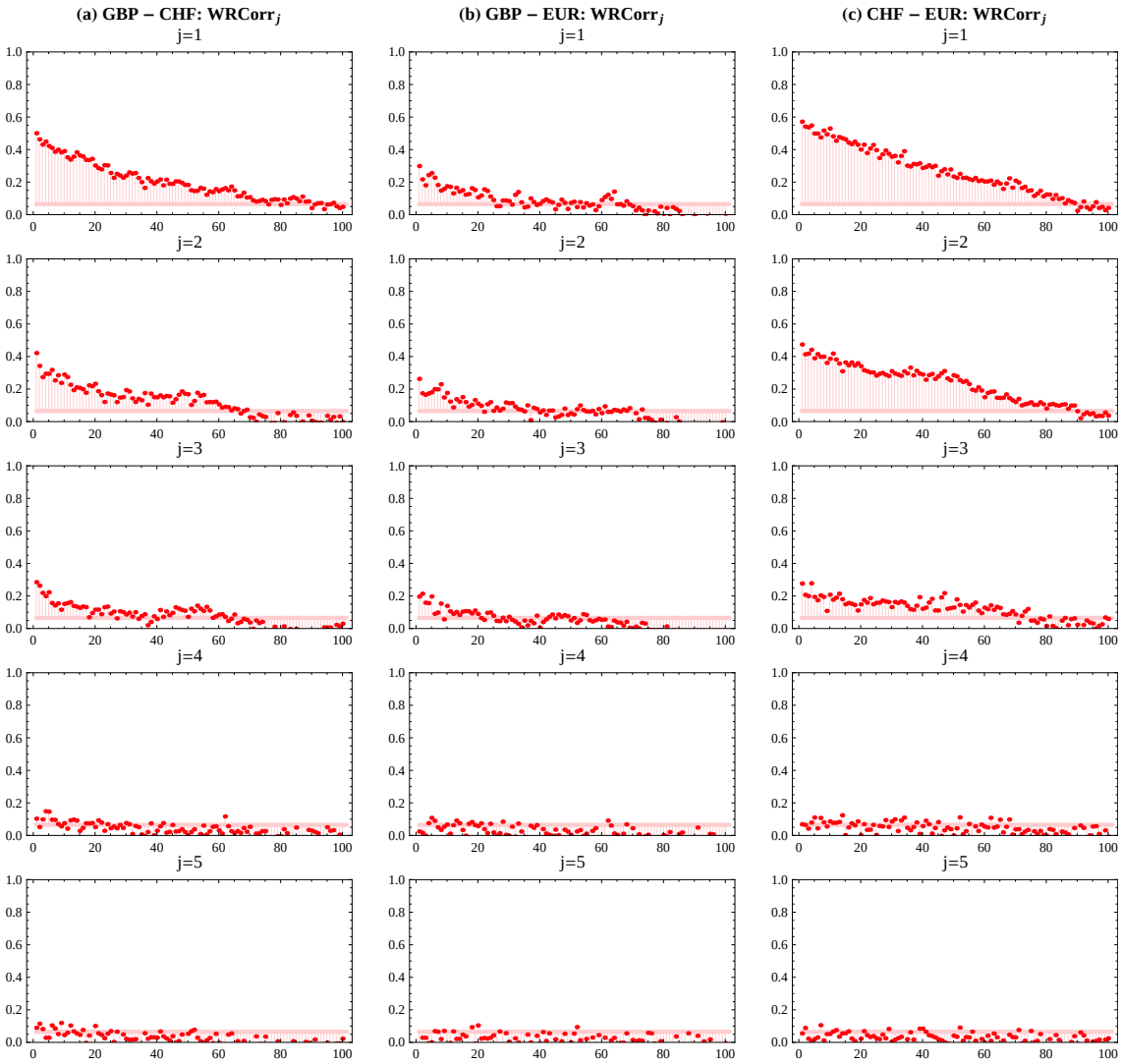


Figure 11.5: Autocorrelation functions of $WRCorr_j$, $j = 1, \ldots, 5$, components of correlations corresponding to investment horizons of 5–10 minutes, 10–20 minutes, 20–40 minutes, 40–80 minutes and 80 minutes up to 1 day for (a) GBP-CHF futures pair (b) GBP-EUR futures pair and (c) CHF-EUR futures pair.